

ABSTRACT

Master's thesis: 93 p., 11 fig., 32 tabl., 40 sources and 2 appendices.

Over the last two centuries, mankind increases data volumes all around the world. This process increases requirements on speed and quality of search engines.

Serious research around search engines started since 1954. In this years has been introduced the first Vector Space model. In 2013 Google announced "Word2Vec" model which is a de-facto standard in Vector Space representation of words.

The task is to devel multimodal text search system for search relevance evaluation. In this work were implemented and examined methods of Multimodal Information Retrieval compared different algorithms for spatial retrieval, frequency counting, and topic modeling.

The object is text search systems. The subject is multimodal text search systems.

Results:

- implemented multimodal text search system, which is based on asynchronous model;
- implemented disk-based spatial retrieval;
- generic implementation of document storage and retrieval.

Results of this work are applicable for as a search engine for Internet-stores, catalogs, informational websites, filesystems and everywhere, where there is need in high qulity search.

In the future it is useful to increase the number of ranking factors and extend existing language for working with data and retrieval operations.

MULTIMODAL INFORMATION RETRIEVAL, INFORMATION RETRIEVAL, MACHINE LEARNING, FULL-TEXT SEARCH, OKAPI-BM25, EM-ALGORITHMS, PROBABILISTIC INFERENCE, SEMANTIC TEXT ANALYSIS