

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра дискретной математики и алгоритмики

Аннотация к магистерской диссертации

**«Автоматический поиск релевантных текстовых документов
в многоязычной Интернет-среде»**

Новик Никита Сергеевич

Научный руководитель – доктор технических наук,
профессор Совпель И. В.

Минск, 2018

Реферат

Магистерская диссертация, 42 страниц, 10 рисунков, 10 источников.

ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА, ВЫДЕЛЕНИЕ КЛЮЧЕВЫХ СЛОВ, ИНТЕРНЕТ, МАШИННЫЙ ПЕРЕВОД, ЛИНГВИСТИЧЕСКИЙ ПРОЦЕССОР, РЕЛЕВАНТНЫЕ ДОКУМЕНТЫ.

Объект исследования – алгоритмы и способы поиска релевантных документов в многоязычной Интернет-среде.

Цель работы – разработать алгоритм выделения ключевых слов, разработать алгоритм перевода ключевых слов на другой язык, разработать систему для автоматического поиска релевантных текстовых документов с помощью некоторой поисковой службы, проанализировать полученные результаты.

В результате исследования разработаны и реализованы алгоритмы выделения ключевых слов и их перевода на другой язык, разработана автоматическая система поиска релевантных текстовых документов по заданному документу.

Методы исследования – анализ литературы, статистический и семантический анализ.

Область применения результатов – системы информационного поиска, классификации и анализ текстовых документов.

Abstract

Master's thesis, 42 pages, 10 pictures, 10 sources.

NATURAL LANGUAGE PROCESSING, KEYWORD EXTRACTION, INTERNET, MACHINE TRANSLATION, LINGUISTIC PROCESSOR, RELEVANT DOCUMENTS.

Object of study – algorithms and ways of searching for relevant documents in a multilingual Internet environment.

Purpose – develop an algorithm for extracting keywords, develop an algorithm for translating keywords into another language, develop a system for automatically searching for relevant text documents with the help of some search service, and analyze the results.

As a result algorithms for extraction keywords and their translation into another language have been developed and implemented, an automatic system for searching relevant text documents for a given document has been developed.

Research methods – literature analysis, statistical and semantic analysis.

Area of applications – information retrieval systems, text documents classification and analysis.