

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ
Кафедра информационных систем управления

Аннотация к дипломной работе

**«CROSS-LANGUAGE ФУНКЦИОНАЛЬНОСТЬ АВТОМАТИЧЕСКОГО
ПОИСКА В СЕТИ INTERNET РЕЛЕВАНТНЫХ ДОКУМЕНТОВ»**

Исаченко Дмитрий Александрович

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

Минск, 2017

РЕФЕРАТ

Дипломная работа, 52 с., 14 рис., 6 табл., 11 источников.

Ключевые слова: ПОИСК, АВТОМАТИЧЕСКАЯ ОБРАБОТКА ТЕКСТОВ, ПЕРЕВОД, РАЗРЕШЕНИЕ ЛЕКСИЧЕСКОЙ МНОГОЗНАЧНОСТИ, МАШИННОЕ ОБУЧЕНИЕ, ТЕЗАУРУС.

Объект исследования – алгоритмы извлечения ключевой информации из документа, алгоритмы разрешения лексической многозначности при переводе.

Цель работы – исследование методов поиска релевантных документов в многоязычной информационной среде, разработка мобильного клиента для поиска актуальных заданному документов, в том числе предоставленных на языке отличном от языка заданного документа.

Методы исследования – методы теории вероятности, математической статистики, интеллектуальный анализ данных, машинное обучение.

Результатом является выполненный обзор существующих решений в области поиска релевантных документов в многоязычной информационной среде, разработанное мобильное приложение под ОС Android, позволяющее для текстового документа/веб-страницы выполнить поиск релевантной информации в сети Internet, в том числе предоставленной на языке отличном от языка заданного текстового документа/веб-страницы.

Область применения: информационный поиск.

РЭФЕРАТ

Дыпломная работа, 52 с., 14 мал., 6 табл., 11 крыніц.

Ключавыя словы: ПОШУК, АЎТАМАТЫЧНАЯ АПРАЦОЎКА ТЭКСТУ, ПЕРАКЛАД, ДАЗВОЛ ЛЕКСІЧНАЙ ШМАТЗНАЧНАСЦІ, МАШЫННАЕ НАВУЧАННЕ, ТЭЗАЎРУС.

Аб'ект даследавання - алгарытмы здабывання ключавой інфармацыі з дакумента, алгарытмы дазволу лексічнай шматзначнасці пры перакладзе.

Мэта работы - даследаванне метадаў пошуку рэлевантных дакументаў у шматмоўным інфармацыйным асяроддзе, распрацоўка мабільнага кліента для пошуку актуальных дадзенаму дакументаў, у тым ліку на выдатных ад дадзенага дакумента мовах.

Метады даследавання - метады тэорыі верагоднасці, матэматычнай статыстыкі, інтэлектуальны аналіз дадзеных, машыннае навучанне.

Вынікам з'яўляецца выкананы агляд існуючых рашэнняў у вобласці пошуку рэлевантных дакументаў у шматмоўным інфармацыйным асяроддзі, распрацаванае мабільнае прыкладанне пад АС Android, якое дазваляе для тэкставага дакумента/вэб-старонкі выканаць пошук рэлевантнай інфармацыі ў сетцы Internet, у тым ліку прадстаўленай на мове выдатнай ад мовы дадзенага тэкставага дакумента/вэб-старонкі.

Вобласць прымянення: інфармацыйны пошук.

ABSTRACT

Diploma thesis, 52 p., 14 fig., 6 tabl., 11 references.

Keywords: SEARCH, AUTOMATIC TEXT PROCESSING, TRANSLATION, WORD-SENSE DISAMBIGUATION, MACHINE LEARNING, THESAURUS.

The object of research - algorithms of key information extraction, algorithms solving word-sense disambiguation problem during translation.

The objective is to study various methods of searching for relevant documents in a multilingual information environment, development of a mobile client for searching relevant documents, including those whose language are different from the language of source document.

The methods of research - methods of probability theory, mathematical statistics, data mining, machine learning.

The result is review of existing solutions in the field of searching for relevant documents in a multilingual information environment, developed mobile application for the Android OS, which allows for a text document/webpage to search for relevant information on the Internet, including those whose language are different from the language of the source text document/webpage.

The application area: information retrieval.