

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра технологий программирования

Аннотация к дипломной работе

**«Разработка методов и средств поддержки процессов
создания программного обеспечения анализа текстовых
данных»**

Казачёнок Дмитрий Вадимович

Научный руководитель – кандидат физико-математических
наук, доцент Зубович К.А.

2016

РЕФЕРАТ

Дипломная работа: 69 страниц, 18 таблиц, 22 рисунка, 21 источник, 3 приложения.

Ключевые слова: ОБРАБОТКА ТЕКСТОВЫХ ДАННЫХ, АНАЛИЗ ТЕКСТОВЫХ ДАННЫХ, МОРФОЛОГИЧЕСКИЙ АНАЛИЗ, СИНТАКСИЧЕСКИЙ АНАЛИЗ, СЕМАНТИЧЕСКИЙ АНАЛИЗ, ЛЕММАТИЗАЦИЯ, МАШИННОЕ ОБУЧЕНИЕ.

Объект исследования – методы и средства анализа текстовых данных.

Цель работы – исследовать методы и подходы в разработке программного обеспечения анализа текстовых данных, разработать программные средства поддержки процессов создания программного обеспечения морфологического анализа текстовых данных.

Методы исследования: изучение литературы, научных публикаций, документации программного обеспечения.

Результаты: исследованы методы и подходы в разработке программного обеспечения анализа текстовых данных; изучены, реализованы и проанализированы три модели лемматизации, применимые к разным языкам.

Область применения: информационно-аналитическая деятельность.

ABSTRACT

The volume of the degree work is 69 pages, 18 tables, 22 figures, 21 literature references, 3 applications.

Keywords: PROCESSING TEXT DATA, THE ANALYSIS OF TEXT DATA, MORPHOLOGICAL ANALYSIS, SYNTACTIC ANALYSIS, SEMANTIC ANALYSIS, LEMMATIZATION, MACHINE LEARNING.

Research object – methods and tools for analysis of text data.

Purpose of the degree work – to explore methods and approaches to software development analysis of text data, to develop software tools to support processes of creation of the software for morphological analysis of text data.

Research methods: the study of literature, scientific publications, software documentation.

Results: methods and approaches in the development of software analysis of text data were investigated; three models of lemmatization applicable to different languages have been studied, implemented and analyzed.

Scope: information-analytical activity.