

ОТЗЫВ

официального оппонента на диссертацию

БОЛЬШАКОВОЙ Светланы Анатольевны на тему:

«Совершенствование методов компьютерной обработки текстовой информации в аспекте задач, связанных с омонимией и синонимией», представленной на соискание ученой степени кандидата технических наук по специальности 2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки)

Актуальность избранной темы

В рамках развития исследований в области искусственного интеллекта, системы обработки естественного языка (NLP) сталкиваются с рядом фундаментальных проблем, среди которых омонимия и синонимия занимают центральное место. Омонимия приводит к множеству интерпретаций на этапе морфологического и синтаксического анализа, что замедляет обработку и снижает точность. Синонимия, как способность языка выражать один и тот же смысл разными словами или конструкциями, является ключевым механизмом для лексической адаптации и упрощения текстов. В русском языке это особенно выражено и достаточно сложно.

Актуальность данной темы обусловлена возрастающими объемами текстовой информации, особенно в условиях цифровизации и бурного роста контента в интернете. Без эффективных методов анализа и интерпретации таких данных невозможно построение качественных систем машинного перевода, поиска, классификации, автоматического реферирования и вопросно-ответных систем.

Таким образом, тема диссертационного исследования Большаковой С.А. «Совершенствование методов компьютерной обработки текстовой информации в аспекте задач, связанных с омонимией и синонимией» является актуальной и имеет важное практическое значение.

Оценка содержания и оформления диссертации

Диссертационная работа содержит 172 страницы машинописного текста и состоит из введения, четырех разделов, заключения, списка литературы из 121 источника на 15 страницах и 7 приложений на 29 страницах. Основной



текст, изложенный на 142 страницах, иллюстрируется 9 рисунками и содержит 19 таблиц.

Во введении обоснована актуальность темы диссертационной работы, сформулирована цель и задачи исследований, представлена научная новизна, теоретическая и практическая значимость работы.

В первом разделе представлен комплексный анализ современных технологий и методов работы с синонимией и омонимией в контексте автоматической обработки текста. Основное внимание уделяется методам автоматической адаптации текста, технологиям морфологического анализа и различным подходам к разрешению омонимии (включая правила, статистику, машинное обучение и нейросетевые модели). Также рассматриваются современные языковые модели и их применение для решения задач обработки естественного языка. Обоснована актуальность разработки специализированных решений для русского языка, т.к. действующие инструменты ориентированы на языки романо-германской группы.

Во втором разделе работы описан процесс создания и реализации алгоритма для определения морфологических характеристик слов и их лемматизации. Описана структура словаря русского языка, организованного в виде эффективного префиксного дерева, и специальный алгоритм индексирования его строк, позволяющий мгновенно находить нужные словоформы.

В третьем разделе диссертации представлен гибридный подход к решению проблемы автоматического разрешения омонимии, объединяющий два ключевых компонента: декларативные знания в форме специализированных словарей и процедурные знания через систему продукционных правил. Разработанный метод направлен на снятие частеречной омонимии для сложных языковых конструкций: предикативов и их словосочетаний, деепричастий, групп «наречие-существительное».

Для оценки эффективности предложенных методов снятия омонимии проведено их сравнение с двумя ведущими системами обработки русского языка – MyStem и Rymorphy2. Предложенные методы показывают существенно более высокий результат (99,3% точности) благодаря учёту особенностей русского языка при обработке различных частей речи.

В четвертом разделе описана реализация единой системы обработки и анализа текстовой информации. Автором описаны разработанные методы

лексической адаптации текста с помощью синонимических замен; автоматического разбиения текста на семантически однородные фрагменты и автоматического создания элементов плана за счет использования словаря отглагольных существительных. Эффективность предложенного метода лексической адаптации текста с использованием базы синонимов и системы продукционных правил показана на примерах из Национального корпуса русского языка. Метод демонстрирует впечатляющую точность (96%) при обработке как отдельных лексических единиц, так и устойчивых словосочетаний. Качество адаптации текста обеспечивается отсутствием смысловых искажений при проведении синонимических замен. Грамматическая корректность преобразований поддерживается на высоком уровне благодаря продуманному механизму замены.

В заключении автор приводит общие выводы по результатам диссертационного исследования. Текст диссертации изложен последовательно, грамотно, корректно, достаточно аргументирован.

Содержание автореферата соответствует содержанию диссертации.

Степень обоснованности, достоверность и новизна научных положений, выводов и рекомендаций

Обоснованность и достоверность научных положений и выводов обеспечивается комплексным подходом к исследованию, использованием современных методов анализа, проверкой на репрезентативном корпусе текстов и подтверждением работоспособности разработанных алгоритмов.

Диссертация содержит оригинальные решения, что позволяет высоко оценить уровень научной новизны работы, подтвержденной публикациями.

Соискателем проведена положительная апробация результатов исследования на международных и всероссийских научных конференциях и в научных публикациях. Всего по теме исследования было опубликовано 17 публикаций, в том числе 2 статьи в ведущих изданиях согласно перечню ВАК РФ. Получено 1 свидетельство о регистрации программы для ЭВМ. Результаты исследования внедрены и использованы в научно-исследовательской деятельности ФГБНУ «ИПИИ» и при обработке массивов текстовой информации в ФГБНУ "РАНИМИ".

Таким образом, научные положения и выводы, которые содержатся в диссертации Большаковой С.А., являются в достаточной степени обоснованными и достоверными.

Замечания

При общей положительной оценке диссертационного исследования Большаковой С.А. отметим ряд дискуссионных вопросов и замечаний.

1. В разделе о снятии омонимии не представлен детальный анализ возможных ошибок в сложных случаях, когда контекст недостаточно определен.

2. При описании базы синонимов не раскрыт механизм автоматического обновления данных. Актуальность такого словаря требует регулярного пополнения и актуализации.

3. В описании методов адаптации текста не рассмотрены случаи работы с жаргонизмами и профессиональной терминологией.

4. В описании алгоритма разбиения текста на абзацы не учтены особенности работы с текстами разных жанров (научные статьи, художественная литература, новостные материалы).

5. В тексте встречаются лингвистические погрешности, в частности несовпадение падежей в сложных предложениях, пропуски предлогов.

Вышеуказанные замечания и дискуссионные вопросы не снижают научной ценности проведенного исследования и общую положительную оценку диссертационной работы Большаковой С.А.

Заключение

Диссертация является завершенной научной работой, в которой получено решение актуальной и практически полезной научно-технической проблемы совершенствования методов компьютерной обработки текстовой информации в аспекте задач, связанных с омонимией и синонимией. Цель работы по повышению эффективности обработки и анализа текстовой информации достигнута, все поставленные задачи успешно решены.

По области исследований, содержанию научных положений и выводов, существу полученных результатов диссертационная работа соответствует паспорту специальности 2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки) по пп. 3, 4, 5.

Диссертационная работа отвечает требованиям пп. 9, 10, 11, 13, 14 Положения о присуждении ученых степеней, утвержденного постановлением Правительства РФ от 24 сентября 2013 г. № 842 с изменениями (редакция от 16.10.2024), а её автор, Большакова Светлана Анатольевна, заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки).

Официальный оппонент:

Доктор технических наук по специальности 05.13.01 «Системный анализ, управление и обработка информации (по отраслям)», профессор, профессор кафедры «Системы обработки информации и управления» Федерального государственного автономного образовательного учреждения высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»

Варламов Олег Олегович

Я, Варламов Олег Олегович, даю согласие на автоматизированную обработку моих персональных данных, указанных в отзыве, и размещение их на сайте ФГБОУ ВО ДонНТУ.

01.12.2025 г.

Варламов О. О.

Адрес: 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1, Федеральное государственное автономное образовательное учреждение высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)».

Тел.: +7 499 263 63 91, +7 926 276 76 45.

Эл. почта: bauman@bmstu.ru, varlamovoo@bmstu.ru