ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА
кандидата технических наук Кузнецова Владислава Владимировича
на диссертационную работу Воробьева Никиты Григорьевича
на тему: «Методы и алгоритмы рубрикации документов на основе векторно-графовой модели», представленную на соискание ученой степени кандидата технических наук по специальности 2.3.8. «Информатика и информационные процессы».

АКТУАЛЬНОСТЬ ТЕМЫ ДИССЕРТАЦИОННОГО ИССЛЕДОВАНИЯ

Каждые семнадцать лет объём научных публикаций человечества удваивается. Это означает экспоненциальный рост – а значит, и инфраструктура научного знания должна в своём развитии соответствовать таким темпам. Поэтому всё более востребованными становятся методы автоматической классификации документов, выявления скрытых семантических связей и интеллектуального анализа текстовой информации.

Задача автоматической классификации документов не нова, ей уже около семидесяти лет. Первый метод машинного индексирования был разработан в 1957 году и основывался на статистическом анализе частот встречаемости слов. В настоящее время в исследовательских работах для данной задачи адаптируют архитектуру Трансформер и базирующиеся на ней большие языковые модели. Однако такой подход имеет множество недостатков, связанных с особенностями этой архитектуры – и прежде всего её универсальностью. В то же время существуют узкоспециальные методы и алгоритмы индексации статей, использующие тезаурусы и графовые модели, однако их развитие ограничивается сложностью автоматического построения таких структур и проблемами масштабирования.

Одним из направлений решения данной задачи является разработка гибридных нейросетевых моделей, способных объединить преимущества существующих подходов. Такие модели на уровне архитектуры учитывают как семантические связи между языковыми сущностями, так и статистические закономерности. Решению данной научной задачи посвящена диссертационная работа Никиты Григорьевича Воробьёва. В данной работе автором предложена новая специализированная гибридная архитектура, интегрирующая графовый подход и векторное представление слов, а также метод её обучения. Полученные результаты позволяют рассматривать данную работу как актуальное исследование, вносящее вклад в развитие методов обработки естественного языка, интеллектуального анализа текстов и создания лексико-семантических информационных ресурсов.

СТЕПЕНЬ ОБОСНОВАННОСТИ И ДОСТОВЕРНОСТИ НАУЧНЫХ ПОЛОЖЕНИЙ, СФОРМУЛИРОВАННЫХ В ДИССЕРТАЦИИ

Обоснованность научных положений, результатов и выводов, представленных в диссертации Н.Г. Воробьёва, определяется корректностью постановки цели и задач, хорошо структурированной методикой исследования, использованием актуальных методов статистического анализа локальных контекстов, аппарата графовых грамматик и семантических сетей, методов векторного представления языковых сущностей, а также вычислительных экспериментов, подтверждающих достижение поставленных целей.

Диссертационная работа опирается на тщательно выполненный анализ литературы, который охватывает ключевые публикации тематической области, как российские, так и иностранные. Результат анализа изложен последовательно, начиная от фундаментальных работ по автоматической классификации текстов и заканчивая современными работами по генеративным нейросетевым моделям. Решения автора, которые он предложил в своей работе, логично включены в линию эволюции описанных в литературе подходов.

Достоверность полученных результатов обеспечивается применением формализованного математического аппарата, методов статистического анализа локальных контекстов и графового представления семантических структур, а также вычислительными экспериментами. Результаты экспериментов подтверждают эффективность решения в виде графовой языковой модели (GLM) применительно к задаче рубрикации текстовой информации и выбора ответов в диалоговых системах, что позволяет считать сформулированные в диссертации выводы обоснованными и достоверными.

Основные результаты исследования опубликованы в рецензируемых научных изданиях и апробированы на всероссийских и международных научных конференциях. Представленные в диссертации и приложениях к ней данные вычислительных экспериментов подтверждают работоспособность предложенных методов и алгоритмов, согласуются с теоретическими положениями работы и подтверждают достоверность результатов.

Основные научные результаты, представленные в диссертации, получены автором самостоятельно. Диссертация включает введение, четыре главы, заключение, список литературы из 93 наименований и два приложения. Общий объём работы составляет 173 страницы; она содержит 7 таблиц, 34 рисунка и графика, что обеспечивает достаточную полноту и наглядность изложения материала.

Таким образом, научные положения и выводы, сформулированные в диссертации Н. Г. Воробьёва, являются обоснованными, достоверными и не вызывающими сомнений.

НАУЧНАЯ НОВИЗНА И ПРАКТИЧЕСКАЯ ЗНАЧИМОСТЬ РАБОТЫ

К наиболее важным результатам, определяющим научную новизну диссертационной работы Воробьёва Н.Г., можно отнести следующие положения:

1. Предложен метод автоматизированного формирования информационно-поискового тезауруса, обеспечивающий извлечение и систематизацию семантических отношений между терминами предметной области на основе анализа текстовых данных.
2. Разработан подход к оценке качества информационно-поисковых тезаурусов, основанный на анализе их структурных характеристик и позволяющий количественно сравнивать экспертные и автоматически формируемые тезаурусные системы.
3. Создана графовая языковая модель, использующая аппарат графовых грамматик и локальных контекстов для представления лексико-семантических структур и отражения взаимосвязей между языковыми единицами.
4. Предложен алгоритм преобразования графовой языковой модели в векторное пространство, основанный на модифицированной процедуре Вайсфейлера–Лемана и сохраняющий структурные особенности графового представления текста.
5. Разработан метод автоматической рубрикации документов, использующий оценку структурного сходства графов и их спектральных характеристик для повышения точности классификации текстовой информации.
6. Разработан метод объединения графовых и векторных способов представления текстовой информации, обеспечивающий совместное использование их преимуществ в интеллектуальных системах обработки естественного языка.
7. Предложены способы интеграции графовых языковых моделей с современными нейросетевыми архитектурами, ориентированные на решение задач тематической классификации документов и интеллектуального выбора ответов в диалоговых системах.

Практическая значимость диссертационного исследования определяется разработкой методов и алгоритмов, предназначенных для решения задач интеллектуальной обработки текстовой информации. Предложенные автором решения позволяют значительно повысить качество автоматической рубрикации документов, улучшить результаты семантического поиска и обеспечить более эффективное использование лексико-семантических связей при анализе текстов. Разработанные подходы могут быть использованы при создании интеллектуальных поисковых систем, диалоговых сервисов, корпоративных баз знаний, лексико-семантических информационных ресурсов, а также специализированных информационно-аналитических систем.

Предложенные в работе методы ориентированы на прямое практическое применение и могут быть использованы в рамках существующей программной инфраструктуры обработки естественного языка, включая рубрикацию научных статей.

Практическая ценность полученных результатов подтверждается их внедрением в научную и образовательную деятельность Федерального государственного автономного образовательного учреждения высшего образования «Московский политехнический университет», что подтверждено соответствующими актами внедрения.

ЗНАЧИМОСТЬ РЕЗУЛЬТАТОВ ДИССЕРТАЦИОННОЙ РАБОТЫ ДЛЯ НАУКИ И ПРАКТИКИ

Изложенные в диссертационной работе результаты Н.Г. Воробьёва являются новыми и представляют научный интерес. Полученные автором результаты имеют значение для дальнейше-

го развития методов обработки естественного языка, интеллектуального анализа текстовой информации и формирования лексико-семантических информационных ресурсов. Разработанные подходы к построению графовых языковых моделей, оценке качества тезаурусных структур и интеграции графовых и векторных представлений текста расширяют возможности современных методов семантического анализа и открывают перспективы дальнейших исследований в области гибридных языковых моделей.

Значимость полученных результатов определяется не только их научной новизной, но и возможностями практического применения. Предложенные методы и алгоритмы могут найти применение в интеллектуальных поисковых системах, средствах автоматической рубрикации документов, диалоговых сервисах и других системах обработки естественного языка.

АНАЛИЗ ДИССЕРТАЦИИ ПО ГЛАВАМ

Во введении диссертационной работы актуальность выбранной темы, сформулированы цель и задачи исследования, определены объект и предмет исследования, представлены положения, выносимые на защиту, явно сформулирована научная новизна, теоретическая и практическая значимость работы, а также приведены сведения об апробации полученных результатов и публикациях автора.

В первой главе рассмотрены современные методы и модели обработки естественного языка, а также существующие подходы к построению тезаурусов и представлению текстовой информации. Проанализированы методы формирования локальных контекстов и способы семантического представления текстов. На основании проведённого анализа обоснована целесообразность разработки гибридных графовых языковых моделей, сочетающих структурные и статистические способы описания языка.

Во второй главе разработан функционал для оценки структуры и свойств тезаурусов и графовых языковых моделей. Показано, что разработанный функционал является метрикой. В рамках метрики рассмотрены такие характеристики, как размер графа, плотность связей, коэффициент кластеризации, средняя длина пути и энтропия словаря. Проведено сравнение экспертных и автоматически сформированных тезаурусов, что позволило выявить особенности их структуры и подтвердить возможность применения предложенных метрик для анализа и совершенствования графовых моделей.

В третьей главе предложена графовая языковая модель предметной области, основанная на локальных контекстах и графовых грамматиках. Разработаны методы построения графов и алгоритмы их преобразования в векторное представление для последующего использования в нейросетевых системах. Предложены способы применения графовой модели для решения задач рубрикации документов и выбора ответов в диалоговых системах.

В четвёртой главе представлены результаты вычислительных экспериментов и рассмотрены вопросы практического применения разработанных методов. Показана возможность интеграции графовых представлений с существующими интеллектуальными системами обработки текстов. Полученные результаты подтверждают работоспособность предложенных алгоритмов и эффективность их применения при решении прикладных задач.

В заключении обобщены результаты выполненного исследования и сформулированы основные выводы, включая итоги теоретических и экспериментальных исследований, подтверждена эффективность предложенных методов построения графовых языковых моделей и обоснована целесообразность их применения для решения тематических задач.

Текст исследования отличается ясным и корректным изложением материала и в целом логичной структурой. Все главы взаимосвязаны и формируют единую иерархию в рамках поставленной цели. Полученные результаты в совокупности позволяют решить научную задачу, связанную с разработкой методов построения и применения графовых языковых моделей для обработки текстовой информации.

В приложениях приведены дополнительные материалы, дополняющие раскрывающие подробности численных экспериментов работы и подробно раскрывающие отдельные результаты проведённого исследования.

ЗАМЕЧАНИЯ ПО ДИССЕРТАЦИОННОЙ РАБОТЕ

Несмотря на, несомненно, сильные результаты, работа не лишена и недостатков.

1. Разработанная метрика и формулы для количественной оценки эффективности тезаурусов является самостоятельным результатом и не используется при разработке гибридной архитектуры в рамках автоматизированного формирования тезауруса
2. Рисунки 3 и 4, расположенные на странице 74, выбиваются из общего порядка нумерации, будучи расположены после рисунка 6. При этом иллюстрация на рисунке 4 выполнена с неточностями и помарками.
3. На странице 104 в описании анализируемого набора данных дана ссылка на статью с номером 82 – при этом сам набор данных дан в приложениях.
4. Имеются недочёты в оформлении списка использованных источников в записях с номерами 3, 34, 41, 42 и 52, где фамилия автора приведена после инициалов.
5. Исследование бы выиграло от анализа других архитектур, объединяющих графы и большие языковые модели, такие как Graphormer.

ЗАКЛЮЧЕНИЕ

Упомянутые замечания не несут принципиального характера, не снижают научной ценности диссертационной работы и не влияют на её общую положительную оценку. При этом они не снижают научной ценности диссертационного исследования и не меняют общей положительной оценки работы.

В целом диссертация Воробьёва Никиты Григорьевича представляет собой завершённую научно-квалификационную работу, в которой решена актуальная задача разработки методов построения и применения графовых языковых моделей для автоматизированного автоматизированного формирования информационно-поискового тезауруса и интеллектуальной обработки текстовой информации. Полученные автором результаты обладают научной новизной, теоретической и практической значимостью, а выводы и положения диссертации являются обоснованными и достоверными.

По своему содержанию, научному уровню, объёму выполненных исследований и значимости полученных результатов диссертационная работа соответствует требованиям п. 9 «Положения о порядке присуждения учёных степеней», утверждённого постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842, предъявляемым к диссертациям на соискание учёной степени кандидата наук, а её автор, Воробьёв Никита Григорьевич, заслуживает присуждения учёной степени кандидата технических наук по специальности 2.3.8. – «Информатика и информационные процессы».

Официальный оппонент:
кандидат технических наук, ведущий программист федерального государственного учреждения "Федеральный исследовательский центр "Информатика и управление" Российской академии наук"

Подпись _____ Владислав Владимирович Кузнецов

Подпись заверяю _____ заместитель директора ФИЦ ИУ РАН по научной работе Ростислав Валерьевич Разумчик

Дата « 15 » июня 2026г. М.П.

Федеральный исследовательский центр "Информатика и управление" российской академии наук.
Адрес: 119333, Москва, Вавилова, д.44, кор.2.
Телефон: +7 (499) 135-62-60, e-mail: frccsc@frccsc.ru.

Шифр и наименование научной специальности в соответствии с номенклатурой, по которой была защищена диссертация лица, представившего отзыв:
05.13.17 Теоретические основы информатики.

С отзывом ознакомлен 15 июня 2026 г.