ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА 99.2.113.02,
СОЗДАННОГО НА БАЗЕ ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО ОБРАЗОВАТЕЛЬНОГО УЧРЕЖДЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ «РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ В.Ф. УТКИНА» МИНОБРНАУКИ РОССИИ, ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО ОБРАЗОВАТЕЛЬНОГО УЧРЕЖДЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ «ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ» МИНОБРНАУКИ РОССИИ ПО ДИССЕРТАЦИИ НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ КАНДИДАТА НАУК

Аттестационное дело № _____
решение диссертационного совета от 29.06.2026 №18

о присуждении Воробьеву Никите Григорьевичу, гражданину Российской Федерации, учёной степени кандидата технических наук.

Диссертация «Методы и алгоритмы рубрикации документов на основе векторно-графовой модели» **по специальности** 2.3.8. – «Информатика и информационные процессы» (технические науки) **принята к защите** 27.04.2026 г. (протокол заседания № 11) диссертационным советом 99.2.113.02, **созданным на базе** федерального государственного бюджетного образовательного учреждения высшего образования «Рязанский радиотехнический университет имени В.Ф. Уткина» Минобрнауки России, 390005, Рязанская обл., г. Рязань, ул. Гагарина, 59/1 и федерального государственного бюджетного образовательного учреждения высшего образования «Пензенский государственный технологический университет» Минобрнауки России, 440039, Пензенская обл. Пенза, проезд Байдукова/ул. Гагарина, д. 1А/11, приказ о создании диссертационного совета № 533/НК от 24 марта 2023 г

Соискатель Воробьев Никита Григорьевич, 28 января 1998 года рождения. В 2025 году соискатель **окончил** аспирантуру в Федеральном

государственном автономном образовательном учреждении высшего образования «Московский политехнический университет» по специальности 09.06.01 «Информатика и вычислительная техника».

Работает преподавателем кафедры «Инфокогнитивные технологии» в Федеральном государственном автономном образовательном учреждении высшего образования «Московский политехнический университет».

Диссертация выполнена на кафедре «Инфокогнитивные технологии» Федерального государственного автономного образовательного учреждения высшего образования «Московский политехнический университет».

Научный руководитель – Филиппович Юрий Николаевич, кандидат технических наук, доцент, профессор кафедры «Инфокогнитивные технологии» Федерального государственного автономного образовательного учреждения высшего образования «Московский политехнический университет».

Официальные оппоненты:

Конявский Валерий Аркадьевич, доктор технических наук, Академик РАЕН, академик АЭН РФ, Федеральное государственное автономное образовательное учреждение высшего образования «Московский физико-технический институт (национальный исследовательский университет)», г. Долгопрудный, кафедра Защиты информации, заведующий кафедрой;

Кузнецов Владислав Владимирович, кандидат технических наук, Федеральное государственное учреждение "Федеральный исследовательский центр "информатика и управление" Российской Академии Наук", г. Москва, ведущий программист

дали положительные отзывы на диссертацию.

Ведущая организация Федеральное государственное автономное образовательное учреждение высшего образования "Московский государственный технический университет имени Н.Э. Баумана (Национальный исследовательский университет)", г. Москва, **в своем положительном отзыве**, обсужденном и одобренном на заседании кафедры

систем обработки информации и управления 28 мая 2026 г., протокол №10, **подписанном** заведующим кафедрой систем обработки информации и управления, кандидатом технических наук, доцентом Тереховым Валерием Игоревичем и утвержденным первым проректором по науке и цифровому развитию ФГАОУ ВО «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)», д.э.н., профессором Дроговозом Павлом Анатольевичем, указал, что диссертационная работа соответствует указанной специальности 2.3.8. – «Информатика и информационные процессы» (технические науки), удовлетворяет требованиям «Положения о присуждении ученых степеней», предъявляемым ВАК при Минобрнауки России к диссертациям на соискание ученой степени кандидата технических наук, является законченной научно-квалификационной работой, содержащей решение актуальной научной задачи разработки гибридных графово-векторных языковых моделей для повышения эффективности автоматической обработки текстов в задачах рубрикации документов и семантического анализа, что заключается в развитии методов графового моделирования языка и разработке гибридного подхода к построению языковых моделей и обеспечивает повышение точности рубрикации документов и может быть интегрировано в существующие системы обработки естественного языка и искусственного интеллекта и имеет важное значение для развития соответствующей области исследования, а ее автор заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.8. – «Информатика и информационные процессы» (технические науки).

Соискатель имеет 7 опубликованных работ, в том числе по теме диссертации опубликовано 7 работ, из них в рецензируемых научных изданиях опубликовано 2 работы, среди которых 2 статьи опубликованы в изданиях, входящих в перечень ВАК РФ, 4 свидетельства о регистрации программ для ЭВМ. Общий объём опубликованных работ по теме диссертации составляет 2,41 печатных листов, авторский вклад – 71,44%.

Недостоверных сведений об опубликованных соискателем работах в диссертации не обнаружено.

Наиболее значимые работы:

1. Воробьев Н.Г. СРАВНЕНИЕ ЭФФЕКТИВНОСТИ МЕТОДОВ СБОРА И ИСПОЛЬЗОВАНИЯ ЕСТЕСТВЕННО-ЯЗЫКОВЫХ МОДЕЛЕЙ В МЕДИЦИНЕ / Н.Г. Воробьев // Вестник РГРТУ. – 2025. – № 93. – с. 200-212.

2. Воробьев Н.Г. ПРИМЕНЕНИЕ МЕТОДОВ RETRIEVAL-AUGMENTED GENERATION ДЛЯ АВТОМАТИЗАЦИИ АНАЛИЗА СЕТЕВОЙ ИНФРАСТРУКТУРЫ / Н.Г. Воробьев // XXI век: ИТОГИ 5 ПРОШЛОГО И ПРОБЛЕМЫ НАСТОЯЩЕГ плюс. — 2025. — №3(71) т.14, 2025. - с. 65-70

3. Воробьев Н.Г. The comparison of local context collection systems and methods / Н. Г. Воробьев, Ю. Н. Филиппович, Е. А. Пшехотская, А. Ю. Филиппович // Third International Conference on Optics, Computer Applications, and Materials Science (CMSD-III 2023). — Dushanbe, 20–22 декабря 2023 года. Vol. 13065. – Washington: SPIE-SOC PHOTO-OPTICAL INSTRUMENTATION ENGINEERS, 2024. — P. 1306514. – DOI 10.1117/12.3025183.

Воробьевым Н.Г. исследованы методы создания информационно поисковых тезаурусов и предложен алгоритм создания тезауруса предметной области на основе метода обратного индекса статистических данных вхождения слов в локальный контекст.

На диссертацию и автореферат поступили отзывы:

1. Федеральное государственное бюджетное образовательное учреждение высшего образования "МИРЭА - Российский технологический университет, г. Москва. Отзыв подписан Берковым Николаем Андреевичем, кандидатом технических наук, доцентом кафедры высшей математики.

2. Общество с ограниченной ответственностью «АНАЛИТНАУЧЦЕНТР», г. Москва. Отзыв подписан Варварицей Владиславом Петровичем, доктором технических наук, профессором.

3. Публичное акционерное общество «Сбербанк России», Блок Технологии, г. Москва. Отзыв подписан Рабиновичем Борисом Ильичом, кандидатом технических наук.

4. Федеральное государственное бюджетное образовательное учреждение высшего образования «Калужский государственный университет им. К.Э. Циолковского». Отзыв подписан Раевским Владимиром Алексеевичем, кандидатом технических наук, доцентом кафедры Информатики и информационных технологий.

5. Общество с ограниченной ответственностью «ГАММАФЛЕКС», г. Щёлково Московской области. Отзыв подписан Дубашовым Алексеем Евгеньевичем, кандидатом технических наук, начальником информационно-технического отдела.

6. Акционерное общество «Атомкомплекс», г. Москва. Отзыв подписан Вашуркиным Дмитрием Владимировичем, кандидатом технических наук.

7. Акционерное общество «Научно-производственное предприятие «Радий», г. Москва. Отзыв подписан Бавижевым Михаилом Данильевичем, доктором физико-математических наук, профессором, Лауреатом Государственной премии РФ, вице-президентом АО «НПП «Радий», директором НИЦ.

8. ФГБОУ ВО «Московский автомобильно-дорожный государственный технический университет (МАДИ)». Отзыв подписан Исмаиловым Мухамаджоном Идибоевичем, кандидатом технических наук, доцентом, доцентом кафедры «Автоматизированные системы управления».

В поступивших отзывах отмечается новизна представленных методов и алгоритмов, научная обоснованность положений, выносимых на защиту, теоретическая и практическая ценность работы.

В отзывах указаны следующие замечания: представляется целесообразным более подробно раскрыть влияние параметров графовой модели (пороговой фильтрации, глубины контекстного окна, предварительной обработки текстов), принципы формирования

семантических связей, механизмы обработки многозначных лексических единиц, а также оценить чувствительность разработанных методов к выбору этих параметров; экспериментальную часть работы можно было бы дополнить более подробным анализом характеристик использованных корпусов, причин возникновения ошибок рубрикации, устойчивости предложенных алгоритмов к шумовым и неполным данным, вычислительных затрат, масштабируемости метода, а также сравнением с дополнительными современными моделями обработки естественного языка; в работе недостаточно полно раскрыты вопросы использования графовой языковой модели в задачах интеллектуального поиска, генерации и аннотирования текстов, адаптации к динамически изменяющимся предметным областям, совместного обучения с современными нейросетевыми архитектурами и сопоставления разработанных методов с современными графовыми и трансформерными подходами; автором предложен оригинальный подход к объединению графовых и векторных представлений текста. Однако в диссертации ограниченное внимание уделено сравнению разработанного метода с современными трансформерными архитектурами последнего поколения, что могло бы дополнительно продемонстрировать преимущества и ограничения предложенного решения; в диссертации приведены результаты оценки качества рубрикации документов, однако анализ устойчивости предложенных алгоритмов к наличию шумовых данных и ошибок в исходных текстах представлен в ограниченном объёме; в работе показана эффективность разработанных методов для тематической рубрикации документов. Вместе с тем дополнительный интерес представляла бы оценка возможностей их применения для решения смежных задач обработки естественного языка, таких как автоматическое аннотирование.

Выбор официальных оппонентов и ведущей организации обосновывается их признанным научным авторитетом в рассматриваемой области науки. Они имеют значительный публикационный опыт по тематике диссертационного исследования и способны компетентно оценить его

теоретическую новизну и практическую значимость. Оппонент Кузнецов В.В. является специалистом в сфере информационных технологий и активно исследует их в тематике медицинских систем. За последние 5 лет им опубликовано 7 работ, непосредственно связанных с темой исследований диссертанта. Оппонент Конявский В.А. является экспертом в области информационных технологий, языковых моделей и информационной безопасности. За последние 5 лет им опубликовано 14 научных работ по теме оппонируемой работы.

Ведущая организация Федеральное государственное автономное образовательное учреждение высшего образования "Московский государственный технический университет имени Н.Э. Баумана (Национальный исследовательский университет)" (г. Москва) известен исследованиями NLP, языковых моделей и методов формализации естественного языка. Сотрудники данной организации (Винокуров А. А., Осипов А. А., Вершинин Е. В., Иванов Н. В., Косова К. А., Громоздов Д. Р., Гапанюк Ю. Е. и др.) имеют обширный научный опыт и значительное количество публикаций, относящихся к тематике диссертационного исследования.

Выбор официальных оппонентов и ведущей организации обосновывается также отсутствием совместных с соискателем проектов и печатных работ.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

разработан метод автоматического построения графовой языковой модели предметной области на основе локальных контекстов и тезаурусных связей, обеспечивающий формирование структурированного представления семантических отношений между терминами текстового корпуса,

предложены:

- метрика оценки эффективности тезауруса, позволяющая вычислить эффективность информационно-поискового тезауруса предметной области и сравнить новое полученное представление с эталонным,
- метод графового представления языковых данных в форме контекстного семантического графа, обеспечивающий явное описание типов и силы семантических взаимодействий между языковыми единицами и повышение качества рубрикации документов на величину до 26%,
- метод приведения графовой языковой модели к векторному представлению, позволяющий сохранить до 90% структурных характеристик изначальной модели,

доказаны ограничения существующих распределённых языковых моделей, связанные с неявным представлением семантических отношений между терминами,

введены интегральная метрика оценки качества информационно-поисковых тезаурусов, метод построения графовой языковой модели предметной области, включающий автоматическое формирование семантических связей, насыщение графовой структуры тезаурусными отношениями и её последующую интеграцию с векторными представлениями.

Теоретическая значимость исследования обоснована тем, что:

доказана эффективность методики формализации естественно-языковых данных с помощью графовой языковой модели в задаче рубрикации текстовых документов,

применительно к проблематике диссертации результативно (эффективно, то есть с получением обладающих новизной результатов) использованы методы теории вероятностей и математической статистики для обработки экспериментальных данных, методы экспериментальных исследований, алгоритмические методы анализа последовательностей событий,

изложены этапы создания графовой языковой модели из текстового корпуса, этапы приведения графовой языковой модели к векторному представлению,

раскрыты противоречия между традиционными методами моделирования языковых данных, учитывающими только один тип семантических связей и необходимостью структурного графового моделирования, учитывающего множество типов семантических связей и представляющего их в явном виде,

изучены методы теории алгоритмов и вычислительной сложности, включая оценку заполнения векторных и графовых языковых моделей, методы установления типов семантических связей в текстовых корпусах, методы векторного представления графовых структур включая процедуру переобозначения (релабелинга),

проведена модернизация алгоритма Вайсфейлера–Лемана для задач векторизации графовых языковых моделей посредством введения процедуры пороговой фильтрации вершин и учёта семантических характеристик графовой структуры.

Значение полученных соискателем результатов исследования для практики подтверждается тем, что:

разработаны и внедрены в учебный процесс Московского политехнического университета по направлениям подготовки 09.04.01. В учебном процессе нашли отражения следующие методики: представления локальных контекстов в виде взвешенной графовой естественно-языковой модели предметной области; Метод формализации естественно-языковых текстовых данных в виде графовых моделей с возможностью насыщения различными типами связей; алгоритм приведения языковой графовой модели к векторному представлению на основе алгоритма Вайсфейлера-Лемана, позволяющий сохранить структурные связи с множеством ребер в векторном формате, были внедрены при выполнении научного проекта на тему

«Разработка методов сбора и оценки контекстных данных на основе машинного обучения» в рамках гранта В.Е. Фортова,

определены перспективы применения разработанных методов создания и применения графовых языковых моделей на практике, метода приведения разработанной графовой модели к векторной форме на практике,

создан программный комплекс, обеспечивающий создание графовых языковых моделей на основе предложенных текстовых корпусов,

представлены методические материалы для использования в образовательных целях, методические рекомендации по внедрению разработанной методики в процессы разработки систем автоматического рубрицирования.

Оценка достоверности результатов исследования:

для экспериментальных работ использовался разработанный программный комплекс, с помощью которого проводились эксперименты на изолированной вычислительной платформе с фиксированной конфигурацией оборудования и программной среды, в результате которых продемонстрирована высокая воспроизводимость результатов при повторных запусках, а также при варьировании структуры клиентского кода; полученные данные обладают статистической устойчивостью и подтверждают достоверность выводов исследования,

теория построена на известных положениях об алгоритмической сложности, теории графов, включая известные закономерности роста вычислительной сложности, частотных свойствах естественного языка,

идея базируется на современном состоянии исследований в области сбора и использования естественно-языковых данных в языковых моделях, учете выявленных ограничений существующих методов формализации текстовых данных и понимании тенденций развития интеллектуальных систем, требующих более структурируемых и интерпретируемых языковых моделей,

использованы современные методы и подходы к моделированию текстовой информации, результаты отечественных и зарубежных исследований в области обработки естественного языка, графового представления знаний и интеллектуального анализа текстов, а также проведено сопоставление разработанных решений с существующими языковыми моделями Word2Vec и BERT и опубликованными научными результатами,

установлены количественные преимущества предложенных автором алгоритма и метода рубрикации текстов,

использованы методы планирования и проведения экспериментов, методы анализа экспериментальных данных, современные методики сбора естественно-языковых данных.

Личный вклад соискателя состоит в постановке задач исследования, разработке методов автоматического построения тезаурусов и графовых языковых моделей, создании алгоритма приведения графовой модели к векторному представлению, проведении вычислительных экспериментов, обработке и интерпретации результатов, подготовке публикаций и внедрении полученных результатов.

В ходе защиты были высказаны следующие критические замечания:

1. Вопросы математического обоснования процедуры вычисления собственных значений матриц весов графовой модели могли бы быть раскрыты более подробно, что способствовало бы более полному восприятию предложенного метода.

2. Вопросы оценки вычислительной сложности разработанных методов могли бы быть рассмотрены более подробно применительно ко всем этапам обработки комплексно, что позволило бы полнее охарактеризовать эффективность предложенного подхода при его практической реализации.

3. Несмотря на проведенные экспериментальные исследования, в работе было бы полезно более подробно исследовать масштабируемость

предложенных методов и алгоритмов при обработке текстовых корпусов большого объема, включающих документы различных предметных областей.

Соискатель Воробьев Н.Г. **ответил на заданные ему в ходе заседания вопросы и привёл собственную аргументацию**, с частью замечаний согласился.

На заседании 29 июня 2026 года **диссертационный совет принял решение** за решение научной задачи разработки методов и алгоритмов построения графовых языковых моделей и их интеграции с векторными представлениями текста, обеспечивающих повышение качества автоматической рубрикации документов и интеллектуального анализа текстовой информации, что имеет существенное значение для развития методов обработки естественного языка и информационных процессов, **присудить** Воробьеву Н.Г. **учёную степень кандидата технических наук**.

При проведении тайного голосования диссертационный совет в количестве 10 человек, из них 10 докторов наук по специальности 2.3.8. Информатика и информационные процессы, участвовавших в заседании, из 14 человек, входящих в состав совета, проголосовали: за – 10, против – нет, недействительных бюллетеней – нет.

Председатель диссертационного
совета, д.т.н., проф. Костров Борис Васильевич

Ученый секретарь диссертационного
совета, д.т.н., доц. Колесенков Александр Николаевич

29.06.2026 г.