УТВЕРЖДАЮ
Первый проректор ФГАОУ ВО
«Московский государственный
технический университет имени Н.Э.
Баумана (национальный
исследовательский университет)»
д.э.н., профессор.

_____ П.А. Дроговоз

«29» декабря 2026 г.

ОТЗЫВ

ведущей организации на диссертационную работу Воробьева Никиты Григорьевича на тему «Методы и алгоритмы рубрикации документов на основе векторно-графовой модели», представленную на соискание учёной степени кандидата технических наук по специальности 2.3.8. – Информатика и информационные процессы

Актуальность темы диссертационного исследования обусловлена ростом объёмов текстовой информации и необходимостью совершенствования методов её автоматической обработки в задачах рубрикации документов и семантического анализа. Векторные модели обладают высокой вычислительной эффективностью, однако не обеспечивают явного представления семантических связей, тогда как графовые подходы отличаются интерпретируемостью, но ограничены в масштабируемости. В связи с этим разработка гибридных графово-векторных языковых моделей представляет собой актуальную научную задачу.

Вопросы, изложенные в диссертации Воробьева Н.Г., являются актуальными и своевременными для технических наук и практического применения.

Область исследования соответствует научной специальности ВАК РФ 2.3.8. Информатика и информационные процессы, в частности п. 1 «Разработка компьютерных методов и моделей описания, оценки и оптимизации информационных процессов и ресурсов, а также средств анализа и выявления закономерностей на основе обмена информацией пользователями и возможностей используемого программно-аппаратного обеспечения» и п. 4 Разработка методов и технологий цифровой обработки аудиовизуальной информации с целью обнаружения закономерностей в данных, включая обработку текстовых и иных изображений, видеоконтента. Разработка методов и моделей распознавания, понимания и синтеза речи, принципов и методов извлечения требуемой информации из текстов».

Структура диссертации обусловлена целью, задачами и логикой проведения исследования. Диссертационная работа изложена на 178 страницах, состоит из введения, четырех глав (21 параграф), заключения, библиографического списка

(всего 93 источника), приложений (всего 3 приложения), содержит 27 рисунков, 6 диаграмм и 9 таблиц.

Во введении обосновывается актуальность темы, формулируется цель работы, состав решаемых задач, приводится перечень основных результатов, и излагается краткое содержание глав диссертации.

В первой главе выполнен анализ современных подходов к обработке естественного языка, включая методы рубрикации документов и построения диалоговых систем. Проведено сравнение экспертных и автоматически формируемых тезаурусов, а также языковых моделей Word2Vec и BERT, выявлены их преимущества и ограничения. Обоснована целесообразность использования графовых языковых моделей, основанных на графовых грамматиках, и предложен метод автоматического построения информационно-поискового тезауруса. Показано, что комбинирование графовых и векторных представлений позволяет повысить точность и интерпретируемость семантического анализа текстов.

Во второй главе разработаны и обоснованы методы количественной оценки качества тезаурусов на основе метрик мощности и связности. Проведено сравнение экспертных и автоматически построенных тезаурусов, показавшее, что экспертные структуры обладают более высокой семантической связностью и устойчивостью. Полученные результаты используются для обоснования необходимости разработки гибридных языковых моделей, сочетающих структурную точность тезаурусных подходов и масштабируемость автоматических методов.

В третьей главе разработана графовая языковая модель (GLM), основанная на графовых грамматиках и объединяющая графовые и векторные подходы к представлению лексико-семантических структур. Предложены методы формирования графов локальных контекстов, их преобразования в векторное представление и алгоритм рубрикации документов на основе структурной и спектральной близости графов. Показано, что предложенный подход позволяет учитывать топологические и контекстные свойства языка и обеспечивает высокую точность семантического анализа текстов.

В четвертой главе рассмотрены вопросы интеграции разработанных графовых моделей с современными нейросетевыми архитектурами и представлены результаты их применения в задачах рубрикации документов и выбора ответов в диалоговых системах. Показано, что использование графовых представлений позволяет повысить точность обработки текстовой информации. Также обоснована возможность практического применения предложенного подхода в интеллектуальных поисковых и диалоговых системах.

В целом можно констатировать, что работа имеет завершенный характер, соблюдены научный стиль и логика изложения материала. Каждая глава завершена соответствующими выводами.

Научная новизна Научная новизна диссертационного исследования заключается в разработке соискателем методов и алгоритмов построения графовой языковой модели для рубрицирования документов и семантического анализа текстовой информации, а также в создании подходов к интеграции графовых и векторных представлений языка в интеллектуальных системах обработки естественного языка. Наиболее существенные положения диссертационного

исследования Воробьёва Н.Г., определяющие его теоретическую и практическую значимость и обладающие научной новизной, состоят в следующем.

1. Разработана интегральная метрика оценки эффективности тезаурусных структур, основанная на совокупности топологических характеристик графа (мощность, средняя степень вершины, плотность, коэффициент кластеризации, энтропия), которая позволяет количественно сравнивать экспертные и автоматически сформированные тезаурусы и обеспечивает дифференциацию структур по нормализованному вектору признаков;

2. Предложен алгоритм автономной оценки качества тезауруса на основе интегральной метрики и весовой матрицы, обеспечивающий выявление некорректных структур при предельных значениях плотности ($D \to 0$, $D \to 1$);

3. Разработан метод построения графовой языковой модели (GLM) на основе графовых грамматик и локальных контекстов, обеспечивающий явное представление семантических связей при обработке текстовых корпусов объёмом до 1000 документов;

4. Предложен метод насыщения графовой языковой модели дополнительными тезаурусными рёбрами, повышающий структурную связность модели и улучшающий её топологические характеристики (в частности, коэффициент кластеризации и энтропию связей) по сравнению с базовой графовой структурой;

5. Разработан гибридный подход к моделированию контекста, в котором вершинами графа выступают векторные представления документов размерности 300 признаков, что обеспечивает сопоставимость с существующими векторными моделями и позволяет обрабатывать текстовые документы объёмом 3000–5000 символов (400–800 токенов) без снижения точности;

6. Предложен алгоритм приведения графовой модели к векторной форме на основе модифицированного алгоритма Вайсфейлера–Лемана с пороговой фильтрацией, позволяющий интегрировать графовые структуры в нейросетевые архитектуры и обеспечивающий сохранение не менее 90% структурной информации при преобразовании;

7. Экспериментально установлено преимущество разработанной графовой языковой модели в задачах рубрикации документов по сравнению с векторными моделями (Word2Vec, BERT), выражающееся в повышении точности рубрикации (accuracy до 0.94 против 0.78, прирост до 16%), увеличении значения F1 (macro) до 0.97 и устойчивом улучшении качества рубрикации на корпусе из 1000 документов. Обобщённый прирост точности составляет до 26%.

Можно констатировать, что полученные в диссертации результаты вносят значимый вклад в решение научно-методических вопросов построения и применения графовых языковых моделей, а также повышения эффективности рубрицирования документов и семантического анализа текстовой информации в интеллектуальных системах обработки естественного языка.

Достоверность и обоснованность научных положений, выводов и рекомендаций, сформулированных в диссертации, подтверждается апробацией

результатов исследования на научных конференциях, а также в практической и научной деятельности образовательной организации.

Соискатель корректно использует современные методы научного исследования, включая методы обработки естественного языка, теории графов, машинного обучения, статистического анализа и вычислительного моделирования, что обеспечило обоснованность и достоверность предложенных в диссертации решений, а также согласованность полученных результатов с современными исследованиями в области компьютерной лингвистики и интеллектуального анализа текстов. В работе использована теоретико-методическая и информационно-эмпирическая база, включающая фундаментальные научные труды в области обработки естественного языка, графовых грамматик, векторных языковых моделей и методов семантического анализа текстов, а также современные программные и алгоритмические средства построения интеллектуальных систем.

По теме диссертации выполнено 7 научный публикаций, из них две статьи представлены в изданиях, рекомендованных ВАК при Минобрнауки РФ.

Теоретическая и практическая значимость диссертационного исследования. Полученные соискателем результаты развивают теоретические и практические подходы к построению и применению графовых языковых моделей, предназначенных для рубрицирования документов и семантического анализа текстовой информации в интеллектуальных системах обработки естественного языка.

Теоретическая значимость диссертационной работы заключается в развитии методов графового моделирования языка и разработке гибридного подхода к построению языковых моделей, объединяющего графовые и векторные представления. Полученные результаты вносят вклад в развитие теории графовых грамматик, компьютерной лингвистики и методов семантического анализа текстов.

Практическая значимость работы заключается в возможности применения разработанных методов при создании интеллектуальных систем анализа текстов, поисковых и диалоговых систем. Предложенные решения обеспечивают повышение точности рубрикации документов и могут быть интегрированы в существующие системы обработки естественного языка и искусственного интеллекта.

Рекомендации по использованию результатов и выводов, приведенных в диссертации. Научные результаты и выводы диссертации Воробьева Н.Г. могут быть использованы:

– научно-исследовательскими организациями, ИТ-компаниями и предприятиями, занимающимися разработкой интеллектуальных систем обработки естественного языка, поисковых сервисов, рекомендательных и диалоговых систем;

– в учреждениях высшего образования при преподавании дисциплин, связанных с обработкой естественного языка, искусственным интеллектом, машинным обучением, компьютерной лингвистикой и интеллектуальным анализом данных, а также в системе дополнительного профессионального образования при подготовке и переподготовке специалистов в области информационных технологий и искусственного интеллекта.

Разработанную автором метрику эффективности тезауруса можно применять для оценки качества собираемых информационно-поисковых тезаурусов специализированных предметных областей.

Замечания по диссертационной работе. Диссертационное исследование Воробьева Н.Г. не лишено недостатков, среди которых отметим следующие.

1. В диссертационной работе подробно рассмотрены методы построения графовой языковой модели и алгоритмы рубрикации документов на основе графовых и векторных представлений текста (стр. 71–103). Вместе с тем недостаточное внимание уделено анализу вычислительной сложности предложенных алгоритмов при обработке сверхбольших текстовых корпусов и вопросам масштабируемости модели в распределённых вычислительных средах.

2. В работе подробно рассмотрены методы построения графовой языковой модели и её интеграции с векторными представлениями текста. Вместе с тем, следовало бы более подробно представить сравнительный анализ предложенного подхода с современными графовыми нейронными сетями (Graph Neural Networks), активно применяемыми в задачах обработки естественного языка.

3. В разделе, посвящённом экспериментальной апробации (стр. 94–113), представлены результаты сравнения графовой модели с Word2Vec и BERT. Вместе с тем в работе можно было бы более подробно рассмотреть влияние выбора параметров графа и пороговых коэффициентов на итоговую точность рубрикации документов.

4. Представленный метод приведения графовой языковой модели к векторному представлению обладает практической значимостью для интеграции с нейросетевыми архитектурами (стр. 77–88). Однако, на наш взгляд, следовало бы дополнительно рассмотреть возможности применения предложенного подхода совместно с современными трансформерными моделями и большими языковыми моделями (LLM).

Приведенные замечания носят частный характер, выступая в качестве возможного направления для дальнейших изысканий по данной тематике, и не снижают общей высокой оценки проделанного исследования, которое представляет собой актуальный и самостоятельный научный труд.

Заключение о соответствии критериям Положения о присуждении ученых степеней

Диссертация Воробьева Н.Г. выполнена на высоком научном уровне, обладает научной новизной, теоретической значимостью и практической ценностью. Структура диссертации представляется логичной и отражает весь процесс проведения исследования. Выводы и рекомендации достаточно обоснованы. Язык и стиль диссертации, оформление представленного текста соответствуют нормам изложения результатов научных работ.

Публикации автора достаточно полно отражают содержание работы и новые научные результаты, автореферат и диссертация удовлетворяют требованиям, предъявляемым к диссертациям на соискание ученой степени кандидата

технических наук по специальности 2.3.8. – Информатика и информационные процессы. Содержание автореферата отражает основные положения диссертации в части исследуемых задач, полученных автором новых научных результатов, оценки их теоретической значимости и практической ценности. По теме диссертационного исследования соискателем опубликованы 7 научных работ, в том числе 2 научные работы в журналах из Перечня рецензируемых научных изданий.

Диссертация Воробьева Н.Г. является самостоятельной и завершенной, научно-квалификационной работой и содержит решение научной задачи, имеющей значение для развития технических систем. Диссертация полностью соответствует всем критериям Положения о присуждении ученых степеней, утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842 (в ред. от 01 января 2025 г.), а ее автор – Воробьев Никита Григорьевич – заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.8. – Информатика и информационные процессы.

Отзыв ведущей организации по диссертации Воробьева Н.Г. обсужден и одобрен на заседании кафедры систем обработки информации и управления МГТУ им. Н.Э. Баумана 28 мая 2026 г., протокол № 10.

Доктор технических наук, доцент, заведующий кафедрой систем обработки информации и управления, федерального государственного автономного образовательного учреждения высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»

Терехов Валерий Игоревич

«28» мая 2026 г.

105005 г. Москва, 2-я Бауманская ул., д. 5, стр. 1
тел.: +7 (499) 267-54-34 эл. почта: terekchow@bmstu.ru

Сведения о ведущей организации:
Федеральное государственное автономное образовательное учреждение высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»
Адрес: 105005, г. Москва, 2-я Бауманская ул., д. 5, с. 1
Телефон: +7 (499) 261-17-43
E-mail: bauman@bmstu.ru
Официальный сайт: https://bmstu.ru/

С отзывом ознакомлен 2 июня 2026 г.