

ЗАКЛЮЧЕНИЕ ОБЪЕДИНЕННОГО СОВЕТА ПО ЗАЩИТЕ
ДИССЕРТАЦИЙ НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ КАНДИДАТА
НАУК, НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ ДОКТОРА НАУК
99.2.107.02 НА БАЗЕ ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО
АВТОНОМНОГО ОБРАЗОВАТЕЛЬНОГО УЧРЕЖДЕНИЯ ВЫСШЕГО
ОБРАЗОВАНИЯ «ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»,
ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО
ОБРАЗОВАТЕЛЬНОГО УЧРЕЖДЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
«ЮЖНО-РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ПОЛИТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ (НПИ) ИМЕНИ М.И. ПЛАТОВА», МИНИСТЕРСТВА
НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ,
ПО ДИССЕРТАЦИИ НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ
КАНДИДАТА НАУК

аттестационное дело №

решение диссертационного совета от 26.06.2025 г. № 29

О присуждении Мансур Али Махмуд, гражданину Сирийской Арабской Республики, ученой степени кандидата технических наук.

Диссертация «Модель, метод и алгоритмы Data Mining для интеллектуальной обработки и анализа текстов на естественном языке» по специальности 1.2.1. Искусственный интеллект и машинное обучение (технические науки) принята к защите 17.04.2025 г., протокол заседания № 22, объединенным диссертационным советом 99.2.107.02, созданным на базе федерального государственного автономного образовательного учреждения высшего образования «Южный федеральный университет», федерального государственного бюджетного образовательного учреждения высшего образования «Южно-Российский государственный политехнический университет (НПИ) имени М.И. Платова», Министерства науки и высшего образования Российской Федерации, 347922, пер. Некрасовский, 44, г. Таганрог, Ростовская область, Россия, приказ № 223/нк от 14.02.2023 г.

Срок полномочий совета – на период действия Номенклатуры научных специальностей.

Соискатель Мансур Али Махмуд, «26» августа 1991 года рождения, в 2019 году с отличием окончил магистратуру федерального государственного бюджетного образовательного учреждения высшего образования «Донской государственный технический университет» по направлению подготовки 09.04.02 Информационные системы и технологии, с присуждением квалификации магистра.

В период подготовки диссертации с 2019 по 2023 гг. Мансур Али Махмуд обучался (форма обучения – очная) в аспирантуре федерального государственного автономного образовательного учреждения высшего образования «Южный федеральный университет» по направлению подготовки 09.06.01 Информатика и вычислительная техника, специальность 05.25.05 Информационные системы и процессы (технические науки). После утверждения новой номенклатуры научных специальностей Мансур Али Махмуд успешно сдал кандидатский экзамен по специальности 1.2.1. Искусственный интеллект и машинное обучение (технические науки).

В 2023 году соискатель окончил аспирантуру с присвоением квалификации – Исследователь. Преподаватель-исследователь.

Работает в настоящее время программистом Института компьютерных технологий и информационной безопасности федерального государственного автономного образовательного учреждения высшего образования «Южный федеральный университет», Минобрнауки России.

Диссертация выполнена на кафедре систем автоматизированного проектирования Института компьютерных технологий и информационной безопасности федерального государственного автономного образовательного учреждения высшего образования «Южный федеральный университет», Минобрнауки России.

Научный руководитель – доктор технических наук, доцент Кравченко Юрий Алексеевич, федеральное государственное автономное образовательное учреждение высшего образования «Южный федеральный университет», Институт компьютерных технологий и информационной

безопасности, кафедра систем автоматизированного проектирования, профессор.

Официальные оппоненты:

1. Кравец Алла Григорьевна, доктор технических наук, профессор, федеральное государственное бюджетное образовательное учреждение высшего образования «Волгоградский государственный технический университет», кафедра «Системы автоматизированного проектирования и поискового конструирования», профессор;

2. Вишняков Юрий Муссович, доктор технических наук, федеральное государственное бюджетное научное учреждение «Кубанский государственный университет», кафедра вычислительных технологий, профессор;

дали положительные отзывы на диссертацию.

Ведущая организация – федеральное государственное бюджетное образовательное учреждение высшего образования «Воронежский государственный технический университет» (г. Воронеж) в своем положительном отзыве, подписанным Гусевым Павлом Юрьевичем, кандидатом технических наук, доцентом, заведующим кафедрой искусственного интеллекта и цифровых технологий и Рындиным Никитой Александровичем, доктором технических наук, доцентом, профессором кафедры искусственного интеллекта и цифровых технологий, указала, что диссертационное исследование Мансур Али Махмуд является самостоятельной, завершённой научно-квалификационной работой, в которой решена актуальная научная задача разработки моделей, методов и алгоритмов Data Mining для интеллектуальной обработки и анализа текстов на естественном языке. Решение данной научной задачи имеет принципиальное значение для развития технологий искусственного интеллекта и машинного обучения, так как позволяет существенно снизить частоту ошибок при классификации и кластеризации текстовых данных в условиях их экспоненциального роста.

Диссертационная работа Мансур Али Махмуд на тему «Модель, метод

и алгоритмы Data Mining для интеллектуальной обработки и анализа текстов на естественном языке», представленная на соискание ученой степени кандидата технических наук по специальности 1.2.1. Искусственный интеллект и машинное обучение, обладает значительной научной новизной и сочетает в себе теоретическую ценность и практическую значимость. Научные результаты исследования опубликованы в ведущих научных изданиях: рецензируемых журналах Перечня ВАК; а также в международных изданиях, индексируемых в базах данных Scopus и Web of Science. Кроме того, основные положения работы были успешно представлены и получили положительную оценку на ряде авторитетных международных и всероссийских научных конференций.

Диссертационная работа в полной мере соответствует пунктам 4 и 5 паспорта научной специальности 1.2.1. Искусственный интеллект и машинное обучение (технические науки).

Хотя в работе имеются отдельные замечания, диссертация Мансур А.М. по всем ключевым параметрам: актуальности; научной новизне; объему исследований и практической значимости – полностью отвечает требованиям пп. 9-14 Положения о присуждении ученых степеней, утвержденного Постановлением Правительства РФ от 24 сентября 2013 г. № 842 «О порядке присуждения ученых степеней» (с изменениями и дополнениями) в редакции от 16 октября 2024 г., предъявляемым ВАК РФ к диссертациям на соискание ученой степени кандидата технических наук.

Автор работы, Мансур Али Махмуд, достоин присуждения ему ученой степени кандидата технических наук по специальности 1.2.1. Искусственный интеллект и машинное обучение.

Соискатель имеет 20 опубликованных работ, в том числе по теме диссертации опубликовано 17 работ, из них в рецензируемых научных изданиях опубликовано 5 работ, из которых: 3 статьи опубликованы в издании из перечня, утвержденного ВАК РФ, рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций, в т.ч. 1 научная работа принадлежит лично автору;

2 статьи опубликованы в научных изданиях, индексируемых международными базами данных, перечень которых определен в соответствии с рекомендациями ВАК РФ; 12 статей опубликованы в других изданиях. Имеется 2 свидетельства о государственной регистрации программ для ЭВМ. Общий объем опубликованных по диссертации работ – 10.7 п.л., авторский вклад – 7,58 п.л. (авторских 70,84 %).

Полнота изложения соискателем материалов диссертации подтверждается публикацией указанных научных работ. В диссертации отсутствуют недостоверные сведения об опубликованных соискателем ученой степени работах, в которых изложены научные результаты диссертации.

Наиболее значимые работы:

1. Mansour, A. Harnessing Key Phrases in Constructing a Concept-Based Semantic Representation of Text Using Clustering Techniques / A. Mansour, J. Mohammad, Y. Kravchenko, N. Silega, D. Kravchenko // Lecture Notes in Computer Science. – LNCS. – Vol. 14335. – 2023. – P. 190-201. DOI 10.1007/978-3-031-49552-6_17. (Scopus, Q2)
2. Mansour, A. Algorithm for Optimization of Keyword Extraction Based on the Application of a Linguistic Parser / A. Mansour, J. Mohammad, D. Kravchenko, Y. Kravchenko & N. Pavlov // Informatics and Automation. – 2024. – Vol. 23. – No 2. – P. 467-494. DOI 10.15622/ia.23.2.6. (Scopus, Q4)
3. Мансур, А. М. Модифицированный метод устранения неоднозначности смысла слов, основанный на методах распределенного представления / А.М. Мансур, Ж.Х. Мохаммад, Ю.А. Кравченко // Известия ЮФУ. Технические науки. – 2021. – № 3 (220). – С. 92-101. DOI 10.18522/2311-3103-2021-3-92-101. (ВАК, K2)
4. Мансур, А. М. Векторизация текста с использованием методов интеллектуального анализа данных / Ю.А. Кравченко, А.М. Мансур, Ж.Х. Мохаммад // Известия ЮФУ. Технические науки. – 2021. – № 2 (219). – С. 154-167. DOI 10.18522/2311-3103-2021-2-154-167. (ВАК, K2)

5. Мансур А. М. Алгоритм на основе трансформеров для классификации длинных текстов /А. М. Мансур // Известия ЮФУ. Технические науки. – 2024. – №. 3. – С. 196-187. DOI 10.18522/2311-3103-2024-3-187-196. (ВАК, К2)

На диссертацию и автореферат поступили отзывы:

ведущей организации – ФГБОУ ВО «Воронежский государственный технический университет». Отзыв подписали: заведующий кафедрой искусственного интеллекта и цифровых технологий ФГБОУ ВО «ВГТУ», кандидат технических наук, доцент Гусев Павел Юрьевич и профессор кафедры искусственного интеллекта и цифровых технологий, доктор технических наук, доцент Рындин Никита Александрович, 7 замечаний;

официального оппонента – профессора кафедры «Системы автоматизированного проектирования и поискового конструирования» ФГБОУ ВО «Волгоградский государственный технический университет» (г. Волгоград), доктора технических наук, профессора Кравец Аллы Григорьевны, 5 замечаний;

официального оппонента – профессора кафедры «вычислительных технологий» ФГБОУ ВО «Кубанский государственный университет» (г. Краснодар), доктора технических наук, профессора Вишнякова Юрия Муссовича, 3 замечания;

ФГБОУ ВО «Томский государственный университет систем управления и радиоэлектроники» (г. Томск). Отзыв подписал профессор кафедры комплексной информационной безопасности электронно-вычислительных систем факультета безопасности Института системной интеграции и безопасности, доктор технических наук, профессор Ходашинский Илья Александрович, 2 замечания;

ФГБОУ ВО «Ивановский государственный энергетический университет имени В. И. Ленина» (г. Иваново). Отзыв подписал проректор по научной работе, доктор технических наук, профессор Тютиков Владимир Валентинович, 2 замечания;

ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (г. Санкт-Петербург). Отзыв подписал

руководитель лаборатории «Интегрированных систем автоматизации» главный научный сотрудник, доктор технических наук, профессор Смирнов Александр Викторович, 2 замечания;

ФГБОУ ВО «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича» (г. Санкт-Петербург). Отзыв подписал профессор кафедры «сетей связи и передачи данных (СС и ПД)», доктор технических наук, профессор Мутханна Аммар Салех Али, 2 замечания;

ФГБОУ ВО «Донской государственный технический университет» (г. Ростов-на-Дону). Отзыв подписал профессор кафедры «Информационная безопасность в вычислительных системах и сетях», доктор технических наук, профессор Шевчук Петр Сергеевич, 2 замечания;

ФГБОУ ВО «Национальный исследовательский университет «МЭИ» (г. Москва). Отзыв подписал доктор технических наук, профессор кафедры прикладной математики и искусственного интеллекта, профессор Еремеев Александр Павлович, 2 замечания.

Все **отзывы положительные**, во всех отзывах отмечено, что работа соответствует специальности 1.2.1. Искусственный интеллект и машинное обучение (технические науки).

Наиболее существенные замечания:

1. В работе автор анализирует два подхода к извлечению концептов и останавливается на применении текстового подхода, однако не проводит сравнение с методами, основанными на знаниях, а это могло бы усилить обоснованность выбора и повысить объективность результатов;

2. Хотя автором декларируется, что интерпретируемость улучшает производительность алгоритмов, отсутствует количественная оценка этого влияния;

3. Целесообразным было бы привести оценку временной трудоемкости предложенных алгоритмов при обработке больших объемов данных.

Выбор официальных оппонентов обосновывается их компетентностью в области искусственного интеллекта и машинного обучения, методов обработки и анализа текстов на естественном языке, методов интеллектуального анализа данных, а также наличием значительного количества публикаций в данной сфере исследования.

Выбор ведущей организации обосновывается широкой известностью своими научными и практическими достижениями в области методов и алгоритмов обработки и анализа текстов на естественном языке, решения задач векторизации и представления текстов, задач анализа текстов, таких как классификация текстов, кластеризация и извлечение из них значимых закономерностей, алгоритмов обработки данных, обучения и разработки методов машинного обучения, а также наличием значительного количества публикаций в данной сфере исследования.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

Разработан метод генерации векторных представлений документов на основе построенной модели векторизации, отличающийся применением интерпретируемых признаков при векторизации и позволяющий снизить частоту ошибок алгоритмов классификации и кластеризации текстовых документов;

Разработан алгоритм извлечения и фильтрации ключевых фраз на основе частоты их появления, отличающийся применением функции парсера для разметки частей речи, что позволяет извлекать ключевые фразы с правильной грамматической структурой;

Разработан алгоритм построения концептов из семантически близких фраз, отличающийся решением задачи кластеризации фраз с учетом контекстуальной семантической близости, что позволяет повысить однородность кластеров, содержащих концепты.

предложена математическая модель векторизации текстов на основе концептов, отличающаяся применением новых правил построения эталонных концептов и новых функций определения их весов и позволяющая снизить

размерность векторного пространства и улучшить дискриминационную способность результирующих векторов признаков;

доказана перспективность использования разработанных модели, метода и алгоритмов при построении малоразмерных, информативных и интерпретируемых векторных представлений текстов, что позволяет применять их для решения различных задач в системах искусственного интеллекта для обработки и анализа текстов на естественном языке;

введены измененные трактовки понятий «эталонного концепта», используемого для построения вектора текстового документа, и «интерпретируемости векторов», которое дает возможность определения значения и веса каждого элемента в векторе представления текста.

Теоретическая значимость исследования обоснована тем, что:

доказана возможность применения разработанных модели, метода и алгоритмов обработки и анализа текстов на естественном языке в системах искусственного интеллекта и машинного обучения для повышения эффективности алгоритмов классификации и кластеризации текстов, а также обнаружения скрытых закономерностей;

применительно к проблематике диссертации результативно использованы фундаментальные положения искусственного интеллекта и интеллектуального анализа данных, теорий алгоритмов и объектно-ориентированного программирования;

изложена идея построения интерпретируемых векторных представлений текстов с использованием методов обработки естественного языка и машинного обучения. Это обеспечивает возможность обнаружения ошибок, что увеличивает доверие пользователей к таким моделям анализа текстов, а также позволяет снизить частоту ошибок алгоритмов классификации и кластеризации текстовых документов;

раскрыты вопросы обработки и анализа неструктурированных текстовых данных для выявления внутренних структур, таких как паттерны в словах и фразах, присутствующих в документах, а также объединения семантически близких слов в тематические группы для формирования

концептуальных кластеров повышенной однородности и сниженной семантической неоднозначности;

изучены основные ограничения и противоречия, связанные с выбором методов и алгоритмов, применяемых для построения интерпретируемого малоразмерного векторного представления текстов с сохранением его дискриминационной способности;

проведена модернизация существующих методов векторизации текстов на основе использования алгоритмов построения эталонных концептов из семантически близких ключевых фраз вместо униграмм и лингвистического анализатора для исключения слов, не несущих смысловой нагрузки и приводящих к формированию зашумленных концептов.

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

разработаны и внедрены теоретические и практические результаты в информационные процессы ИТ-компании ООО «Ит-Эффект» (г. Москва). Полученные в работе научные результаты позволили повысить эффективность решения задач классификации, кластеризации и извлечения ключевых фраз в рекомендательной системе, реализующей технологию «Look-a-like» (поиск целевой аудитории для эффективного масштабирования деловой активности предприятия). Результаты работы также используются в учебном процессе института компьютерных технологий и информационной безопасности Южного федерального университета;

определены перспективы практического использования разработанных модели, метода и алгоритмов в составе алгоритмического обеспечения рекомендательных систем, реализующих технологию «Look-a-like»;

создано программное приложение, позволяющее использовать разработанные модель, метод и алгоритмы обработки и анализа текстов на естественном языке в системах искусственного интеллекта для снижения частоты появления ошибок в задачах классификации и кластеризации с учётом условий снижения размерности векторного представления текстов и сохранения его интерпретируемости;

представлены предложения по дальнейшему совершенствованию разработанных модели, метода и алгоритмов, включающие оптимизацию векторизации текстов для документов разного размера и применение алгоритмов построения концептов и извлечения ключевых фраз в профилировании пользователей при автоматизации построения расширенных словарей эталонных концептов за счёт внешних источников.

Оценка достоверности результатов исследования выявила:

для экспериментальных работ результаты проведенных вычислительных экспериментов согласуются с известными из литературы результатами тестовых примеров;

теоретические основы построены на достоверных данных и фактах и согласуются с опубликованными материалами по теме диссертации;

идея базируется на результатах анализа существующих методов векторизации текстов для построения интерпретируемых, низкоразмерных векторных представлений без ущерба для дискриминационной способности векторов, что позволяет обнаруживать ошибки в алгоритмах классификации и кластеризации текстов, а также выявлять и понимать их логику, тем самым повышая доверие пользователей к этим алгоритмам;

использованы результаты сравнительного анализа полученных автором решений с известными моделями и методами векторизации текстов;

установлено количественное совпадение полученных автором решений с результатами, полученными и опубликованными другими исследователями в рецензируемых научных изданиях, применительно к оценке качества методов и алгоритмов векторизации, классификации и кластеризации текстов на естественном языке;

использованы процедуры предварительной обработки текстовых данных и фильтрации ключевых фраз с использованием набора лингвистических анализаторов, которые следуют различным подходам к анализу, в дополнение к использованию статистических мер для оценки точности и эффективности разработанного метода и алгоритмов.

Личный вклад соискателя состоит в реализации всех этапов проведенных исследований. Сформулированы цель и задачи диссертации, обоснован выбор основных направлений исследований. Построена математическая модель векторизации текстов на основе применения новых правил построения эталонных концептов и новых функций определения их весов, которая позволяет снизить размерность векторного пространства и улучшить дискриминационную способность результирующих векторов признаков. Разработанный модифицированный метод генерации векторных представлений текстов на основе построенной модели векторизации позволяет снизить частоту ошибок алгоритмов классификации и кластеризации текстовых документов. Разработанный соискателем алгоритм извлечения и фильтрации ключевых фраз на основе применения функций парсера для разметки частей речи позволяет извлекать ключевые фразы с правильной грамматической структурой. Создано программное приложение, позволяющее использовать разработанные модель, метод и алгоритмы обработки и анализа текстов на естественном языке в системах искусственного интеллекта для снижения частоты появления ошибок в алгоритмах классификации и кластеризации с учётом условий снижения размерности векторного представления текстов и сохранения его интерпретируемости. Соискатель принял непосредственное участие: в получении исходных данных и проведении вычислительных экспериментов, показавших эффективность разработанных модели, метода и алгоритмов; в обработке и интерпретации экспериментальных данных; в апробации результатов исследования; в подготовке материалов для публикаций по выполненной работе.

В ходе защиты диссертации были высказаны следующие критические замечания: 1. Недостаточно обоснована эффективность характеристики интерпретируемости векторных представлений в процессе выявления ошибок алгоритмов обработки и анализа текстов на естественном языке; 2. Не в полной степени аргументирован выбор косинусной меры сходства для оценки семантической близости.

Соискатель Мансур А.М. ответил на задаваемые ему в ходе заседания вопросы и привел собственную аргументацию.

Диссертация Мансур А.М. представляет собой завершенную научно-квалификационную работу, которая соответствует критериям, установленным пп. 9-11, 13, 14 «Положения о присуждении ученых степеней», утвержденного постановлением Правительства РФ от 24.09.2013 г. № 842 (ред. от 16.10.2024 г., с изм. и доп., вступ. в силу с 01.01.2025 г.), предъявляемым ВАК РФ к диссертациям на соискание ученой степени кандидата технических наук.

На заседании 26.06.2025 г. диссертационный совет принял решение за решение научной задачи интеллектуальной обработки и анализа неструктурированных текстовых данных на естественном языке в условиях экспоненциального роста их объёма, имеющей значение для развития искусственного интеллекта и машинного обучения, присудить Мансур А.М. ученую степень кандидата технических наук.

При проведении тайного голосования диссертационный совет в количестве – 11 человек, из них – 4 доктора технических наук по научной специальности рассматриваемой диссертации, участвовавших в заседании, из – 15 человек, входящих в состав совета, проголосовали: за – 11, против – нет, недействительных бюллетеней – нет.

ПРЕДСЕДАТЕЛЬ
диссертационного совета



 Веселов Геннадий Евгеньевич

УЧЕНЫЙ СЕКРЕТАРЬ
диссертационного совета

 Кравченко Юрий Алексеевич

27.06.2025 г.