

ОТЗЫВ

на автореферат диссертации *Мансур Али Махмуд* на тему: «*Модель, метод и алгоритмы Data Mining для интеллектуальной обработки и анализа текстов на естественном языке*», представленной на соискание ученой степени *кандидата технических наук* по специальности *1.2.1 Искусственный интеллект и машинное обучение*

Тема кандидатской диссертации Мансур Али Махмуд посвящена решению актуальной научной задачи – разработке моделей, методов и алгоритмов интеллектуального анализа текстов на естественном языке с применением искусственного интеллекта (ИИ) и машинного обучения. Основной фокус работы направлен на повышение эффективности алгоритмов классификации и кластеризации текстов в ИИ-системах за счет усовершенствования методов векторного представления текстов (векторизации). Эти методы критически важны для машинного обучения, поскольку лежат в основе работы соответствующих алгоритмов.

Для построения интерпретируемых, информативных векторных представлений текстовых документов соискателем были получены следующие результаты:

1) математическая модель векторного представления текстовых документов на основе применения новых правил построения эталонных концептов и новых функций определения их весов позволяет снизить размерность векторного пространства и улучшить дискриминационную способность результирующих векторов признаков;

2) модифицированный метод генерации векторных представлений документов на основе построенной модели векторизации позволяет снизить частоту ошибок алгоритмов классификации и кластеризации документов;

3) алгоритм извлечения и фильтрации ключевых фраз на основе применения функций парсера для разметки частей речи позволяет извлекать ключевые фразы с правильной грамматической структурой;

4) алгоритм построения концептов из семантически близких фраз позволяет повысить однородность кластеров, представляющих концепты.

Все полученные автором основные результаты направлены на решение важной научной задачи разработкой модели, алгоритмов и метода обработки и анализа текстов на естественном языке в условиях экспоненциального роста объемов текстовых данных.

Практическая значимость работы заключается в создании программного приложения для проведения вычислительного эксперимента, подтверждающего преимущества полученных автором научных результатов по сравнению с существующими аналогами. Разработанное приложение также позволяет решать задачи классификации и кластеризации текстов, извлечения ключевых фраз и концептов из текстов на естественных языках. Практическая значимость также подтверждается внедрением результатов работы в коммерческой организации и в образовательной сфере, а также получением автором двух свидетельств о государственной регистрации программ для ЭВМ.

Работа прошла достаточную апробацию, результаты докладывалась и обсуждалась в рамках мероприятий ряда всероссийских и международных научно-технических конференций. Содержание диссертации отражено в 17 печатных работах, в т.ч. 3 – в изданиях, рекомендованных ВАК РФ, 2 – в изданиях из международных баз данных Scopus и/или Web of Science.

По представленному автореферату имеются следующие замечания:

- в автореферате не представлено обоснование использования косинусной меры сходства в качестве метрики для оценки семантической близости векторов;
- отсутствие рекомендаций о вычислительных ресурсах, таких как время обучения для моделей встраивания и требования к оборудованию, которые важны для практической реализации приложения;
- в тексте автореферата отсутствует ссылка на рисунок 5, этот рисунок был бы более информативен, если ось ординат задать в логарифмическом масштабе.

Данные замечания не снижают высокого уровня проведенных автором исследований и разработок. Представленный автореферат позволяет утверждать, что диссертация Мансур А.М. представляет собой законченную научно-квалификационную работу, в которой решена задача создания системы искусственного интеллекта и машинного обучения для обработки и анализа текстов на естественном языке, работа удовлетворяет требованиям пунктов 9-14 «Положения о присуждении учёных степеней», утверждённого Постановлением Правительства РФ от 24 сентября 2013 г. № 842 (в редакции от 16.10.2024 г.), установленным для кандидатских диссертаций, а её автор, Мансур Али Махмуд, заслуживает присуждения учёной степени кандидата технических наук по специальности 1.2.1 Искусственный интеллект и машинное обучение.

Профессор кафедры компьютерных систем в управлении и проектировании
факультета вычислительных систем

ФГАОУ ВО «Томский государственный университет
систем управления и радиоэлектроники»,

доктор технических наук, профессор

Ходашинский Илья Александрович

634045, г. Томск, ул. Вершинина, д. 74,

Тел.: (3822) 41-47-17 (внутр. 2095), e-mail: ilia.a.khodashinskii@tusur.ru

Персональная страница: <https://directory.tusur.ru/people/241>

Специальность: 05.13.18 - Математическое моделирование,
численные методы и комплексы программ

Я, Ходашинский Илья Александрович, даю согласие на включение моих персональных данных в документы, связанные с работой диссертационного совета, и их дальнейшую обработку.

«26» мая 2025 г.



И.А. Ходашинский

Подпись *И.А. Ходашинский*
УДОСТОВЕРЯЮ

Ученый секретарь

Е.В. Прокопьев, Е.В. Прокопьев

