

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертацию Секерина Алексея Викторовича

«Методы и алгоритмы автоматической идентификации и классификации идеологем
в русскоязычных общественно-политических текстах»

на соискание ученой степени кандидата технических наук по специальности 2.3.8 –
Информатика и информационные процессы.

Актуальность темы диссертационной работы

Современное информационное пространство характеризуется беспрецедентной интенсивностью идеологических противоборств. Ключевым инструментом концептуализации и манипуляции общественным сознанием выступает идеологема-лексико-фразеологическая единица, обладающая высоким оценочным и прагматическим потенциалом и действующая на эмоционально-подсознательном уровне. В условиях лавинообразного роста объемов текстовых данных, публикуемых в сети Интернет, ручной лингвистический анализ становится невозможным, что объективно требует разработки автоматизированных методов идентификации и классификации идеологем для противодействия деструктивным ценностным установкам и обеспечения информационно-психологической безопасности населения.

Таким образом, разработка методов и алгоритмов автоматической идентификации и классификации идеологем, позволяющих выявлять как прямые, так и контекстуальные идеологемы без необходимости формирования тематических наборов документов, представляет собой актуальную научно-практическую задачу, имеющую существенное значение для информатики, компьютерной лингвистики и обеспечения информационной безопасности.

Анализ содержания диссертационной работы

Диссертационная работа состоит из введения, трех глав, заключения, списка сокращений и условных обозначений, списка литературы и трех приложений. Общий объем диссертации составляет 148 страниц основного текста. Список литературы содержит 148 наименований.

Во введении обоснована актуальность, определены объект, предмет, цель и задачи исследования, сформулированы научная новизна, теоретическая и практическая значимость, положения, выносимые на защиту.

В первой главе проведен анализ современных подходов к автоматической идентификации и классификации идеологем. Рассмотрены различные определения понятия "идеологема", модели идейно-политического спектра, статистические и нейросетевые методы анализа политических текстов, граф-ориентированные подходы и принципы построения электронных тезаурусов. На основе анализа выявлены этапы процесса, подлежащие автоматизации, и обоснованы направления диссертационного исследования.

Во второй главе разработан общий алгоритм идентификации и классификации идеологем. Предложены: модификация нейросетевого метода преобразования текстовых последовательностей для генерации религиозно-политических ассоциаций; модифицированное структурное предсказание для разрешения кореференции; усовершенствованный метод классификации понятий на основе анализа ассоциативных обобщающих цепочек с использованием векторных представлений слов; а также модификация нейросетевых методов для выявления мнений в общественно-политических текстах.

В третьей главе представлена разработка интеллектуальной системы идентификации и классификации идеологем: спроектированы архитектура и алгоритм функционирования системы, разработаны и оценены отдельные компоненты (модели выявления именованных сущностей, генератор ассоциаций, модель разрешения кореференции, модели выявления мнений, понятийный классификатор), выполнена программная реализация и проведен анализ работы системы с привлечением экспертной оценки. Экспериментально подтверждена эффективность предложенных методов.

В заключении сформулированы основные результаты работы и выводы, обозначены перспективы дальнейших исследований.

Содержание диссертации логично структурировано, материал изложен последовательно – от обзора проблемы и теоретических основ до разработки конкретных методик и их практической апробации.

Научная новизна

Диссертационная работа отличается высокой научной новизной.

1. Автором предложен способ использования нейросетевого метода преобразования текстовых последовательностей для построения религиозно-политических ассоциаций русскоязычного дискурса, а также алгоритм, реализующий его моделью машинного обучения преобразования последовательности.

2. Усовершенствован метод классификации идеологем на основе анализа ассоциативных обобщающих цепочек и реализован алгоритм, обеспечивающий формирование ассоциативных религиозно-политических словарей с использованием векторных представлений слов.

3. Разработана интеллектуальная система идентификации и классификации идеологем, построенная на основе предложенных методов и алгоритмов, интегрирующая совокупность русскоязычных моделей машинного обучения для выявления прямых и контекстуальных идеологем и обеспечивающая определение целевых фрагментов текста без необходимости формирования тематических наборов документов.

Перечисленные решения являются оригинальными, и получены лично соискателем.

Теоретическая и практическая значимость

Теоретическая значимость работы состоит в развитии научных представлений о методах и алгоритмах автоматической обработки русскоязычных общественно-политических текстов применительно к задаче идентификации и классификации идеологем. Предложенные в диссертации методы и алгоритмы расширяют теоретическую базу в области компьютерной лингвистики и интеллектуального анализа текстов. Работа обобщает и систематизирует ранее разрозненные подходы, связывая классические методы статистического анализа и построения тезаурусов с современными нейросетевыми технологиями обработки естественного языка. Таким образом, результаты исследования обогащают научно-методический аппарат информатики и информационных процессов новыми знаниями о механизмах функционирования идеологем в русскоязычном дискурсе и о том, как эти знания могут быть использованы для автоматизации их выявления и классификации в целях обеспечения информационно-психологической безопасности.

Практическая значимость результатов диссертации является высокой. Разработанные методы и алгоритмы могут найти применение в системах мониторинга информационного пространства и обеспечения информационно-психологической безопасности. Предложенные автором подходы позволяют создавать инструменты для автоматического выявления идеологем в общественно-политических текстах, в том числе

контекстуальных, что актуально для противодействия распространению деструктивных ценностных установок и экстремистских материалов. Методы классификации идеологем и выявления авторской позиции могут использоваться при проведении лингвистических экспертиз, мониторинге СМИ и социальных сетей, а также в исследовательских проектах по изучению общественного мнения.

Практическая ценность подтверждается созданием программного обеспечения (получены 3 свидетельства о регистрации программ для ЭВМ и 1 свидетельство о регистрации базы данных), внедренного в деятельность трех организаций: АНО «Институт развития цифровых и социокультурных компетенций», КРОО «Объединенный центр «МОНОЛИТ», АНО «Ресурсный центр добровольчества Курской области». Разработанная система обеспечивает обработку текста со скоростью, превышающей эксперта-лингвиста в 6,96 раз, и достигает F1-меры идентификации идеологем 76,4%, что подтверждает ее практическую пригодность для решения задач мониторинга информационного пространства.

Обоснованность и достоверность результатов

Полученные в диссертации научные результаты и выводы являются обоснованными и достоверными. Это обеспечено корректным применением современных методов исследований – математической статистики, машинного обучения, анализа графов и вычислительных экспериментов при решении поставленных задач. Автор провел обширные экспериментальные исследования на репрезентативных русскоязычных корпусах текстовых данных (включая наборы ru_news, NEREL, RuCoCo, CABSAR, RuSentNE-2023), что позволило проверить работоспособность предложенных методов. Использование метрик оценки качества (METEOR, WER, seqeval-F1, LEA, F1-macro/micro) и сравнение с базовыми подходами подтверждают правильность сделанных выводов. Достоверность результатов также подтверждена согласованностью экспертных оценок (рассчитаны коэффициенты ICC(1) и AC2 Гвета). В диссертации прослеживается согласованность полученных результатов с выдвинутыми гипотезами и целями исследования.

Полнота опубликования результатов работы

Основные результаты диссертационного исследования опубликованы в достаточном объеме, что соответствует установленным требованиям ВАК: по теме диссертации автором опубликовано 13 печатных работ, в том числе 3 статьи в рецензируемых научных изданиях из перечня ВАК РФ, 6 работ в материалах всероссийских и международных конференций, получено 3 свидетельства о государственной регистрации программы для ЭВМ и 1 свидетельство о государственной регистрации базы данных; результаты докладывались на 7 научно-технических конференциях (в том числе 2 международных), что в полной мере отражает основное содержание диссертационного исследования.

Соответствие автореферата содержанию диссертации

Автореферат диссертации полностью соответствует ее содержанию. В автореферате отражены все основные положения, выводы и результаты исследования, представленные в полном тексте диссертации. Материалы автореферата структурированы аналогично диссертации, что позволяет получить корректное представление о цели, задачах, методах и результатах работы.

Замечания по диссертационной работе

Вместе с тем по диссертационной работе имеются следующие замечания:

1. Для классификации понятий используется fastText (2020 г.), который автор сам называет «устаревшим» (стр. 84). При этом в работе нет экспериментального сравнения с

контекстуальными эмбедингами (BERT, RoBERTa). Без такого сравнения утверждение о «достаточности» статической модели не доказано.

2. Все метрики (METEOR, F1, LEA, precision) приведены как точечные оценки. Доверительные интервалы не рассчитаны, что не позволяет судить о статистической значимости различий между моделями и стабильности результатов.

3. Неясно, как разработанный алгоритм отличает идеологему от смежных понятий (политическая лексема, идеологический концепт), и существуют ли случаи, когда лексема в одном контексте является идеологемой, а в другом - нейтральным термином, и как это обрабатывается?

4. Исходный набор *NEREL* содержит лишь 524 упоминания идеологических категорий. Автоматическая разметка 4 млн новостей на такой основе может приводить к накоплению систематической ошибки. Неясно, как это контролировалось.

5. Система представляет собой последовательную или параллельно-последовательную цепочку из 6 моделей (рисунки 3.13–3.21). Ошибка на раннем этапе может искажать результаты всех последующих этапов. Исследовалось ли влияние качества отдельных компонентов на итоговую F-меру идентификации (76,4%), и предусмотрены ли механизмы обнаружения и компенсации ошибок на ранних этапах?

6. Для проверки качества понятийного классификатора привлекалось 10 респондентов, оценивавших 100 слов ($ICC(1) = 0,436$). Для задач NLP в международной практике приняты более масштабные эксперименты. Не повышает ли малый объем выборки риск непреднамеренной предвзятости респондентов, и рассматривалась ли возможность масштабирования экспертной оценки?

7. Полученная скорость обработки 230 слов в минуту остается недостаточной для мониторинга СМИ в реальном времени. Основным «узким горлышком», как следует из анализа архитектуры, выступает этап генеративного разрешения кореференции на основе модели FRED-T5-1.7B. Рассматривались ли способы ускорения системы (дистилляция модели, замена на менее ресурсоемкую архитектуру, кэширование результатов для часто встречающихся конструкций, использование эффективных методов внимания), и каковы их перспективы?

8. В работе представлены количественные оценки скорости обработки текстов (230 слов/мин) и объема потребляемой оперативной памяти (20 Гб). Однако в тексте диссертации отсутствует информация о конфигурации вычислительного стенда (процессоре, количестве ядер, объеме оперативной памяти, модели GPU) на котором проводились эксперименты.

9. Фактически, автором создана универсальная технологическая платформа для выявления и структурирования скрытых смысловых связей в тексте. Однако автор ограничивается политической сферой, не рассматривая другие возможные области использования. В каких предметных областях предлагаемые методы могут найти применение и проводились ли соответствующие эксперименты?

10. Значение 762,05 для CNN-NER (таблица 3.2) выглядит аномальным. При стандартной многоклассовой кросс-энтропии такая величина потерь означает, что модель систематически с высокой степенью уверенности предсказывала неверный класс. Однако метрики Precision и Recall при этом указаны на уровне 98%, что не согласуется с интерпретацией значения функции потерь как меры ошибки классификации. Вопрос - каким образом рассчитывался loss для CNN-NER и по какой причине он

приводится в одной таблице с loss трансформерной модели без оговорок о несопоставимости?

Отмеченные недостатки носят частный характер и не снижают общей положительной оценки работы.

Заключение

Диссертационная работа Секерина Алексея Викторовича является завершенным научно-квалификационным исследованием, выполненным на высоком теоретическом и методологическом уровне. В работе решена актуальная научно-практическая задача - автоматизация процесса идентификации и классификации идеологем в русскоязычных общественно-политических текстах с применением нейросетевых технологий.

Автором лично получены следующие результаты:

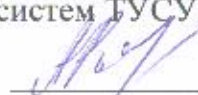
- предложен и экспериментально обоснован способ использования нейросетевого метода преобразования последовательностей для генерации религиозно-политических ассоциаций;
- усовершенствован метод классификации идеологем на основе ассоциативных обобщающих цепочек;
- разработан многоэтапный алгоритм и интеллектуальная система, интегрирующая совокупность моделей машинного обучения.

По объему исследований, научной новизне, обоснованности выводов и практической ценности работа полностью удовлетворяет требованиям п. 9–14 «Положения о присуждении ученых степеней», утвержденного Постановлением Правительства Российской Федерации № 842 от 24 сентября 2013 г., а ее автор, Секерин Алексей Викторович, заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.8. Информатика и информационные процессы.

Я, Романов Александр Сергеевич, даю свое согласие на включение своих персональных данных в документы, связанные с работой диссертационного совета, и их дальнейшую обработку.

Официальный оппонент,
доктор технических наук, доцент, доцент каф. комплексной информационной безопасности электронно-вычислительных систем ТУСУРа

«07» сентября 2026

 Романов Александр Сергеевич

Диссертация на соискание ученой степени доктора технических наук защищена по специальности 2.3.6 – Методы и системы защиты информации, информационная безопасность (2025 г.).

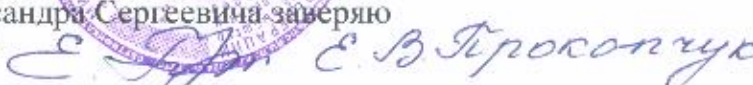
ФГАОУ ВО «Томский государственный университет систем управления и радиоэлектроники»

634050, г. Томск, пр. Ленина, 40

Тел: +7 (952) 895-25-52

Email: alexx.romanov@gmail.com

Подпись Романова Александра Сергеевича заверяю


Ученый секретарь  Е. В. Тирокопчук