

ОТЗЫВ

**Официального оппонента д.т.н. Н. В. Лукашевич
на диссертационную работу Чистовой Елены Викторовны на тему
«Методы анализа риторической структуры текстов на русском языке»,
представленную на соискание учёной степени
кандидата технических наук по специальности
1.2.1 – Искусственный интеллект и машинное обучение**

Актуальность темы исследования обусловлена ростом объёмов текстовой информации и необходимостью расширения лексико-семантического анализа текстов риторическим анализом, моделирующим дискурсивную организацию текста, для решения практических задач понимания связного текста. Большинство существующих систем риторического анализа ориентировано на английский язык; для русского языка до настоящего времени автоматических средств выделения риторических структур не существовало. Актуальной является также проблема использования ограниченных ресурсов экспертной лингвистической разметки. Корпуса с риторической разметкой малы, неоднородны по жанрам и используют разные варианты формализации Теории риторических структур, что затрудняет построение универсальных моделей и применение методов глубокого обучения. В диссертационном исследовании предложены решения описанных проблем, в том числе разработаны методы риторического анализа для русского языка, а также кросс-языковые и кросс-жанровые методы риторического анализа, что соответствует ключевым тенденциям развития искусственного интеллекта и обработки естественного языка.

Научная новизна заключается в разработке комплекса методов автоматического риторического анализа текстов на русском языке, включающего методы выделения локальных и полнотекстовых дискурсивных структур; до настоящего времени подобные методы для русского языка отсутствовали. Предложены методы машинного обучения анализаторов на объединённых корпусах, использующих различные варианты интерпретации теории риторических структур, что обеспечивает кросс-жанровую и кросс-языковую универсальность разрабатываемых моделей. Предложены и экспериментально обоснованы методы решения прикладных задач (классификации текстов, разрешения кореференции и извлечения структур аргументации), использующих информацию о дискурсивной организации текста; продемонстрировано улучшение качества относительно базовых моделей на основе современных языковых моделей. В совокупности полу-

ченные результаты развивают теорию и практику дискурсивного анализа и расширяют возможности разработки методов искусственного интеллекта для русского языка.

Достоверность и обоснованность научных результатов подтверждается корректно примененным математическим аппаратом, лежащим в основе разработанных методов обработки текстов. Экспериментальные исследования проведены с использованием открытых общепризнанных корпусов экспертной разметки, а также с применением строго определённых метрик, что позволяет сопоставлять полученные результаты с другими работами в этой области. Предложенные методы реализованы в виде открытого программного обеспечения, что дополнительно подтверждает воспроизводимость результатов. Основные положения диссертации обсуждались на международных конференциях серий ACL, AIST и Диалог, а также изложены в публикациях в журналах из перечня ВАК; получено свидетельство о регистрации программы для ЭВМ.

Теоретическая значимость работы состоит в развитии методов риторического анализа и дискурсивного моделирования для русского языка. В работе предложены гибридные архитектуры риторического анализа, использующие преимущества как интерпретируемых лингвистических признаков и обучаемых векторных представлений. Впервые исследованы возможности переноса моделей полного риторического анализа между языками и жанрами, в том числе при наличии существенно отличающихся схем разметки. Продемонстрировано влияние использования дискурсивного анализа на решение прикладных задач; уточнено представление о роли риторической структуры в моделировании смысла текста. Полученные модели и экспериментальные выводы могут служить основой для дальнейших исследований по универсализации риторического анализа, использованию дискурсивной информации в больших языковых моделях и разработке интерпретируемых систем анализа текста.

Практическая значимость состоит в разработке программных средств анализа текста, реализующих методы риторического анализа текстов на русском языке. Показано, что учёт риторической структуры позволяет повысить качество автоматической классификации текстов в задачах анализа тональности и аргументации по сравнению с классическими моделями на основе кодирующих языковых моделей. Указанное программное обеспечение используется в составе средств лингво-статистического и психолингвистического анализа текстов, что

подтверждено документально. Разработанные методы и модели доступны для свободного использования другими исследователями.

Содержание работы

Диссертация состоит из введения, пяти глав, заключения, библиографического списка и приложений.

Во введении обосновываются актуальность тематики, формулируются цель и задачи исследования, перечисляются основные положения, выносимые на защиту, описываются научная новизна и практическая значимость работы.

В первой главе представлен обзор теоретических основ дискурсивного анализа и Теории риторических структур, рассмотрены модели представления дискурса, методы автоматического анализа, корпуса риторической разметки для разных языков. Сформулированы актуальные проблемы: отсутствие средств риторического анализа для русского языка, низкая согласованность разметки и трудности кросс-языковой обобщаемости автоматического риторического анализа.

Вторая глава посвящена разработке и экспериментальному исследованию методов классификации риторических отношений и нуклеарности в русскоязычном письменном дискурсе. Подробно описаны используемые признаки дискурсивных единиц, приведено сопоставление различных алгоритмов машинного обучения, предложены ансамбли моделей, проведён анализ влияния признаков и объёма обучающих данных на качество классификации.

Третья глава содержит описание методов построения поверхностных риторических структур в текстах на русском языке. Предложена система, включающая дискурсивный сегментатор на основе BiLSTM-CRF, классификаторы структуры и типа риторических отношений с ансамблированием методов классического машинного и глубокого обучения. Система реализует алгоритмы жадного восходящего разбора структуры и нисходящего разбора с лучевым поиском. Показано, что включение локального нисходящего разбора абзацев приводит к улучшению качества риторического анализа на уровне предложений, абзацев и документа.

В четвёртой главе описан усовершенствованный метод полнотекстового нисходящего риторического анализа на базе модели DMRST. Приведены описание архитектуры анализатора на основе глубокого обучения и метода динамического взвешивания потерь. Описано создание параллельного корпуса RRG, приведены оценки кросс-языковой обобщаемости риторического анализа. Предложен метод обучения на смешанных корпусах, позволяющий получать жанрово-универсальные модели.

Пятая глава посвящена применению риторического анализа в прикладных задачах. Предложена двухэтапная схема классификации текстов с агрегированием оценок по риторическому дереву; приведен анализ влияния дискурсивных признаков на качество разрешения кореференции; предложен анализатор структуры аргументации поверх риторической структуры. Экспериментально показано улучшение качества по сравнению с базовыми методами в каждой из рассмотренных задач.

В заключении приведены основные результаты работы.

Замечания

По диссертационной работе имеются следующие замечания:

1. В главе 5 рассматривается применение риторических структур для решения прикладных задач, а именно: извлечение аргументов, извлечение позиции автора по некоторой проблеме и разрешение референции. Показано, что добавление риторических структур к базовому классификатору улучшает результаты. Однако не приводятся и не обсуждаются лучшие результаты, которые достигаются на этих же данных альтернативными методами.

Указанное замечание не ставит под сомнение ценность основных результатов работы.

Общее заключение

Диссертационная работа Е.В. Чисовой «Методы анализа риторической структуры текстов на русском языке» представляет собой завершённое научно-квалификационное исследование, в котором решены актуальные задачи, объединённые общим подходом. Полученные результаты обладают научной новизной, теоретической и практической значимостью, подтверждены экспериментальными исследованиями и внедрением программного обеспечения.

Содержание работы Е.В. Чисовой «Методы анализа риторической структуры текстов на русском языке» соответствует паспорту специальности 1.2.1 «Искусственный интеллект и машинное обучение». Работа соответствует требованиям ВАК Минобрнауки России, предъявляемым к диссертациям на соискание степени кандидата наук, а её автор заслуживает присуждения искомой учёной степени кандидата технических наук по указанной специальности.

Официальный оппонент
доктор технических наук,
ведущий научный сотрудник лаборатории
анализа информационных ресурсов
Научно-исследовательского
вычислительного центра МГУ
имени М.В. Ломоносова
Адрес: 119899, Москва, Ленинские горы, 1, стр. 4
27.01.2026 г.

 / Н. В. Лукашевич

Подпись Лукашевич Н.В. заверяю



Директор НИВЦ МГУ,
Член-кор. РАН В.В.Воеводин