

Министерство науки и высшего образования
Российской Федерации

Федеральное государственное автономное
образовательное учреждение
высшего образования

«Московский физико-технический институт
(национальный исследовательский университет)»
(МФТИ)

Юридический адрес: 117303, г. Москва,
ул. Керченская, дом 1А, корпус 1

Почтовый адрес: 141700, Московская обл.,
г. Долгопрудный, Институтский переулок, дом 9
Тел.: +7 (495) 408-57-00, факс: +7 (495) 408-68-69
info@mipt.ru

УТВЕРЖДАЮ

Проректор по научной работе



Баган Виталий
Анатольевич

2026 г.

на № 28.01.2026 № 26-17-560
от _____

ОТЗЫВ

ведущей организации на диссертацию **Чистовой Елены Викторовны**
«Методы анализа риторической структуры текстов на русском языке»,
представленную на соискание ученой степени кандидата технических наук
по специальности 1.2.1. Искусственный интеллект и машинное обучение

Актуальность темы диссертации

Диссертационная работа посвящена современным методам анализа риторической структуры текста, разработке методов анализа риторической структуры для русского языка и использованию риторических структур в задачах обработки текстов. Актуальность темы исследования обусловлена необходимостью разработки моделей, способных учитывать не только поверхностное содержание текста, но и способ его организации: связность, логику аргументации и коммуникативную функцию отдельных высказываний.

Предложенные в работе методы полнотекстового анализа риторической структуры, основанные на алгоритме нисходящего разбора, позволяют достичь высокого качества разбора дискурсивных структур для русского языка, а также демонстрируют надёжность на гетерогенных кросс-языковых и кросс-жанровых данных. Работа также формирует научно обоснованный подход к интеграции дискурсивной информации в современные модели

современные модели обработки естественного языка. Показано, что реализация предложенных методов риторического анализа открывает возможности для улучшения качества решения задач классификации, извлечения структур аргументации и разрешения кореференции.

Таким образом, результаты, представленные в диссертации, формируют научно и практически значимый вклад в развитие инструментов обработки естественного языка. Работа имеет высокую актуальность как для научных исследований в области компьютерной лингвистики, так и для практических приложений, требующих глубокой интерпретации текстовой структуры.

Структура и содержание диссертации

Диссертация состоит из введения, пяти глав, заключения, списка литературы и приложений. Общий объём работы составляет 217 страниц, включая 25 рисунков и 36 таблиц. Список литературы содержит 201 наименование.

Во **введении** обоснованы актуальность темы исследования, научная новизна и практическая значимость результатов, полученных в работе, представлены выносимые на защиту научные положения. **Первая глава** посвящена теоретическим основам анализа дискурса. В ней приведены детальный обзор теории риторических структур, характеристика риторических отношений, методы их формального описания, а также существующие корпуса разметки. Особое внимание уделено специфике русскоязычного дискурса и проблемам кросс-языкового и кросс-жанрового переноса. Во **второй главе** исследуются риторические отношения в русскоязычном письменном дискурсе. Представлено признаковое описание дискурсивных единиц, рассмотрены методы классификации риторических отношений на основе машинного обучения и приведены результаты экспериментального анализа на корпусе русскоязычных текстов. В **третьей главе** предложены методы поверхностного разбора риторической структуры текста на русском языке. Описаны подходы к автоматической дискурсивной сегментации, представлены модели классификации риторических отношений, включая методы глубокого обучения и ансамблевые методы, приведены разработанные алгоритмы построения риторического леса. Описаны результаты оценок качества решения подзадач и полного риторического анализа текста. В **четвёртой главе** предложены методы полнотекстового разбора риторической структуры. Описан гибридный

нисходящий метод анализа, рассмотрены подходы к кросс-языковой адаптации риторического анализа и методы использования разнородной риторической разметки для повышения кросс-жанровой обобщаемости. Представлены результаты экспериментов на параллельных и смешанных корпусах риторической разметки для русского и английского языков. **Пятая глава** посвящена приложениям анализа риторической структуры в практических задачах обработки естественного языка. Представлены подходы к использованию дискурсивной информации в современных методах классификации текстов, разрешения кореференции и извлечения структуры аргументации; проведены соответствующие экспериментальные исследования и сопоставление с существующими решениями. В **заключении** обобщены результаты исследования, сформулированы основные выводы и обозначены направления дальнейшего развития методов риторического анализа.

Основные результаты диссертации и их научная новизна

- Реализован полный набор методов риторического анализа русскоязычных текстов от выделения элементарных дискурсивных единиц до построения полнотекстового риторического дерева. Впервые для русского языка разработаны методы анализа локальных и полнотекстовых риторических структур.
- Предложен оригинальный метод риторического анализа посредством глубокого обучения на разнородных данных, включающих корпуса с различными наборами риторических отношений и жанровой принадлежностью.
- Проведено исследование кросс-языковой обобщаемости анализа риторических структур на большом параллельном русско-английском корпусе.
- Разработан и экспериментально исследован метод классификации текстов с учётом риторической структуры, обеспечивающий повышение качества в задачах классификации тональности и аргументации.
- Предложен метод разрешения кореференции в русскоязычных текстах, основанный на использовании метрик расстояния, учитывающих ядерность высказываний в риторическом дереве.
- Разработан оригинальный метод извлечения из текста структур аргументации поверх риторической структуры. Установлено, что

использование нескольких вариантов риторического разбора позволяет улучшить качество решения задачи извлечения структуры аргументации.

Соответствие паспорту научной специальности

Область исследования и содержание диссертации соответствуют паспорту специальности 1.2.1 «Искусственный интеллект и машинное обучение» в следующих пунктах:

Пункт 4: «Разработка методов, алгоритмов и создание систем искусственного интеллекта и машинного обучения для обработки и анализа текстов на естественном языке, для изображений, речи, биомедицины и других специальных видов данных».

Пункт 5 в части «Методы и технологии поиска, приобретения и использования знаний и закономерностей, в том числе – эмпирических, в системах искусственного интеллекта».

Пункт 7 в части «Разработка специализированного математического, алгоритмического и программного обеспечения систем искусственного интеллекта и машинного обучения».

Достоверность полученных результатов

Достоверность результатов диссертационной работы обеспечивается корректностью постановки задач, обоснованным выбором методов исследования и применением современных алгоритмов машинного обучения и глубокого обучения, адаптированных для задач риторического анализа. Использованные модели обучались и оценивались на корпусах риторической разметки, соответствующих действующим требованиям к качеству разметки и принятым в международной практике. Результаты подтверждаются воспроизводимостью вычислительных экспериментов, проведённых на независимых выборках и включающих стандартные метрики качества дискурсивного анализа. Дополнительная проверка достоверности обеспечена применением предложенных методов в прикладных задачах классификации текстов, разрешения кореференции и анализа аргументации, где полученные результаты согласуются с теоретическими ожиданиями и подтверждаются экспериментально. Внедрение разработанного программного обеспечения в действующие аналитические системы также подтверждает корректность и надежность разработанных методов. Полученные результаты представлены научному сообществу в виде публикаций и докладов на ведущих

международных конференциях по компьютерной лингвистике, что свидетельствует о независимой экспертной оценке и признании научной новизны работы.

Значимость полученных результатов

Научные результаты, полученные в диссертационной работе, имеют важное значение для дискурсивного анализа текстов, компьютерной лингвистики и разработки интеллектуальных систем. Созданные методы формируют основу для дальнейших исследований риторических структур русского языка, для которого до настоящего времени не существовало инструментов полнотекстового дискурсивного анализа.

Теоретическая значимость диссертационной работы состоит в развитии методологии дискурсивного анализа в рамках Теории риторических структур применительно к русскоязычному тексту. Теоретической значимостью также обладает разработка методологии совместного обучения на гетерогенных и неполностью совместимых корпусах риторической разметки, позволяющей применять теорию риторических структур к широкому спектру текстов и решающей актуальную задачу использования разнородных данных в нейросетевом моделировании.

Практическая значимость диссертационной работы заключается в создании набора методов и программных компонентов, предназначенных для интеграции в реальные системы обработки текстовой информации. Разработан открытый программный комплекс риторического анализа текстов для русского языка, который может использоваться как автономный инструмент, либо как модуль в составе прикладных систем обработки текстов. Полученные результаты доказали практическую важность риторической структуры как источника признаков для решения задач классификации, разрешения кореференции и извлечения аргументации, что расширяет возможности создания интерпретируемых и надёжных систем искусственного интеллекта в научных и прикладных областях. Внедрение разработанного программного комплекса в исследовательские и прикладные системы подчёркивает актуальность и востребованность научных результатов.

Рекомендации по использованию результатов

Результаты работы могут быть использованы при разработке интеллектуальных систем обработки текстовой информации, в которых

требуется реконструкция структуры рассуждения, выделение ключевых фрагментов и интерпретация логических связей в документах; в системах автоматического извлечения аргументов, тезисов и связей, применяемых в юридическом анализе и общественно-политической аналитике; в решениях, связанных с построением баз знаний и онтологий на основе связанных документов, где важна корректная обработка сущностей и событий; в образовательных и экспертных системах, ориентированных на анализ аргументации и оценку письменных работ, включая автоматизированные системы проверки изложения, эссе и научных текстов.

Методы, разработанные в диссертации, могут использоваться для повышения интерпретируемости моделей искусственного интеллекта и улучшения качества решений в задачах, требующих глубокого анализа структуры текста. Разработанное программное обеспечение может служить ядром для построения новых модулей анализа текста, расширяя функциональные возможности существующих языковых моделей.

Замечания

В тексте Введения, а именно в формулировке цели, задач, научной новизны и положениях, выносимых на защиту используются формулировки, такие как "Методы риторического дискурсивного анализа", "метод построения структур" и другие. Данные формулировки могут быть также применимы к лингвистическим задачам, выполняемым вручную. Рекомендуется уточнять, что рассматривается именно автоматическое выполнение данных задач с помощью методов машинного обучения.

В некоторых таблицах Главы 3 не приведено значение стандартного отклонения по нескольким запускам эксперимента (Таблицы 9, 10, 11). Также не указано явно, что представляет собой метод "baseline".

В тексте автореферата упоминается задача разработки "универсального риторического анализатора, т.е. построение универсальных систем риторического анализа не только для множества языков, но и для множества жанров внутри одного языка". В какой мере данная задача была выполнена в рамках данной диссертации и какие существуют недостающие этапы решения данной задачи?

Опишите область применимости и перспективы развития узкоспециализированных обучаемых систем для риторического дискурсивного анализа по сравнению с большими языковыми моделями, в

том числе применимость последних для разметки и генерации обучающих данных.

Заключительная оценка

Результаты диссертации изложены в двух публикациях в изданиях, рекомендованных ВАК. Шесть публикаций проиндексированы в международной наукометрической базе Scopus. Результаты представлены на ведущих международных конференциях по компьютерной лингвистике. Опубликованные работы полно и адекватно отражают содержание диссертации, в тексте диссертации приведены корректные ссылки на использованные источники и соблюдены требования к оформлению научно-квалификационной работы.

Диссертационная работа Чистовой Е. В. «Методы анализа риторической структуры текстов на русском языке» является научно-квалификационной работой, в которой изложены новые научно обоснованные технические и технологические решения и разработки, обеспечивающие автоматический риторический анализ текстов множества жанров на русском языке. Диссертация соответствует требованиям к кандидатским диссертациям, установленным «Положением о присуждении ученых степеней», утвержденным Постановлением Правительства РФ от 24.09.2013 №842, а её автор, Чистова Елена Викторовна, заслуживает присуждения учёной степени кандидата технических наук по специальности 1.2.1 «Искусственный интеллект и машинное обучение».

Диссертационная работа и отзыв рассмотрены и обсуждены на научном семинаре Лаборатории нейронных систем и глубокого обучения МФТИ, протокол №1 от 03 декабря 2025 года.

Кандидат физ.-мат. наук  Куратов Юрий Михайлович

Почтовый адрес: 141700, Московская область, г. Долгопрудный, Институтский пер.,9

Телефон: +7 911 502 9223,

Адрес электронной почты: yurii.kuratov@phystech.edu

Организация – место работы: федеральное государственное автономное образовательное учреждение высшего образования «Московский физико-технический институт (национальный исследовательский университет)», Лаборатория нейронных систем и глубокого обучения

Должность: Старший научный сотрудник

Web-сайт организации: <https://mipt.ru>