

ОТЗЫВ

Официального оппонента к.т.н. Д. А. Ильвовского
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

Содержание работы

Диссертация включает введение, пять глав, заключение, список литературы и два приложения.

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

Первая глава носит обзорный характер и посвящена дискурсу как объекту моделирования, теории риторических структур и существующим подходам к автоматическому риторическому анализу. Отдельно выделены актуальные проблемы низкой согласованности разметки корпусов и трудности кросс-языкового переноса моделей.

Во второй главе предложены методы классификации риторических отношений и ядерности дискурсивных единиц в русскоязычных текстах. Описано построение признакового пространства, рассмотрены различные алгоритмы машинного обучения и их ансамбли, приведены результаты экспериментов на корпусе риторической разметки русскоязычного письменного дискурса RRT и описаны выявленные закономерности.

Третья глава описывает архитектуру поверхностного риторического анализатора для русского языка, включающего модель дискурсивной сегментации, структурный классификатор и классификатор типа отношения и ядерности. Показаны преимущества ансамблирования моделей с интерпретируемыми признаками и языковых моделей при решении задач дискурсивной классификации. Предложены стратегии построения так называемого леса риторических структур и их комбинация с нисходящим разбором абзацев на основе нейронной модели.

В четвёртой главе подробно рассматривается метод полнотекстового нисходящего анализа, его модификации и применение для кросс-языкового и кросс-жанрового риторического анализа. Описано создание параллельного корпуса RRG, обучение двуязычных моделей. Приведены кросс-языковой и

кросс-жанровой обобщаемости, разработана модель риторического анализа для русского языка на объединении непосредственно несовместимых корпусов RRT и RRG.

Пятая глава посвящена использованию риторических структур в прикладных задачах. Предложен метод, улучшающий классификацию дискурсивно сложных текстов за счет использования риторического анализа. Исследовано влияние риторической структуры на качество разрешения кореференции. Разработаны методы построения структуры аргументации в тексте «поверх» риторической структуры.

В заключении описаны основные результаты исследования.

Актуальность темы исследования обусловлена востребованностью технологий, позволяющих учитывать не только локальные свойства текста, но и его дискурсивную организацию. Для русского языка до настоящего времени отсутствовали автоматические методы риторического анализа, что существенно ограничивало применение современных методов анализа связных документов. Работа Елены Чистовой направлена на устранение этого пробела – она предлагает методы риторического анализа текстов на русском языке, включая кросс-языковые и кросс-жанровые подходы, и развивает как теоретические, так и прикладные аспекты компьютерной лингвистики.

Научная новизна работы заключается в разработке методов автоматического риторического анализа для русскоязычных текстов, включающих модели дискурсивной сегментации, классификации риторических отношений и построения иерархических структур. Предложены подходы к поверхностному разбору риторической структуры, сочетающие классические методы машинного обучения и нейросетевые модели. Сформирован и описан первый крупный параллельный русско-английский корпус риторической разметки RRG, позволяющий проводить исследования кросс-языковой обобщаемости методов риторического анализа и обучать двуязычные модели. Предложены методы, которые позволяют обучать полнотекстовые анализаторы на смешанных данных из нескольких формально несовместимых корпусов риторической разметки. Эти методы обеспечивают жанровую и стилистическую обобщаемость моделей при сохранении различий в интерпретации риторических отношений. Разработаны и исследованы методы применения риторической структуры в задачах классификации текстов, разрешения кореференции и построения структур аргументации.

Достоверность и обоснованность научных результатов подтверждена корректно примененным математическим аппаратом, лежащим в основе методов

машинного обучения, использованием общепринятых метрик качества и стандартных наборов данных для исследуемых задач. Корректность выводов подтверждается сопоставлением полученных оценок с результатами других авторов, в том числе в рамках международных бенчмарков, а также апробацией работы на российских и международных лингвистических конференциях самого высокого уровня (Диалог, ACL и другие). Содержание диссертационной работы изложено в 2 работах, опубликованных в изданиях ВАК, в 6 публикациях в изданиях, индексируемых в Scopus, зарегистрирована 1 программа для ЭВМ.

Теоретическая значимость диссертации заключается в развитии и углублении методологии дискурсивного анализа в рамках Теории риторических структур. Еленой предложены интерпретируемые признаки дискурсивных единиц, отражающие связи между синтаксисом, семантикой и риторической организацией текстов на русском языке. Разработаны и обоснованы подходы к кросс-языковой и кросс-жанровой унификации автоматического риторического анализа без примитивизации экспертной разметки. Продемонстрировано, каким образом риторическая структура может выступать промежуточным уровнем представления текста при решении прикладных задач: от разрешения кореференции до анализа аргументации. Таким образом, диссертация вносит существенный вклад в теорию дискурсивного анализа и в методологию построения многоуровневых представлений текста в системах искусственного интеллекта.

Практическая значимость работы выражается в создании совокупности методов и программных средств, применимых в реальных системах обработки текстовой информации. Разработан и опубликован открытый программный комплекс риторического анализа для русского языка, включающий сегментатор, модули классификации отношений и ядерности, а также полнотекстовый анализатор структуры. Разработаны модели разрешения кореференции, использующие характеристики положения упоминаний в риторическом дереве, что может быть востребовано в системах извлечения фактов и построения знаний; показано, что учёт риторической структуры позволяет повысить качество автоматической классификации текстов в задачах анализа тональности и аргументации по сравнению с классическими моделями на основе BERT-подобных представлений; предложены методы построения структур аргументации для текстов-рассуждений на русском и английском языках, которые могут применяться в экспертных системах, образовательных платформах и системах поддержки принятия решений. Разработанное программное обеспечение внедрено в инструменты лингво-статистического и психолингвистического анализа

текстов ООО «РИ ТЕХНОЛОГИИ», что подтверждено справкой об использовании. Могу также добавить, что мы с коллегами активно использовали разработанные Еленой инструменты в своих исследованиях.

Замечания

В тексте работы содержится некоторое количество опечаток. Также прослеживается разная стилистика оформления некоторых таблиц и разные варианты перевода близких терминов с английского языка на русский. Кроме того, в работе отсутствуют наглядные примеры работы (как «хорошей», так и «плохой») предлагаемых методов на конкретных текстах – это немного затрудняет чтение и понимание диссертации. В целом, отмеченные замечания не являются критичными и никак не снижают ценность представленной работы.

Заключение

Диссертационная работа Е.В. Чистовой представляет собой законченное научно-квалификационное исследование, в котором решена важная научно-техническая задача разработки методов автоматического анализа риторической структуры русскоязычных текстов и исследования возможностей их кросс-языкового и кросс-жанрового использования в системах искусственного интеллекта. Полученные результаты имеют существенное значение для дальнейшего развития методов обработки естественного языка.

На мой взгляд, диссертационная работа соответствует требованиям Положения о присуждении ученых степеней, утвержденного постановлением Правительства РФ от 24 сентября 2013 года № 842, а её автор заслуживает присуждения учёной степени кандидата технических наук по специальности 1.2.1 «Искусственный интеллект и машинное обучение».

Официальный оппонент
кандидат технических наук,
научный сотрудник международной
лаборатории интеллектуальных систем
и структурного анализа
Национального исследовательского
университета «Высшая школа экономики»
Адрес: Москва, Покровский бульвар, 11
16.01.2026 г.



/ Д. А. Ильвовский

