

**МУЛЬТИКОРПУСНОЕ ИССЛЕДОВАНИЕ АСПЕКТНОГО АНАЛИЗА
ТОНАЛЬНОСТИ РУССКОЯЗЫЧНЫХ ТЕКСТОВ**

Абрегова З.Х. (Университет ИТМО)

Научный руководитель – Двойникова А.А.
(Университет ИТМО)

Введение. В последние годы анализ тональности текста стал одной из ключевых задач обработки естественного языка (NLP). Аспектный анализ тональности – важная задача обработки естественного языка, позволяющая определить эмоциональную окраску текста относительно конкретных аспектов. Однако традиционные методы аспектного анализа тональности часто сталкиваются с трудностями при работе с многозначными словами, идиоматическими выражениями и контекстуальными особенностями языков. Мультикорпусный подход предлагает решение этих проблем за счет использования нескольких корпусов данных, что повышает точность и надежность результатов.

Основная часть. Мультикорпусное исследование аспектного анализа тональности предполагает использование множества источников данных для выявления эмоциональной окраски текста. Это может включать в себя различные типы текстов, такие как новостные статьи, отзывы пользователей, сообщения в социальных сетях и другие. Такой подход позволяет учитывать особенности каждого источника и получать более точные результаты. В данном исследовании рассматриваются такие корпуса данных как CABSAR (Corpus for Aspect-Based Sentiment Analysis in Russian-language) [1], RuTermEval-2024 [2], SentiRuEval-2016 [3].

Одним из преимуществ мультикорпусного подхода является возможность использования различных методов анализа тональности, включая машинное обучение, статистические методы и лексический анализ. Это позволяет адаптироваться к специфике каждого корпуса и улучшать качество анализа.

Кроме того, мультикорпусный подход позволяет учитывать культурные и языковые различия между различными источниками данных. Например, одно и то же слово может иметь разную эмоциональную окраску в разных языках или культурах. Использование мультикорпусного подхода помогает избежать ошибок, связанных с такими различиями.

В данной работе рассматриваются методы и подходы к аспектному анализу тональности, применяемые к русскоязычным текстам. В первую очередь, проведен анализ существующих алгоритмов и моделей машинного обучения, используемых для определения тональности в текстах. Особое внимание уделяется различиям в лексических и синтаксических структурах, которые могут влиять на результаты анализа. Далее рассмотрено применение мультикорпусного подхода для выявления культурных и языковых нюансов, которые могут оказывать значительное влияние на восприятие тональности.

Обучая модели на различных наборах данных, можно добиться большей устойчивости к шуму и вариативности языка. Это особенно важно для русского языка, который имеет свои уникальные грамматические и лексические особенности. Кроме того, такой подход позволяет исследовать влияние культурных и социальных факторов на восприятие тональности. Например, одно и то же высказывание может восприниматься по-разному в зависимости от контекста, в котором оно было сделано. Мультикорпусное исследование помогает лучше понять эти контексты и адаптировать модели для более точного анализа.

Выводы. Мультикорпусное исследование аспектного анализа тональности представляет собой перспективное направление в области NLP. Данный подход демонстрирует значительные преимущества в понимании эмоциональной окраски контента. Во-первых, использование нескольких корпусов позволяет учитывать разнообразие стилей и

жанров текстов, что делает анализ более универсальным и точным. Во-вторых, мультикорпусный подход способствует улучшению качества алгоритмов машинного обучения.

Список использованных источников:

1. Moloshnikov I. et al. Named Entity-Oriented Sentiment Analysis with text2text Generation Approach //Proceedings of the International Conference “Dialogue. – 2023. – Т. 2023 (дата обращения: 02.10.2024).

2. Codalab [Электронный ресурс] //— URL: https://codalab.lisn.upsaclay.fr/competitions/19601#learn_the_details (дата обращения: 04.10.2024).

3. Диалог-21. Оценка системы Dialog-21 (2016): результаты оценки системы анализа тональности [Электронный ресурс] //— URL: <https://www.dialog-21.ru/evaluation/2016/sentiment/> (дата обращения: 04.10.2024).