

РАЗРАБОТКА МЕТОДА АНАЛИЗА ТОНАЛЬНОСТИ ТЕКСТА НА ОСНОВЕ СЛОВАРНОГО ПОДХОДА ДЛЯ РУССКОГО ЯЗЫКА

Спевак К.А. (Университет ИТМО)

Научный руководитель – д.т.н., доцент Котельников Е.В.

(Университет ИТМО)

В данном исследовании рассматривается словарный подход к задаче анализа тональности текстов на корпусах русского языка. Особое внимание уделяется реализации методов анализа тональности текста. Также акцент сделан на разработке и реализации программного инструмента для анализа тональности текстов на основе правил с учётом особенностей русского языка.

Ключевые слова: словарный подход, анализ тональности, подход на основе правил, словари оценочной лексики

Введение. В современном мире генерируется большое количество информации за счёт использования социальных сетей и различных видов платформ, на которых люди могут взаимодействовать. На многие продукты, сервисы и организации можно найти обзоры и отзывы. Комментарии и записи в социальных сетях можно рассматривать как рейтинговые системы, потому что по поводу любого события любой человек может выразить своё мнение. Данная информация представляет интерес для обычных пользователей, крупных корпораций и государства, так как она является базой данных для аналитики, позволяющей персонализировать и точно улучшать показатели качества производимых продуктов и предоставляемых услуг.

Существует три основных подхода к автоматическому анализу тональности текстов – на основе традиционного машинного обучения, на основе глубокого обучения и с использованием словарей оценочной лексики. Первые два подхода обладают высокой точностью, но их результаты сложно интерпретировать. Словарный подход имеет преимущества перед машинным обучением в скорости работы и в отсутствии необходимости обучения (использования специально подготовленных данных для обучения, время для реализации данного процесса). При этом он уступает в качестве анализа тональности текста и ограниченности верных предсказаний. Многие ошибки связаны с отсутствием правил обработки токенов и предложений согласно правилам русского языка. Обработка некоторых лингвистических явлений (отрицание, сарказм, употребление предлогов, многозначность слов, специфичные для предметной области термины, построение текста) правилами может улучшить качество анализа тональности текста.

В настоящее время известны исследования, в которых описываются результаты работы словарных подходов на текстовых корпусах русского языка, однако отсутствуют исследования о специфичных для языка ошибках. Также отсутствует метод, который позволил бы учесть лингвистические особенности для русского языка. Такой метод может позволить повысить качество результатов анализа тональности русскоязычных текстов.

Цель данного исследования – разработать и реализовать метод анализа тональности текстов на основе правил для русского языка.

Для реализации данной цели были поставлены следующие задачи:

- 1) обзор существующих словарных подходов и словарей оценочной лексики для русского языка;
- 2) разработка и реализация инструмента для анализа тональности русскоязычных текстов;

3) анализ результатов работы описанного метода на нескольких корпусах и словарях оценочной лексики и выявление распространённых ошибок.

Основная часть. В рамках исследования дано описание метода анализа тональности текстов на основе правил. Также представлен обзор методов и инструментов для анализа тональности текста для английского и русского языков. В практической части представлен анализ результатов применения разработанного метода на текстовых корпусах русского языка и выявления ошибок классификации текстов на три класса – позитивный, негативный и нейтральный.

Выводы. В результате исследования установлено, что в настоящее время отсутствуют программные инструменты в рамках словарного подхода, учитывающие специфику анализа тональности текстов для русского языка. Также выполнен анализ ошибок, выявленных при применении разработанного метода анализа тональности текстов. Использование метода анализа тональности текстов на основе правил показывает лучшие результаты по сравнению со словарными методами, не учитывающими особенности русского языка

Спевак К.А. (автор)

Подпись

Котельников Е.В. (научный руководитель)

Подпись