

УДК 004.89

## АНАЛИЗ ТОНАЛЬНОСТИ ТЕКСТА НА ОСНОВЕ СЛОВАРНОГО ПОДХОДА И МАШИННОГО ОБУЧЕНИЯ

Спевак К.А. (Университет ИТМО)  
Научный руководитель – д.т.н., доцент  
Котельников Е.В.  
(Университет ИТМО)

В данном исследовании рассматриваются разные подходы к задаче анализа тональности текстов на корпусах английского и русского языков. Особое внимание уделяется словарному подходу и реализации методов на его основе для русского языка. Также акцент сделан на сравнительный анализ результатов работы словарного подхода и моделей глубокого нейросетевого обучения.

Ключевые слова: словарный подход, анализ тональности, глубокое обучение, VADER, BERT.

**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 и результаты сложно интерпретировать. Словарный подход менее точен, но при его использовании не требуются размеченные по тональности тексты, процесс обучения отсутствует, а результаты классификации легко объясняются.

В настоящее время известны работы, в которых сравниваются модели глубокого обучения в задаче анализа тональности русскоязычных текстов, однако отсутствуют исследования, посвящённые сравнению глубокого обучения и словарных методов.

Цель данного исследования – представить аналитический обзор моделей и методов анализа тональности текстов в рамках словарного подхода и глубокого обучения.

Для реализации данной цели были поставлены следующие задачи:

- 1) обзор задач анализа тональности и подходов к её решению;

- 2) обзор существующих словарей оценочной лексики и инструментов для анализа тональности;
- 3) сравнение результатов работы словарного метода и моделей глубокого обучения.

**Основная часть.** В рамках исследования дано описание словарей английского и русского языков. Также предоставлен обзор словарей неанглоязычных языков и способов их создания. Практическая часть заключается в сравнении работы инструмента словарного подхода VADER и модели BERT на текстовых корпусах английского и русского языков.

**Выводы.** В результате исследования установлено, что в настоящее время имеет место недостаток обзорных исследований, которые дают сравнение существующих методов анализа тональности текстов и отдельных словарей оценочной лексики на разных языках. Также результатом исследования является сравнительный анализ результатов работы глубокой нейросетевой модели BERT и инструмента на основе словарного подхода VADER.