

УДК 004.934.2

**СЕНТИМЕНТ-АНАЛИЗ РУССКОЯЗЫЧНЫХ ТЕКСТОВ С ИСПОЛЬЗОВАНИЕМ
МЕТОДОВ АВТОМАТИЧЕСКОГО ПЕРЕВОДА НА АНГЛИЙСКИЙ ЯЗЫК**

Двойникова А.А.

(Санкт-Петербургский институт информатики и автоматизации Российской академии наук,
Университет ИТМО)

Научный руководитель – д.т.н., доцент, Карпов А.А.

(Санкт-Петербургский институт информатики и автоматизации Российской академии наук)

В работе рассматривается метод sentiment-анализа русскоязычных текстов с использованием методов автоматического перевода на английский язык. Проводятся экспериментальные исследования анализа тональности текстов с помощью русскоязычных и англоязычных тональных словарей.

Актуальной задачей в области компьютерной лингвистики является выявление эмоциональной окрашенной лексики в тексте. Несмотря на это, системы sentiment-анализа русскоязычных текстов дают довольно низкую точность по сравнению с англоязычными. В работе произведен анализ тональности русскоязычных текстов с использованием методов автоматического перевода на английский язык.

Для проведения экспериментальных исследований в работе использовался текстовый корпус RuSentiment. Он содержит в себе примерно 31 тыс. постов из социальной сети «ВКонтакте». Sentiment-анализ текстов производился при помощи метода, основанного на тональном словаре. В настоящее время в свободном доступе можно найти 4 тональных словаря для анализа русскоязычных текстов: RuSentiLex, LinisCrowd, WordNetAffect (русскоязычная версия), Словарь Белякова. Первым этапом производилась предобработка текстов и строился тернарный классификатор для оценки тональности с использованием всех 4 русскоязычных словарей. Далее текстовый корпус RuSentiment был переведен на английский язык при помощи методов автоматического перевода и также была проведена тернарная классификация с использованием уже англоязычных тональных словарей. Для анализа англоязычных текстов существуют следующие словари эмоциональных слов: WordNetAffect (англоязычная версия), SentiWordNet, SenticNet.

Предложенный метод для sentiment-анализа русскоязычных текстов позволяет повысить точность распознавания эмоций в тексте.

Данное исследование проводится при поддержке РФФИ (проекты № 18-07-01407_а и 19-29-09081_мк)