

Автоматическая оценка впечатлений обучающихся методами анализа тональности (на материале отзывов на онлайн-курсы на русском и английском)

Кирина М.А.

Национальный исследовательский университет
«Высшая школа экономики» (ФГАОУ ВО НИУ ВШЭ)
г. Санкт-Петербург, Российская Федерация
ORCID: <https://orcid.org/0000-0002-7381-676X>
e-mail: mkirina@hse.ru

Тельнина Л.Д.

Национальный исследовательский университет
«Высшая школа экономики» (ФГАОУ ВО НИУ ВШЭ)
г. Санкт-Петербург, Российская Федерация
ORCID: <https://orcid.org/0000-0003-2725-1902>
e-mail: ldtelnina@edu.hse.ru

В статье описывается эксперимент, направленный на сравнение эффективности инструментов анализа тональности для оценки пользовательского опыта на материале публичных отзывов на онлайн-курсы на образовательной платформе Stepik. Рассматриваются результаты автоматического извлечения сентимент-оценок пользователей на соответствующие курсы как на русском, так и на английском языках. Для русскоязычных текстов обсуждается применение словаря эмотивной лексики «КартаСловСент» и предобученной на датасете RuSentiment модели, реализованной в библиотеке *dostoevsky*. Для отзывов на английском языке апробируются такие популярные NLP-библиотеки, как *TextBlob* и *VADER*. Прослеживается, как соотносится оценка за онлайн-курс, выставляемая пользователями по окончании обучения – по шкале от 1 до 5 – с эмоциональной оценкой, полученной на основе анализа их отзывов. Анализируется лексика, характерная для отзывов, описывающих положительный и негативный опыт от обучения. Предполагается, что совмещение инструментов позволит в дальнейшем разработать программу для интеллектуального анализа впечатлений пользователей от онлайн-курсов с целью улучшения предлагаемых образовательных продуктов.

Ключевые слова: анализ тональности, моделирование пользовательского опыта, онлайн-образование, автоматическая обработка естественного языка, смысловая компрессия текста, компьютерная лингвистика.

Финансирование. Публикация подготовлена в результате проведения работы по проекту № 21–04–053 «Методы искусственного интеллекта для филологических исследований» в рамках Программы «Научный фонд Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ)» в 2022 г.

Благодарности. Авторы благодарят за помощь в сборе данных А.Д. Москвину.

Для цитаты:

Кирина М.А., Тельнина Л.Д. Автоматическая оценка впечатлений обучающихся методами анализа тональности (на материале отзывов на онлайн-курсы на русском и английском) // Цифровая гуманитаристика и технологии в образовании (ДНТЕ 2022): сб. статей III Всероссийской научно-практической конференции с международным участием. 17–18 ноября 2022 г. / Под ред. В.В. Рубцова, М.Г. Сороковой, Н.П. Радчиковой. М.: Издательство ФГБОУ ВО МГППУ, 2022. 355–374 с.

Введение

Анализ тональности, или сентимент-анализ (*sentiment analysis*), представляет собой одно из активно развивающихся направлений в области автоматической обработки текстов на естественном языке. Он включает в себя класс методов контент-анализа, целью которых можно назвать смысловую компрессию информации, содержащейся в тексте. Автоматический анализ тональности широко применяется в таких областях, как компьютерная лингвистика [9], социология [2], политология [9], продуктовая аналитика [6], маркетинг [11], психология [13] и цифровая гуманитаристика [4].

Результатом применения методов анализа тональности называют «выявление мнения автора текста по поводу предмета, обсуждаемого в тексте» [1]. По этой причине в фокусе исследований, связанных с задачей автоматического определения сентимента в тексте, оказывается так называемый пользовательский контент (*user-generated content, UGC*) – все то, что может содержать мнение людей о качестве или отдельных характеристиках какого-либо продукта [12]. Для моделирования пользовательского опыта используются данные социальных медиа, среди которых особенно часто анализируются отзывы на товары и услуги. Опубликованные на официальных и/или вторичных сайтах, представленные в виде комментариев в социальных сетях, отзывы пользователей помогают компаниям принимать решения по улучшению предлагаемого продукта.

Интерес вызывают также и отзывы, оставляемые пользователями на образовательные онлайн-курсы, в ходе обработки которых, как правило, извлекается мнение участников как о процессе и

результатах обучения, так и об их опыте взаимодействия непосредственно с платформой [8]. Очевидно, что учет мнений пользователей, причем выраженных в текстах отзывов, а не только в виде количественной оценки, которую традиционно предлагают поставить курсу по завершении, позволяет разработчикам развивать продукт, внося более ориентированные на целевую аудиторию изменения. При этом важным кажется не только определение количества позитивных и негативных отзывов, но и выявление конкретных характеристик продукта, которые вызвали соответствующее настроение пользователя.

Целью исследования является проведение эксперимента, состоящего в сравнении инструментов анализа тональности для оценки пользовательского опыта с точки зрения его положительности или отрицательности применительно к публичным отзывам на онлайн-курсы, размещенные на образовательной платформе Stepik (<https://stepik.org>). При этом рассматриваются результаты автоматического извлечения оценок пользователей на соответствующие курсы как на русском, так и на английском языках. Кроме того, прослеживается, как соотносится оценка за онлайн-курс, выставляемая пользователями по окончании, с эмоциональной оценкой, полученной на основе анализа их отзывов, а также определяется лексика, характерная для отзывов разной полярности – положительной и отрицательной. Предполагается, что совмещение инструментов позволит в дальнейшем разработать программу для интеллектуального анализа впечатлений пользователей от онлайн-курсов. Это даст возможность авторам онлайн-курсов улучшать предлагаемые образовательные продукты, отталкиваясь от эмоциональных оценок, извлеченных напрямую из отзывов обучающихся.

Материал

Stepik – российская образовательная платформа, предлагающая обучающимся онлайн-курсы широкой тематики. При этом важно отметить, что автором курса может выступить не только образовательная организация, но и любой пользователь. Несмотря на то, что большинство курсов представлено на русском языке, часть публикуется и на английском – для обеспечения инклюзивности иностранных студентов или в соответствии со спецификой самого онлайн-курса (н-р, языковой онлайн-курс «English for law students»). По завершении обучения каждому пользователю предлагается оценить онлайн-курс (по шкале от 1 до 5) и оставить отзыв.

Далее будут описаны этапы sentiment-анализа и применяемые для этого библиотеки для сформированных выборок отзывов на русско- и англоязычные онлайн-курсы соответственно. Очевидно, что пользовательские оценки не обязательно будут совпадать с sentiment-оценкой, извлекаемой на основе отзывов, однако выдвигается предположение о том, что количественная и «текстовая» оценки пользователей обладают определенной согласованностью.

Методика анализа русскоязычных отзывов

Объем исследуемой выборки на русском языке составляет 5994 отзыва, соответствующих 7 онлайн-курсам разной тематики – от введения в программирование до подготовки к ЕГЭ по русскому языку. Отметим, что вместе с текстами отзывов выгружается также и информация об оценках, оставленных пользователями онлайн-курсу в целом. Критерии отбора онлайн-курсов для исследования были следующие: 1) на курс должно быть оставлено более 400 отзывов; 2) курсы различаются между собой средней оценкой, т.е. обеспечивается наличие в выборке как положительных, как и отрицательных оценок (рис. 1).

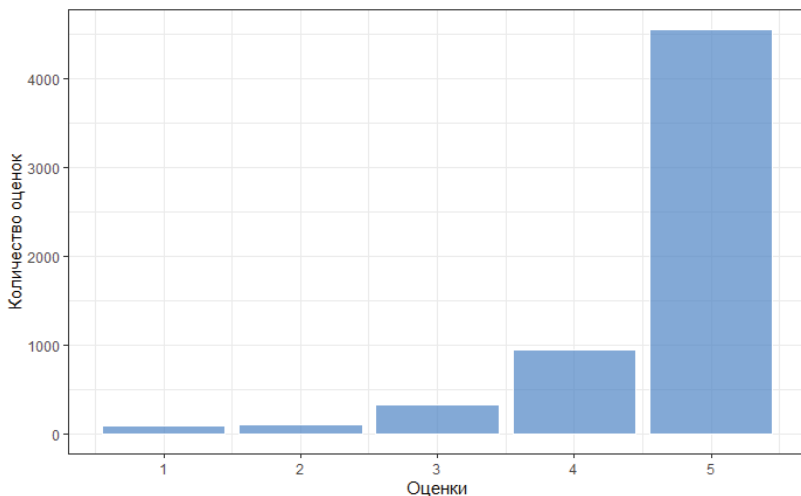


Рис. 1. Распределение отзывов по пользовательским оценкам

Тональность собранных текстов определялась с помощью двух подходов: подхода на основе словарей и подхода на основе машинного обучения. Первый предполагает использование эмоционально-

размеченных словарей, присваивающих лексемам в тексте (при условии, что они совпадают с лексемами из словаря) или определённое числовое значение в рамках некоторого диапазона (например, от -1 до 1), или значение, указывающее на принадлежность к классу («позитивный», «негативный» и др.). Количество и типы классов определяются составителями словарей. Второй подход – на основе машинного обучения – заключается в применении модели, предобученной на выборке размеченных текстов, к новым данным, либо обучении кастомного классификатора.

Для реализации подхода на основе словарей тексты отзывов были предобработаны: из них удалялись знаки препинания, цифры и лишние пробелы; все текстовые символы были приведены к нижнему регистру. Затем с помощью *mystem* (<https://yandex.ru/dev/mystem>) была проведена лемматизация – автоматическое приведение словоформы к её начальной форме. Далее из текстов были удалены стоп-слова – служебные части речи и прочие, не являющиеся информативными лексемами. После, к датафрейму был применён готовый словарь эмотивной лексики «КартаСловСент» [5]. Значения эмоциональной оценки в нём находятся в непрерывном диапазоне от -1 до 1, где -1 – максимально отрицательная полярность, а 1 – максимально положительная. В результате для каждого отзыва было получено среднее значение эмоциональной оценки, а также были выявлены самые частотные отрицательные и положительные слова.

В рамках второго подхода – на основе машинного обучения – было решено воспользоваться готовой моделью *dostoevsky* (<https://github.com/bureaucratic-labs/dostoevsky>), предобученной на крупнейшем для русского языка наборе данных для эмотивной разметки *RuSentiment* (<https://text-machine.cs.uml.edu/projects/rusentiment/>). Модель позволяет предсказать тональность текста, присваивая ему по каждой полярности (негативная, нейтральная, позитивная) значения от 0 до 1. Следует отметить, что при её использовании тексты не требуется предварительно токенизировать и лемматизировать. В результате для каждого отзыва была получена информация о вероятности принадлежности к классу «положительных» и к классу «отрицательных».

Результаты анализа русскоязычных отзывов

При использовании первого подхода были выявлено, что средние значения тональности во многом соотносятся со средними оценками за курсы (рис. 2). Так, у курсов «Подготовка к ЕГЭ...» и «Введение в программирование C++» наивысший средний балл и наивысшая средняя тональность, т.е. на отзывы приходится большее количество

позитивных слов. В то время как для курсов «Введение в базы данных» и «Тестирование ПО...» была получена самая низкая средняя тональность, что также коррелирует со усредненной оценкой за курс, выставляемой пользователем. Однако есть и несоответствия в пользовательских и тональных оценках: для онлайн-курса «Java. Базовый курс» тональная оценка оказалась выше средней пользовательской.

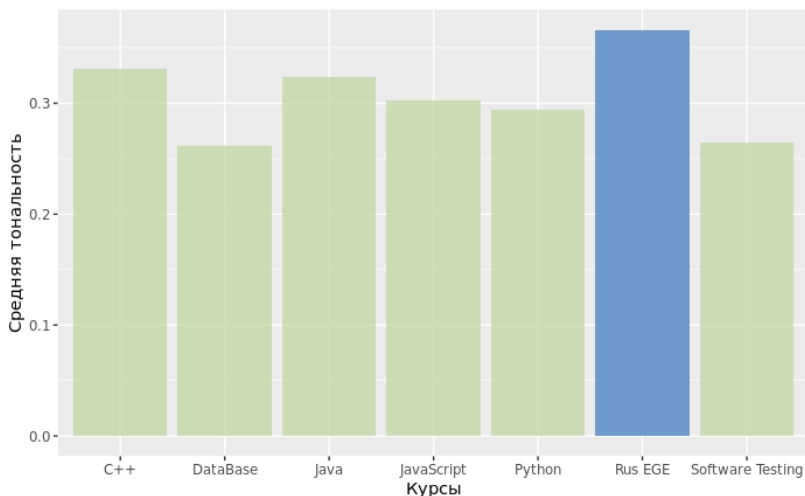


Рис. 2. Распределение средней тональной оценки по курсам

Что касается языковых особенностей рассматриваемых текстов, для самых положительных (88 отзывов) отмечается небольшая длина – в среднем от 2 до 5 слов. То же верно и для самых отрицательных отзывов, и, в целом, они малоинформативны с точки зрения критики конкретных тем или заданий курса, собственно вызвавших негативную оценку обучающихся. Примеры самых положительных и самых отрицательных отзывов представлены в табл. 1.

Таблица 1

Самые положительные и отрицательные отзывы на русскоязычные онлайн-курсы (словарь «КартаСловСент»)

Название онлайн-курса	Оценка пользователя	Отзыв	Тональность отзыва
Введение в программирование C++	5	<i>мне все понравилось</i>	1
Введение в программирование C++	5	<i>обожая его</i>	1

Название онлайн-курса	Оценка пользователя	Отзыв	Тональность отзыва
JavaScript для начинающих	5	<i>годно</i>	1
Введение в программирование C++	5	<i>отличный курс один из лучших на степике</i>	1
Введение в программирование C++	5	<i>понравилось</i>	1
Тестирование ПО: Подготовка к сертификации ISTQB Foundation	1	<i>ужас</i>	-1
JavaScript для начинающих	4	<i>слабоват</i>	-0.82
Подготовка к ЕГЭ по русскому языку	5	<i>бе</i>	-0.79
Введение в программирование C++	5	<i>башка болит после вас</i>	-0.77
JavaScript для начинающих	1	<i>ужас курс javascript для мазохистов</i>	-0.58

Однако если обратить внимание на менее однозначные случаи, то можно обнаружить в текстах с отрицательной полярностью указания на некоторые недостатки рассматриваемых онлайн-курсов:

1. *местами задания нудные и неудобные и побольше бы практики с кодом* («JavaScript для начинающих», -0.495)
2. *худший курс на степике очень непонятный и плохой* («JavaScript для начинающих», -0.424)

В результате применения библиотеки *dostoevsky*, реализующей подход на основе машинного обучения, было также выявлено, что средняя негативная тональность соотносится со средним баллом: чем ниже средний балл, тем выше значение негативной тональности. В целом, тенденция для каждого онлайн-курса сохраняется подобно выявленной в рамках описанного выше эксперимента (рис. 3). Причём точно так же, как и в предыдущем подходе, разница между средними значениями курсов «Введение в базы данных» и «Подготовка к ЕГЭ по русскому» составляет около 0.1.

Распределение средней положительной тональности тоже соотносится со средними баллами (рис. 4). Интересно, что тональность курса, посвящённого программированию на C++ немного выше, чем у курса «Подготовка к ЕГЭ по русскому», хотя при этом значение негативной тональности у него тоже выше. Предположительно,

это можно объяснить тем, что, несмотря на высокий средний балл, в отзывах на этот курс чаще встречаются различные замечания, а также оценки ниже 3. Положительная тональность курсов по Java и Python ожидаемо выше, чем для JavaScript.

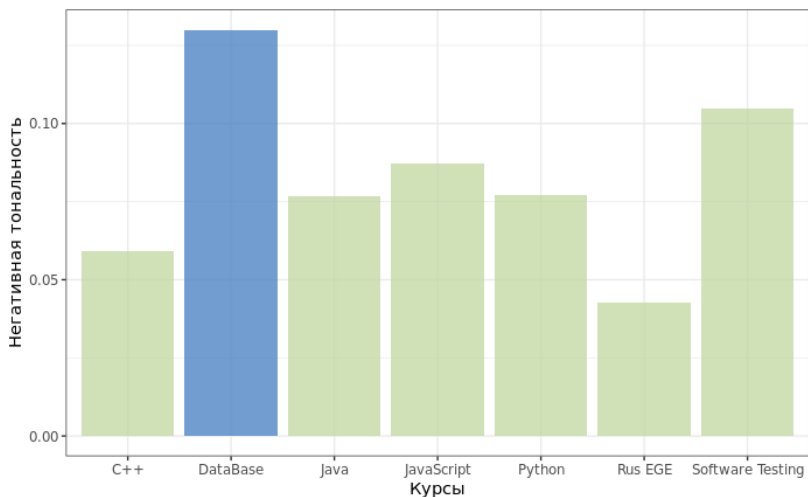


Рис. 3. Распределение средней негативной тональной оценки по курсам

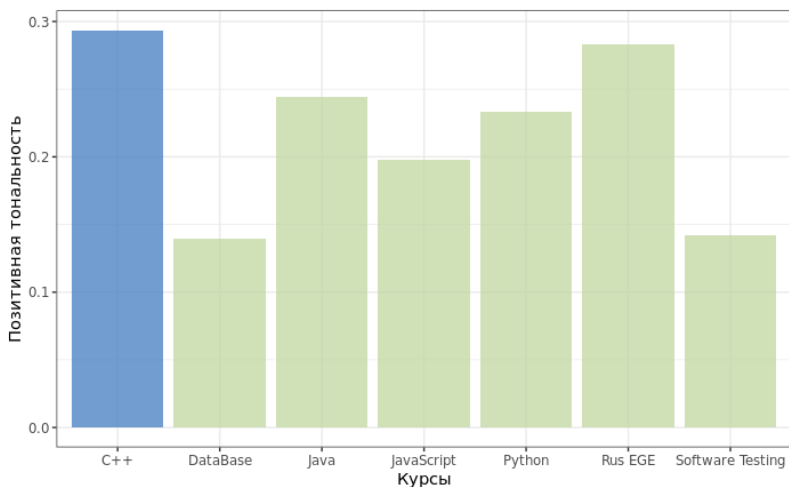


Рис. 4. Распределение средней положительной тональной оценки по курсам

Большинство положительных отзывов содержат от 1 до 3 слов (например, *супер, отличный курс, мне все понравилось*), в связи с чем их трудно назвать информативными. Поэтому для подробного рассмотрения были отобраны наиболее положительные отзывы с 5 и более словами (табл. 2). Ожидаемо, в них встречается яркая оценочная лексика, например: *как прекрасно, в восторге, замечательный, отличный, чудесный*.

Таблица 2

**Самые положительные отзывы на русскоязычные
онлайн-курсы (библиотека dostoevsky)**

Название онлайн-курса	Оценка пользователя	Отзыв	Положительная тональность
JavaScript для начинающих	5	<i>замечательный курс лично мне понравился</i>	0.9978
Введение в программирование C++	5	<i>все хорошо супер даже у меня получилось</i>	0.9975
Введение в программирование C++	4	<i>начало отлично окончание так себе</i>	0.9940
Подготовка к ЕГЭ по русскому языку	5	<i>очень крутой курс я в восторге</i>	0.9937
Подготовка к ЕГЭ по русскому языку	5	<i>чудесный материал шикарная подача всё четко по полочкам ничего лишнего</i>	0.9935

Что касается негативных отзывов, интересно, что среди них оказались и те, которым автор отзыва поставил общую оценку, равную 4 или 5. Обратив внимание на такие отзывы, можно заметить, что они действительно содержат замечания относительно устройства курса (табл. 3).

Таблица 3

**Самые отрицательные отзывы
на русскоязычные онлайн-курсы
(библиотека dostoevsky)**

Название онлайн-курса	Оценка пользователя	Отзыв	Негативная тональность
JavaScript для начинающих	1	<i>ужасно</i>	0.982
Введение в базы данных	2	<i>галопом по европам</i>	0.919
Java. Базовый курс	5	<i>это не базовый курс а только для тех у кого есть опыт как бы грустно это не звучало</i>	0.851

Название онлайн-курса	Оценка пользователя	Отзыв	Негативная тональность
Введение в программирование C++	5	<i>мне зашло но как мне по мне задания сложные но скучноватые</i>	0.831
Введение в программирование C++	4	<i>курс неплохой но иногда обманывают</i>	0.831

Таким образом, оба инструмента позволили проанализировать отзывы с точки зрения их эмоциональности. Однако заметно, что библиотека *dostoevsky* позволяет лучше распознавать негативные сентименты, даже в положительных отзывах, и, следовательно, более точно выявлять замечания студентов.

Методика анализа англоязычных отзывов

Для анализа были выбраны отзывы на англоязычные онлайн-курсы, опубликованные на платформе Stepik. В результате было выгружено 2080 отзывов к 96 онлайн-курсам разной тематики. Несколько на то, что курс находится в категории англоязычных, любой пользователь может его пройти, а также оставить отзыв на родном языке. В этой связи дополнительно проверялся язык отзыва и оставлялись только те, которые действительно были на английском, согласно метаданных о переводе (встроенная функция на платформе Stepik). Так, если перевод для отзыва присутствовал, то отзыв был не на английском и, следовательно, он удалялся из выборки. Итоговый объем выборки составил – 1739 отзывов. Распределение отзывов по оценкам в анализируемой выборке представлено на рис. 5.

В эксперименте применяются такие библиотеки для анализа тональности, как TextBlob (<https://pypi.org/project/textblob>) и VADER (<https://pypi.org/project/vaderSentiment>). Выбор обусловлен тем, что данные инструменты являются примерами реализации подходов на основе правил, отличающихся большей точностью при анализе текстов, содержащих специфическую лексику, а также недостаточным на данном этапе количеством отзывов для обучения кастомного классификатора. Кроме того, эти библиотеки пользуются популярностью среди современных NLP-технологий, применяющихся для анализа тональности.

TextBlob – библиотека с открытым кодом, представляющая собой API для решения широкого спектра задач в области обработки естественного языка [7]. Извлекаемая данной библиотекой тональ-

ная оценка – полярность – характеризует собственно эмоциональность текста и может принимать значения в промежутке от -1 до +1, где тексты с отрицательными полярностями считаются содержащими отрицательные мнения, а с положительными – положительные. Если полярность равна 0, то его можно считать нейтральным. Однако важно уточнить, что присвоение нулевого значения может быть обусловлено тем, что значительное количество слов в анализируемом фрагменте не было обнаружено в словаре.

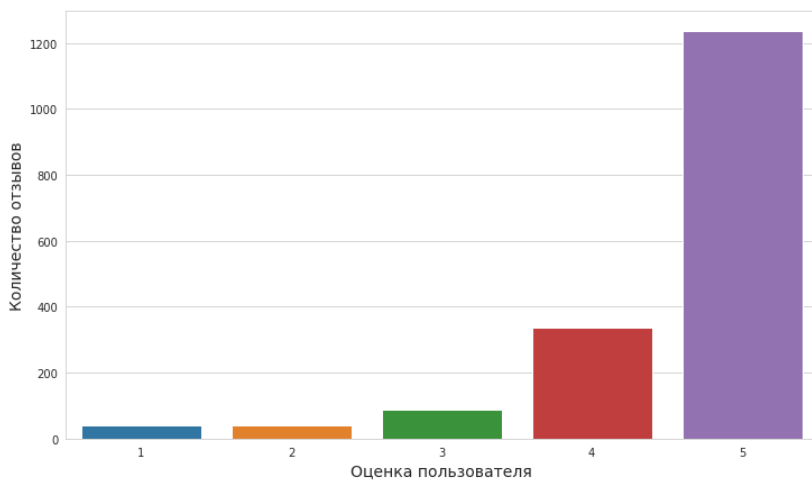


Рис. 5. Распределение отзывов по пользовательским оценкам

При взвешивании слов используется лексикон, содержащий список оценочных прилагательных на английском языке; он доступен в репозитории на GitHub (<https://github.com/sloria/TextBlob/blob/dev/textblob/en/en-sentiment.xml>). Как видно из примера ниже, каждому слову соответствует оценка полярности в зависимости от его значения:

```
<word form="amazing" sense="inspiring awe or admiration or wonder" polarity="0.8" intensity="1.0" confidence="0.9" />
```

В качестве модификаторов для английского используются наречия (например, *very*). Подобные слова с учетом присвоенного им в словаре значения интенсивности, как правило, увеличивают оценку полярности. При взвешивании sentiment-оценки, правила, регулирующие работу этого модуля библиотеки, учитывают также и отрицательные частицы, которые понижают сентимент.

VADER (*Valence Aware Dictionary for sEntiment Reasoning*) – библиотека с открытым кодом, разработанная для анализа тональности текстов социальных медиа [3]. Он, как и TextBlob, сочетает использование лексикона и правил для вывода сентимент-оценки. При этом словарь содержит не только собственно слова, но и обширный лист эмотиконов в западном стиле и акронимов, а также сленговые выражения. В результате применения библиотеки извлекается следующая информация о тексте: *compound* – непосредственно тональная оценка; *pos*, *neu*, *neg* – пропорции текста, относящиеся к соответствующим категориям. Тональная оценка принимает значения в диапазоне от -1 до +1. Чем выше тональная оценка, тем более положителен текст, и наоборот. Обычно используются следующие пороговые значения для отнесения текста к одной из трех категорий:

1. если *compound score* ≥ 0.05 , то текст положителен;
2. если *compound* ≤ -0.05 , то текст негативен;
3. в остальных случаях оценка интерпретируется как нейтральная.

Как и TextBlob, VADER учитывает отрицания и усиливающие сентимент модификаторы. Помимо этого, такие неконвенциональные случаи, как обильное использование пунктуационных знаков и намеренное изменение регистра, воспринимаются анализатором как «сигналы» для повышения сентимента в положительную или отрицательную сторону в зависимости от контекста.

Результаты анализа англоязычных отзывов

К каждому отзыву применялись библиотеки TextBlob и VADER. На рис. 6 представлено распределение по пользовательским оценкам средних сентимент-оценок каждой из библиотек. Можно заметить, что для скорее негативных, т.е. меньше 3, пользовательских оценок результаты «положительнее», по сравнению с VADER, демонстрирует TextBlob. Это соотносится с наблюдениями, сделанными эмпирическим путем. Примечательно, что отзывы тех пользователей, которые впоследствии поставили онлайн-курсу оценку 2, TextBlob выставил нейтральную оценку (*polarity*=0).

При сравнении отзывов, классифицированных библиотеками TextBlob и VADER как самые положительные и самые отрицательные, можно заметить, что длина отзывов значительно различается: TextBlob присваивает тональную оценку безотносительно характера полярности, как правило, более коротким отзывам (табл. 4). Вероятно, это связано с тем, что в основе TextBlob лежит более распространенный словарь и список правил для определения тональности текста.

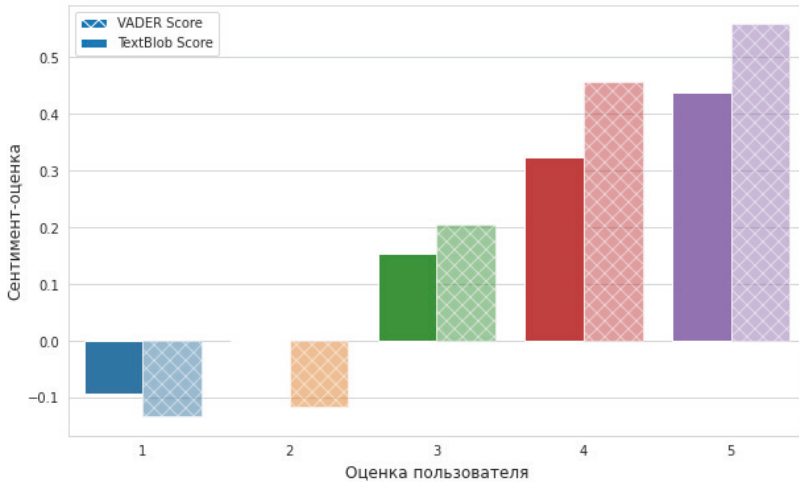


Рис. 6. Распределение средней тональной оценки по пользовательским оценкам

Таблица 4

Самых положительные и отрицательные отзывы на англоязычные онлайн-курсы (библиотека TextBlob)

Название онлайн-курса	Оценка пользователя	Отзыв	Тональная оценка (<i>polarity</i>)
Java. Functional programming [OLD, 2017]	5	<i>The best of the best.</i>	1.0
Basic Life Support and Automated External Defibrillation	5	<i>Excellent I am really impressed.</i>	1.0
Java. Collections Framework.	5	<i>Great course, a lot of practice, like it!</i>	1.0
Artificial Intelligence	1	<i>worst course</i>	-1.0
Introduction to Python	1	<i>this is the worst thing i have ever had to do.</i>	-1.0
Introduction to Python	1	<i>This is awful</i>	-1.0

Далее, если обратить внимание на отрицательные отзывы, можно также увидеть, что VADER не только справляется с более распространенными отзывами, но и лучше улавливает контексты с отрицанием. Помимо этого, VADER чувствителен к эмоджонам и эмодзи, часто встречающимся в подобных текстах (табл. 5).

Таблица 5

**Самых положительные и отрицательные отзывы
 на англоязычные онлайн-курсы (библиотека VADER)**

Название онлайн-курса	Оценка пользователя	Отзыв	Тональная оценка (compound)
Easy Way to Technical Writing	5	<i>I would like to thank you for such amazing, wonderful, and useful training course. And also, I express my appreciation to you for providing us with such wonderful materials. The course is really inspiring. I hope, that the course will be improved in the future and it will make future learners happy as well. Best wishes for you.</i>	0.9914
Data Structures	5	<i>Very good course. I joined to it because wanted to read about hashes but explanation was so perfect that I decided to complete all modules. And didn't regret about it. It would be a good idea to get more coding exercises. Implement part of this algorithms from scratch for example. Maybe authors should think about the second part of this course :)</i>	0.9664
My pronunciation guide	5	<i>Excellent course. I am happy to say that I have really enjoyed in learning this course. Thanks a lot</i>	0.9633
Basic Life Support and Automated External Defibrillation	3	<i>The test is stupid, all answers are wrong, which is the correct then? Huh? Your not letting me get points on it on purpose!</i>	-0.7993
Intro to Programming in Python then Java using Media Computation	1	<i>I don't like stepik because it doesn't offer me enough feedback so I can figure out what I did wrong and fix it. All it says is wrong try again.</i>	-0.8079
LaunchCode's Discovery	2	<i>The information given here works as a primer for coding. However, it doesn't measure up to other tutorial websites. I'm trying to trouble shoot my code, and all I see is «wrong» No hints, no mix of green and red text.</i>	-0.8182

Что касается зависимости тональных оценок от пользовательских, то она присутствует и более выражена для VADER (рис. 7).

При этом сохраняется разброс в оценках обоих инструментов для негативных впечатлений пользователей.

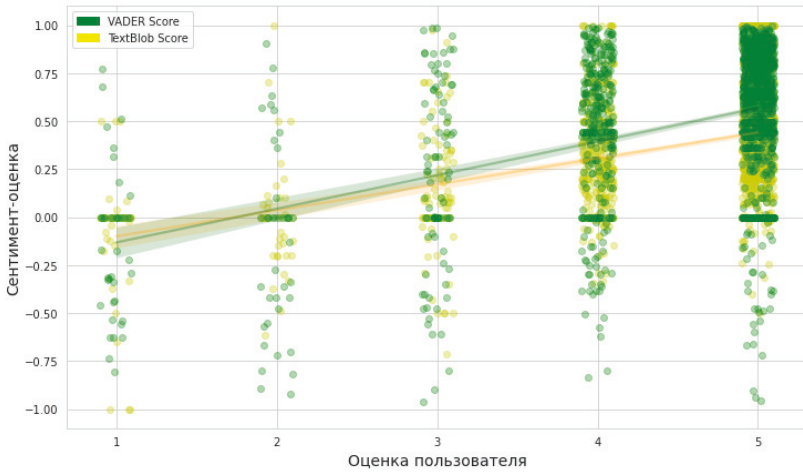


Рис. 7. Взаимосвязь тональных и пользовательских оценок

Обнаруженная предпочтительность VADER перед TextBlob для анализа текстов, которым свойственна неформальная лексика, эмодзи и др., а также имеющих ненормированный объем, коррелирует с наблюдениями, сделанными в [10, 14].

Заключение

В статье описывается серия экспериментов, направленных на сравнение ряда существующих инструментов для анализа тональности применительно к отзывам на онлайн-курсы на русском и английском языках. В качестве материала используются публичные отзывы, размещенные на платформе Stepik. Для русскоязычных текстов обсуждается применение словаря эмотивной лексики «Карта-СловСент» и предобученной на датасете RuSentiment модели, реализованной в библиотеке *dostoevsky*. Для отзывов на английском языке апробируются такие популярные NLP-библиотеки, как TextBlob и VADER. Кроме того, на каждом этапе сравниваются полученные на основе анализа отзывов тональные оценки – положительные и негативные, с соответствующими им количественным оценками (от 1 до 5), выставленных авторами тех же отзывов по завершении курса.

Что касается полученных результатов, для русскоязычного материала отмечается, что в рамках подхода на основе словарей особенно

удобными кажутся специальные словари. Они лучше адаптированы под контекст и могут выдавать более точные результаты, однако для их создания необходима предварительная классификация текстов. Более распространены универсальные словари, которые могут применяться к текстам, не имеющим никакой системы оценивания (например, комментарии к новостной записи) и получать довольно точные результаты. Однако если интерес представляют негативные сентименты, то, как было выявлено в ходе эксперимента, более точный результат позволяет получить использование предобученной модели (тестируется библиотека *dostoevsky*). Применяв ее, удалось обнаружить негативные сентименты в отзывах с положительными оценками, что особенно важно при анализе отзывов обучающихся на онлайн-платформах. Тем не менее и этой модели свойственно ошибаться или присваивать более высокие негативные значения текстам, которые хоть и содержат критику, но являются скорее сдержанными, чем экспрессивными.

В отношении англоязычных отзывов наблюдается преимущество библиотеки VADER над библиотекой TextBlob. VADER хорошо улавливает негативные контексты, «смешанные» отзывы, а также похожие на тексты социальных медиа – избыливающие эмоциями и неформальной лексикой.

Вне зависимости от языка, обнаруживается корреляция между пользовательскими оценками и тональными оценками. При этом, как правило, более высоко оценённые курсы имеют более высокую среднюю позитивную тональность, и наоборот. Это свидетельствует об определённой эффективности и надёжности использованных подходов. Продолжение работы видится в разработке проблемы выбора оптимального порогового значения для классификации отзывов на позитивные и негативные. Предполагается, что выявление условий, наиболее коррелирующих с количественными оценками, может способствовать улучшению результатов.

Литература

1. Автоматическая обработка текстов на естественном языке и анализ данных : учеб. пособие / Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Кльшинский Э.С., Лукашевич Н.В., Сапин А.С. М.: Изд-во НИУ ВШЭ, 2017. 269 с.
2. *Al-Razgan M. et al.* Using diffusion of innovation theory and sentiment analysis to analyze attitudes toward driving adoption by Saudi women // *Technology in Society*. 2021. Vol. 65. P. 101558. doi:10.1016/j.techsoc.2021.101558
3. *Hutto C., Gilbert E.* VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text // *Proceedings of the*

- International AAAI Conference on Web and Social Media. 2014. Vol. 8 (1). P. 216–225.
4. *Kim E., Klinger R.* A Survey on Sentiment and Emotion Analysis for Computational Literary Studies. 2021. P. 38. doi:10.48550/arXiv.1808.03137
 5. *Kulagin D.I.* Publicly available sentiment dictionary for the Russian language KartaSlovSent. 2021. P. 1106–1119. doi:10.28995/2075–7182–2021–20–1106–1119
 6. *Liang Y., Liu Y., Loh H.T.* Exploring Online Reviews For User Experience Modeling. 2013. P. 10.
 7. *Loria S.* textblob Documentation, 2020. P. 73.
 8. *Ngoc T.V., Thi M.N., Thi H.N.* Sentiment Analysis of Students' Reviews on Online Courses: A Transfer Learning Method. 2021. P. 9.
 9. *Phan M.H., Ogunbona P.O.* Modelling Context and Syntactical Features for Aspect-based Sentiment Analysis Online: Association for Computational Linguistics, 2020. P. 3211–3220. doi: 10.18653/v1/2020.acl-main.293
 10. *Pope L.* Comparing VADER and Text Blob to Human Sentiment // Medium [Электронный ресурс]. URL: <https://towardsdatascience.com/comparing-vader-and-text-blob-to-human-sentiment-77068cf73982> (дата обращения: 02.09.2022)
 11. *Reyes-Menendez A., Saura J.R., Filipe F.* Marketing challenges in the #MeToo era: gaining business insights using an exploratory sentiment analysis // Heliyon. 2020. Vol. 6, Issue 3. doi: 10.1016/j.heliyon.2020.e03626
 12. *Smetanin S.* The Applications of Sentiment Analysis for Russian Language Texts: Current Challenges and Future Perspectives // IEEE Access. 2020. Vol. 8. P. 110693–110719. doi: 10.1109/ACCESS.2020.3002215
 13. *Wang X. et al.* A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network // J. Li et al. (eds.), Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. P. 201–213. doi: 10.1007/978–3–642–40319–4–18
 14. *White B.* Sentiment Analysis: VADER or TextBlob? // Medium [Электронный ресурс]. URL: <https://towardsdatascience.com/sentiment-analysis-vader-or-textblob-ff25514ac540> (дата обращения: 02.09.2022)

Информация об авторах

Кирина Маргарита Александровна, магистрант, приглашенный преподаватель департамента филологии, Национальный исследовательский университет «Высшая школа экономики» (ФГАОУ ВО НИУ ВШЭ), г. Санкт-Петербург, Российская Федерация, ORCID: <https://orcid.org/0000-0002-7381-676X>, e-mail: mkirina@hse.ru

Тельнина Людмила Дмитриевна, студент, Национальный исследовательский университет «Высшая школа экономики» (ФГАОУ ВО НИУ ВШЭ), г. Санкт-Петербург, Российская Федерация, ORCID: <https://orcid.org/0000-0003-2725-1902>, e-mail: ldtelnina@edu.hse.ru

Sentiment Analysis for Automatic Assessment of Learners' Experience (on the Basis of Reviews on Online Courses in Russian and English)

Margarita A. Kirina

National Research University Higher School of Economics
Saint Petersburg, Russia
ORCID: <https://orcid.org/0000-0002-7381-676X>
e-mail: mkirina@hse.ru

Ludmila D. Telnina

National Research University Higher School of Economics
Saint Petersburg, Russia
ORCID: <https://orcid.org/0000-0003-2725-190>
e-mail: ldtelnina@edu.hse.ru

The paper describes an experiment aimed at comparing the effectiveness of sentiment analysis tools for evaluating user experience based on public reviews on online courses on Stepik. The results of automatic extraction of sentiment scores for the corresponding reviews both in Russian and in English are considered. With regards to reviews on online courses in Russian, the use of the KartaSlovSent dictionary of emotive lexis and the model pre-trained on the RuSentiment dataset implemented in the dostoevsky library is discussed. Considering the reviews in English, popular NLP libraries, such as TextBlob and VADER, are tested. It is analyzed how the scores, on the scale from 1 to 5, which are given by users after the completion of online course correlate with the sentiment scores obtained on the basis of learners' feedback. The lexical characteristics of reviews describing positive and negative experiences from online education are discussed. It is suggested that the combination of tools will allow developing a program for the text mining of user experiences from online courses in order to improve the educational products in question.

Keywords: sentiment analysis, user experience modelling, online education, natural language processing, semantic compression of text, computational linguistics.

Funding. The publication was prepared within the framework of the Academic Fund Program at the National Research University Higher School of Economics (HSE) in 2022 (grant # 21-04-053 'Artificial Intelligence Methods in Literature and Language Studies').

Acknowledgements. The authors are grateful to Moskvina A.D. for assistance in data collection.

For citation:

Kirina M.A, Telnina L.D. Sentiment Analysis for Automatic Assessment of Learners' Experience (on the Basis of Reviews on Online Courses in Russian and English) // *Digital Humanities and Technology in Education (DHTE 2022): Collection of Articles of the III All-Russian Scientific and Practical Conference with International Participation. November 17–18, 2022 / V.V. Rubtsov, M.G. Sorokova, N.P. Radchikova (Eds). Moscow: Publishing house MSUPE, 2022. 355–374 p. (In Russ., abstr. in Engl.).*

References

1. Avtomaticheskaya obrabotka tekstov na estestvennom yazyke i analiz dannykh: ucheb. posobie [Automatic processing of natural language texts and data analysis] / Bol'shakova E.I., Vorontsov K.V., Efremova N.E., Klyshinskii E.S., Lukashevich N.V., Sapin A.S. Moscow: NIU VShE Publ., 2017. 269 p. (In Russ.).
2. Al-Razgan M. et al. Using diffusion of innovation theory and sentiment analysis to analyze attitudes toward driving adoption by Saudi women. *Technology in Society*, 2021. Vol. 65. p. 101558. doi:10.1016/j.techsoc.2021.101558
3. Hutto C., Gilbert E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, 2014. Vol. 8, no. 1, pp. 216–225.
4. Kim E., Klinger R. A Survey on Sentiment and Emotion Analysis for Computational Literary Studies, 2021. p. 38. doi:10.48550/arXiv.1808.03137
5. Kulagin D.I. Publicly available sentiment dictionary for the Russian language KartaSlovSent, 2021. pp. 1106–1119. doi:10.28995/2075-7182-2021-20-1106-1119
6. Liang Y., Liu Y., Loh H.T. Exploring Online Reviews For User Experience Modeling, 2013. p. 10.
7. Loria S. textblob Documentation, 2020. p. 73.
8. Ngoc T.V., Thi M.N., Thi H.N. Sentiment Analysis of Students' Reviews on Online Courses: A Transfer Learning Method, 2021. p. 9.
9. Phan M.H., Ogunbona P.O. Modelling Context and Syntactical Features for Aspect-based Sentiment Analysis Online: Association for Computational Linguistics, 2020. pp. 3211–3220. doi: 10.18653/v1/2020.acl-main.293
10. Pope L. Comparing VADER and Text Blob to Human Sentiment. Medium [Online resource]. URL: <https://towardsdatascience.com/comparing-vader-and-text-blob-to-human-sentiment-77068cf73982> (Accessed 02.09.2022)
11. Reyes-Menendez A., Saura J.R., Filipe F. Marketing challenges in the #MeToo era: gaining business insights using an exploratory sentiment analysis. *Heliyon*, 2020. Vol. 6, iss. 3. doi: 10.1016/j.heliyon.2020.e03626

12. Smetanin S. The Applications of Sentiment Analysis for Russian Language Texts: Current Challenges and Future Perspectives. *IEEE Access*. 2020, Vol. 8, pp. 110693–110719.
13. Wang X. et al. A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network. In J. Li et al. (ed.), Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. pp. 201–213.
14. White B. Sentiment Analysis: VADER or TextBlob? Medium [Online resource]. URL: <https://towardsdatascience.com/sentiment-analysis-vader-or-textblob-ff25514ac540> (Accessed 02.09.2022)

Information about the authors

Margarita A. Kirina, Master's Student, Visiting Lecturer at Department of Philology, National Research University Higher School of Economics, Saint Petersburg, Russia, ORCID: <https://orcid.org/0000-0002-7381-676X>, e-mail: mkirina@hse.ru

Ludmila D. Telnina, Bachelor's Student, National Research University Higher School of Economics, Saint Petersburg, Russia, ORCID: <https://orcid.org/0000-0003-2725-1902>, e-mail: ldtelnina@edu.hse.ru