УДК 004.89
DOI: 10.18469/ikt.2024.22.3.09

СИСТЕМА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ОПРЕДЕЛЕНИЯ ЭМОЦИОНАЛЬНОЙ ОКРАСКИ ТЕКСТА МЕТОДАМИ NLP

Ульянкин А.Е., Каратаева О.Г., Лумпова П.С., Титов А.Д.
Российский государственный аграрный университет – МСХА имени К.А. Тимирязева, Москва, РФ
E-mail: aeulianckin@rgau-msha.ru, okarataeva@rgau-msha.ru, polinalumpova@gmail.com, a.titov@rgau-msha.ru

Статья посвящена разработке системы искусственного интеллекта, предназначенной для определения эмоциональной окраски текста с использованием методов обработки естественного языка. Авторы представляют различные методы анализа эмоциональной окраски, такие как обучение с учителем и без учителя, использование классификаторов и алгоритмов машинного обучения, таких как метод опорных векторов и наивный байесовский классификатор. Особое внимание уделяется языковым показателям, которые влияют на эмоциональную окраску текста, включая экспрессивно-оценочную окраску слов, интонационные средства, синтаксическую структуру и лексико-грамматические особенности. В статье подчеркивается важность анализа эмоциональной окраски текста для классификации отзывов на маркетплейсах. Кроме того, данную систему можно использовать и для модерации контента в социальных сетях и выявления негативно-оценочных сообщений в сети Интернет.

Ключевые слова: *обработка естественного языка, машинное обучение, эмоциональная окраска текста, отзыв, модель*

Введение

В условиях стремительного роста объемов текстовой информации в цифровом пространстве [1], анализ текстов стал одной из ключевых задач в области информационных технологий. Одним из наиболее востребованных направлений текстового анализа является определение эмоциональной окраски, или тональности, текстов. Способность автоматически оценивать эмоциональные аспекты сообщений предоставляет значительные преимущества в таких сферах, как анализ пользовательских отзывов, мониторинг социальных сетей, автоматизация обслуживания клиентов и маркетинг. Тем не менее, несмотря на активное развитие технологий обработки естественного языка, проблема точного и масштабируемого анализа тональности текстов остается актуальной [2].

Анализ эмоциональной окраски текста представляет собой сложную задачу, требующую не только глубокого понимания контекста, но и учета лексических и синтаксических особенностей естественного языка. Существующие методы NLP (Natural Language Processing – обработка естественного языка), такие как машинное обучение и глубокие нейронные сети, позволяют решать эти задачи с высокой степенью точности. Однако выбор оптимального подхода к разработке системы анализа тональности, а также ее настройка для конкретных задач остаются предметом активных исследований [3; 4].

Решения для анализа тональности текста, такие как TextBlob, Natural Language Toolkit (NLTK) и Transformers, имеют свои преимущества и ограничения. TextBlob использует словарную методику, предоставляет оценку тональности (от -1 до +1) и субъективности (0–1), поддерживает многоязычность, но плохо справляется со сложными текстами. NLTK предлагает широкий набор инструментов для NLP, включая токенизацию, лемматизацию и методы машинного обучения, однако ее производительность снижается при работе с большими объемами данных, и она малоэффективна для решения сложных задач. Transformers от Hugging Face ориентированы на глубокое обучение, работают с моделями типа BERT, GPT и RoBERTa, не требуют предварительной обработки данных, но больше подходят для продвинутых задач. Модель BERT от Google учитывает контекст слов благодаря механизму внимания, что повышает точность анализа, особенно для многозначных слов, поэтому она больше подходит для классификации отзывов потребителей по эмоциональной окраске [5; 6; 7].

Сейчас на рынке существует множество готовых решений для анализа тональности. Они варьируются от полностью автоматических решений, когда достаточно просто настроить транспортировку данных в систему, до более сложных систем, которые требуют настройки и адаптации [8].

Среди готовых решений можно использовать MonkeyLearn или Google Cloud Natural Language, среди русскоязычных предложений можно рассмотреть Brand Analytics [9].

Использование готовых решений для анализа тональности может значительно упростить задачу обработки больших объемов текста, но разработка собственных моделей имеет ряд преимуществ. Первые часто универсальны, однако каждая конкретная область применения может иметь свою специфику, что и может учитывать самостоятельно разработанная модель.

При использовании готового решения существует зависимость от качества модели, которую предоставили разработчики. В случае разработки собственной модели есть возможность контролировать процесс обучения, выбирать данные для тренировки, настраивать гиперпараметры и оценивать качество работы системы на каждом этапе.

Использование сторонних сервисов для анализа тональности требует передачи данных третьим лицам, что может представлять риски с точки зрения конфиденциальности и безопасности.

Создавая собственное решение, компании получают возможность постоянно улучшать и развивать его, добавляя новые функции и возможности по мере необходимости. Это особенно важно в условиях быстро меняющегося рынка и технологий.

Анализ тональности текстов широко применяется в бизнесе, потому что компаниям необходимо знать, что происходит с клиентами и как они к этому относятся. Оценить их отношение к вам или вашему продукту можно через стандартные опросы, в которых будет замеряться уровень удовлетворенности. Однако правильно определить момент, в который клиенту нужно пройти этот опрос, бывает сложно. Помимо этого, зачастую люди игнорируют опросы, так как считают, что это ни на что не повлияет.

В то же время хорошим показателем отношения клиента являются отзывы или обращения в компанию. Они появляются в тот момент, когда это нужно: случилось что-то плохое, клиент что-то не понял, или остался настолько доволен, что хочет поделиться этим со всеми. Скорость обработки подобного взаимодействия с клиентом является очень важной, ведь от нее очень часто зависит, останется клиент с вами или нет.

Компании используют разные подходы к анализу тональности: кому-то достаточно «вручную» читать текст и проставлять оценку, кто-то может купить готовое решение, кто-то реализует свои системы.

Например, компания ОАО «Мегафон» использует инструменты NLP на базе существующих моделей трансформеров. Компания использует свой обучающий набор данных, потому что у нее узкая доменная область, и готовые наборы данных не подходят в полной мере. Компания использует анализ тональности в опросах клиентов и в комплексной оценке бренда. Одна из причин внедрения данного подхода – несоответствие выставляемой оценки пользователя комментарию, который он оставляет. Довольно часто встречаются случаи (в любом бизнесе), когда клиент ставит плохую оценку, но в комментарии пишет, что его все устраивает. Это сильно занижает метрики и при этом не отражает реальную ситуацию. К тому же, если есть текстовые данные, то на их основе проще оценивать, все ли устраивает клиента и если нет, то, что именно [10].

Таким образом, целью данного исследования была разработка системы искусственного интеллекта для автоматического определения эмоциональной окраски текстов с применением методов NLP. Система должна обеспечить возможность эффективной обработки текстов различной тематики и объема, демонстрируя высокую точность при классификации эмоциональных состояний.

Предлагаемый метод

При обработке естественного языка выполняются такие процедуры, как токенизация – разбиение текста на отдельные слова или части речи, стемминг – приведение слов к их корневой форме, лемматизация – возврат слов к исходным формам, частеречная разметка для определения структуры предложений и векторизация текста, которая переводит его в числовые данные для использования в алгоритмах машинного обучения [11; 12].

Текстовые данные, являясь категориальными и неструктурированными, требуют предварительной обработки перед обучением моделей, что помогает компьютеру распознавать смысловую связь между словами. Основные методы включают мешок слов (BoW), который анализирует частоту слов без учета их порядка, TF-IDF (Term Frequency-Inverse Document Frequency), учитывающий частоту слов и их редкость в коллекции текстов, и Word2Vec, использующий нейросети для создания векторов слов на основе больших данных.

Существующий бизнес-процесс можно описать с помощью модели AS IS («как есть») (рисунок 1). Данная модель работает в связке с TO BE, которая отражает то, как должны выглядеть бизнес-процессы.

Как следует из рисунка 1, действующая система аналитики позволяет изучать текущее состояние компании только после обновления данных в базе: они интегрируются по факту обновлений в базе, чтобы не нагружать локальную сеть и видеть данные в реальном времени. Обновле-

ние в базе происходит минимум через 3 дня после выгрузки отзывов из нее. То есть при таком бизнес-процессе компания как минимум отстает на 3 дня. Помимо этого, сотрудник тратит время на вычитку и разметку данных, хотя мог бы только изучать текущие показатели и заниматься генерацией и проверкой гипотез для улучшения бизнеса. Ситуацию можно улучшить, добавив к текущим действующим единицам (сотрудник, система аналитики и база данных) еще одну в виде ML-системы (рисунок 2).

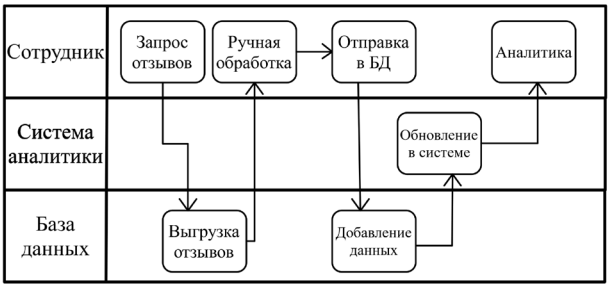

Рисунок 1. Модель AS IS

Общая идея всей системы разбита на 2 блока:

1. Модуль обучения модели. Это блок, в котором будет происходить обучение и подгонка модели для получения наибольшей точности. Предполагается, что обучение нужно будет проводить раз в 2-3 месяца, поэтому исключить этот процесс нельзя, а добавлять в основной функционал нет смысла. На вход она будет получать JSON-файл со следующими полями (таблица 1).

2. Модуль классификации. В этом блоке будет происходить основная часть, указанная ранее на модели TO BE (рисунок 2). Необходимо будет написать запрос к базе данных, чтобы получить отзывы, затем привести их в нужный для модели вид, провести классификацию по тональности и отправить результаты в базу данных.

Таблица 1. Описание полей для обучающего датасета

Поля	Описание	Формат
TEXT	Текст отзыва или обращения	Текстовый
ENTITY	Сущность, главный объект оценки	Текстовый
SUBENTITY	Подсущность, если есть то, что относится к сущности и может быть выделено отдельно	Текстовый
TYPE	Тональность: 1-–очень негативная; 2 – негативная; 3 – нейтральная; 4 – положительная; 5 – очень положительная	Текстовый

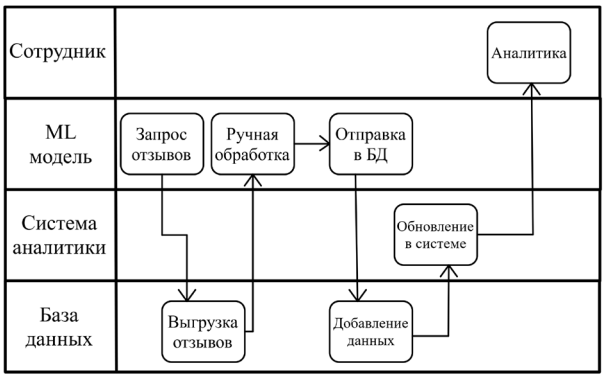

Рисунок 2. Модель TO BE

Программная реализация

Парсинг отзывов на товары был осуществлен на торговой площадке Озон. Весь процесс реализации системы происходил на Apple Macbook Pro 16 (M1 Max, RAM 32 Gb, 1Tb SSD).

Система разработана с учетом модульной архитектуры, где обучение модели и предсказание разделены на независимые процессы. Это позволяет гибко использовать обученную модель для решения задач бизнеса, не затрачивая ресурсы на повторное обучение.

Рассмотрим первый блок – модуль обучения. В нем будут происходить основные процессы:

– загрузка в модель обученных данных. Она реализуется через чтение JSON-файла из определенной папки. При чтении необходимо каждое поле записывать в отдельный список. Такой подход помогает модели в обучении, так как JSON-файл будет содержать внутренние списки, а также в дальнейшем позволит быстро обращаться к различным данным, которые связаны с текстом;

– предобработка данных для BERT (Bidirectional Encoder Representations from Transformers). Здесь будет происходить удаление стоп-слов, то есть слов, которые не несут особого смысла: предлоги, союзы, частицы и т.п. Это происходит за счет функций библиотеки NLTK. Затем с помощью встроенных инструментов BERT запускается процесс токенизации текста. После обрабатываются аннотации и списки преобразуются в тензоры;

– обучение, вывод результатов точности и сохранение модели.

Для начала необходимо загрузить все библиотеки для работы: для чтения файлов, для подключения к базе данных, для работы с фреймами данных и для обучения модели.

Сначала будет происходить загрузка размеченного набора данных. Так как обучение модели очень важный процесс, при загрузке настроены самые основные проверки:

– на наличие файла с набором данных;

– на правильный формат файла с данными;
– на корректное считывание JSON-файла;
– на наличие в файле необходимых данных.

Это сделано для того, чтобы при наличии проблем, они были сразу обозначены.

Далее идет предобработка данных перед обучением. Так как BERT мощная система, то такой строгой обработки, как при использовании нейронных сетей или алгоритмов машинного обучения, делать не нужно. BERT работает с текстом, в его исходном виде. Но как показала практика, небольшую обработку перед загрузкой делать все же нужно: так результаты получаются лучше.

Токенизация нужна для приведения текста к фиксированной длине через паддинг, то есть с отступами, или, наоборот, усечением текста. Также тут происходит обработка сущностей.

Следующий важный этап – это обучение модели. Здесь реализуется класс SentimentDataset, нужный для обработки JSON-файла и генерации входных данных для BERT. Для обучения будет использоваться метод из Fine-Tuning, BertForSequenceClassification. Он подразумевает, что BERT уже предобучен и имеет свои веса. Там же после обучения идет указание метрик точности. Это нужно для того, чтобы получить возможность сразу оценивать, стоит ли дальше обучать модель или нужно сменить параметры, либо подобрать другой набор данных. Мы дообучаем модель на своих данных, тем самым изменяя веса в нужную сторону, что позволяет лучше учитывать семантику.

После того, как модель обучилась и мы получили достаточные результаты ее точности, модель необходимо сохранить. Затем мы снова обращаемся к ней, чтобы оценить ее точность на тренировочном наборе. По факту, был один набор данных , но мы разделили его на обучающую и тестовую выборки в соотношении 80% на 20%, оценивали точность на тестовом наборе данных, после чего проводилась кросс-валидация, также с разделением набора в соотношении 80% на 20%.

Также необходимо оценить правильность определения сущностей, так как именно от них зависит тональность. Для оценки тональностей помимо стандартных оценок точности (presicion):

$$Precision = \frac{TP}{TP + FP},$$

где *TP*, *FP* берутся из матрицы (таблица 2) полноты (recall):

$$Recall = \frac{TP}{TP + FN},$$

где *TP*, *FN* берутся из матрицы (таблица 2) и *F*1-меры:

$$F1score = 2 \cdot \frac{\mathrm{Pr}ecision \cdot \mathrm{Re}call}{\mathrm{Pr}ecision + \mathrm{Re}call},$$

идет оценка по каждой тональности. Это позволяет понять, где модель чаще ошибается, и добавить в ее набор данных тексты именно такой тональности [13].

Таблица 2. Матрица ошибок

		Предсказанные значения	
		Positive	Negative
Фактические значения	Positive	TP	FN
	Negative	FP	TN

Точность модели на тренировочном наборе представлена в таблице 3.

Видно, что полнота повышается в зависимости от тональности: сложнее определять негатив. Точность модели еще не достигает желаемых 95%, однако ближе, чем другие рассматриваемые модели.

Для повышения точности возможно стоит добавить некоторый объем данных, причем с негативной тональностью. Датасет для обучения содержал в одинаковой пропорции разные тональности, с общим объемом свыше 50 тысяч размеченных текстов. Но, возможно, следует несколько изменить распределение. Также можно испробовать разные параметры модели: эпохи, скорость обучения. Обучение заняло чуть больше двух дней.

Далее следует второй блок, реализующий все действия, которые были в модели TO BE, описанной ранее.

Затем настраивается подключение к базе данных для выгрузки исходных данных и загрузки результатов.

На следующем этапе необходимо определить функцию для выгрузки данных, предсказания тональности и сущностей и сохранения полученного результата в нужную таблицу. В завершении можно сформировать отчетность за выбранный период.

Заключение

Разработанная система позволяет:

1. Избавить сотрудника от работы по разметке, выгрузке и загрузке данных в базу. При этом наблюдается ускорение процессов выгрузки, загрузки данных и определения тональности. До этого выгрузка и загрузка происходили через excel-файл, что занимало больше времени. В итоге время на выгрузку и загрузку данных из базы сократилось с 18 минут 30 секунд, то есть на 97,2%.

2. Получить более точную и быструю систему анализа тональности. Модель справляется за 15 минут, в то время как сотруднику требуется в со-

Таблица 3. Результаты расчета точности модели

Type	Precision	Recall	F1-score	Кросс-валидация
Очень негативные	92%	87%	89%	90%
Негативные	90%	88%	89%	89%
Нейтральные	94%	91%	92%	90%
Позитивные	95%	93%	94%	91%
Очень позитивные	96%	96%	96%	92%

вокупности около 6 часов (экономия времени составляет 95,8%), а с перерывами – 3 дня. При этом точность модели при апробации на новых (за отработанные 2 недели) оказалась даже чуть выше, чем на тренировочном наборе: точность – 93%, полнота – 94%, F1-мера – 93%.

3. Создать более узкие категории по тональности пользователей. Помимо предыдущих задач, которые показали лучший результат по точности и времени, прибавился еще один показатель. Человек оставляет свое мнение по причине произошедшего события-триггера. Однако категории, составленные по таким событиям, не всегда описывают то, что именно не понравилось или понравилось человеку. Реализованная система решает эту проблему: не имеет значения, после какого события человек написал. В самом тексте можно уже увидеть то, что ему нравится или нет.

Но есть также проблемы, в основном, все это связано с особенностями языка:

1. Сарказм. Система стабильно оценивает около 50% текстов с сарказмом неверно. На текущий момент это не так критично, потому что таких текстов менее 1% от всего объема, поэтому общая точность не снижается. Для решения данной проблемы выдвинута идея использовать исторический портрет пользователя. Если клиент склонен к подобному формату общения, то целесообразно будет чуть смещать веса в сторону искажения положительного и негативного контекстов.

2. Определение сущностей. Иногда люди пишут о том, что они оценивают, не указывая точно, что это такое. Например, человек оставляет отзыв на сайте магазина под товаром – телефоном – и пишет «Шустро!». В этом случае непонятно, что является объектом комментирования: пользователь оценил быстродействие телефона или скорость доставки? Для решения данной проблемы выдвинута идея о связи дополнительных факторов с сущностями. Но это будет нагружать систему лишними параметрами, что повлечет снижение точности и скорости работы.

Таким образом, авторами были разработаны методические подходы к автоматизации классификации тональности отзывов на маркетплейсах, что позволило значительно сократить затраты времени на аналитику, а также исключить человеческий фактор и снизить ошибки. Кроме того, данную систему можно использовать и для модерации контента в социальных сетях и выявления негативно-оценочных сообщений в сети Интернет.

Литература

1. Скляр М.А., Кудрявцева К.В. Цифровизация: основные направления, преимущества и риски // Экономическое возрождение России. 2019. № 3 (61). С. 103–114.
2. Ермаков С.А., Ермакова Л.М. Методы оценки эмоциональной окраски текста // Вестник Пермского университета. Математика. Механика. Информатика. 2012. № 1 (9). С. 85–90.
3. Богданова Д.Р., Рахимов А.М. Обзор методов оценки эмоциональной окраски текстов // Моделирование систем и процессов. 2021. Т. 14, № 3. С. 11–16.
4. Рубцова Ю.В. Построение корпуса текстов для настройки тонового классификатора // Программные продукты и системы. 2015. № 1. С. 72–78.
5. Маньков А.С., Детков А.А. Анализ тональности текста. Обзорная статья // Весенние дни науки: материалы Международной конференции студентов и молодых ученых. Екатеринбург: ООО Издательский Дом «Ажур», 2023. С. 66–68.
6. Adaskina Y.V., Panicheva P.V., Popov A.M. Syntax-based sentiment analysis of tweets in Russian // Компьютерная лингвистика и интеллектуальные технологии: материалы ежегодной Международной конференции «Диалог». М.: РГГУ, 2015. Т. 2. С. 1–11. URL: https://dialogue-conf.org/media/2778/dialogue-2015_vol2.pdf (дата обращения: 01.11.2024).
7. Баймырадов Г., Шохрадова Дж. Обработка естественного языка (NLP): моделирование языка, анализ текста и системы голосового реагирования // Научный электронный журнал

«Матрица научного познания». 2024. № 10-2. С. 20–23.
8. Семина Т.А. Анализ тональности текста: современные подходы и существующие проблемы // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 6: Языкознание. 2020. № 4. С. 47–63.
9. Павлов Ю.Н., Майструк К.А. Сравнение методов оценки тональности текста // Молодой ученый. 2016. № 12 (116). С. 59–64.
10. МегаФон внедрил искусственный интеллект в работу с персоналом. URL: https://www.cnews.ru/news/line/2024-08-06_megafon_vnedril_iskusstvennyj (дата обращения: 01.11.2024).
11. Боярский К.К. Введение в компьютерную лингвистику: учебное пособие. СПб.: Университет ИТМО, 2013. 73 с.
12. Гречачин В.А. К вопросу о токенизации текста // Международный научно-исследовательский журнал. 2016. № 6-4 (48). С. 25–27.
13. Сметанин С.И. Анализ тональности текстов из социальных сетей на основе методов машинного обучения для мониторинга общественных настроений: дисс. … канд. техн. наук. М., 2022. 162 с.

Дата поступления 13.11.2024
Дата принятия 16.12.2024

Ульянкин Александр Евгеньевич, ассистент кафедры статистики и кибернетики Российского государственного аграрного университета – МСХА имени К.А. Тимирязева (РГАУ-МСХА). 127422, Российская Федерация, г. Москва, ул. Тимирязевская, 49. Тел. +7 499 976-12-53. E-mail: aeulianckin@rgau-msha.ru

Каратаева Оксана Григорьевна, к.э.н., доцент кафедры педагогики и психологии профессионального образования РГАУ-МСХА. 127422, Российская Федерация, г. Москва, ул. Тимирязевская, 49. Тел. +7 499 976-12-53. E-mail: okarataeva@rgau-msha.ru

Лумпова Полина Сергеевна, магистрант кафедры статистики и кибернетики РГАУ-МСХА. 127422, Российская Федерация, г. Москва, ул. Тимирязевская, 49. Тел. +7 499 976-12-53. E-mail: polinalumpova@gmail.com

Титов Артем Денисович, ассистент кафедры статистики и кибернетики РГАУ-МСХА. 127422, Российская Федерация, г. Москва, ул. Тимирязевская, 49. Тел. +7 499 976-12-53. E-mail: a.titov@rgau-msha.ru

UDC 004.89
DOI: 10.18469/ikt.2024.22.3.09

AN ARTIFICIAL INTELLIGENCE SYSTEM FOR DETERMINING THE EMOTIONAL COLORING OF TEXT USING NLP METHODS

Ulianckin A.E., Karataeva O.G., Lumpova P.S., Titov A.D.
Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, Moscow, Russian Federation
E-mail: aeulianckin@rgau-msha.ru, okarataeva@rgau-msha.ru, polinalumpova@gmail.com, a.titov@rgau-msha.ru

The article is devoted to the development of an artificial intelligence system designed to determine the emotional context of a text using natural language processing methods. The authors present various methods of analyzing emotional context, such as teaching with and without participation of the teacher, using classifiers and machine learning algorithms such as SVM and Naive Bayes. Special attention is paid to linguistic indicators that affect the emotional side of the text, including expressive and evaluative coloring of words, intonation means, syntactic structure and lexico-grammatical features. The article emphasizes the importance of analyzing the emotional context of the text in order to classify reviews on marketplaces. In addition, this system can also be used to moderate content on social networks and identify messages with negative evaluation on the Internet.

Keywords: *natural language processing, machine learning, emotional coloring of text, feedback, model*

Ulianckin Alexander Evgeniyevich, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, 49, Timiryazevskaya Street, Moscow, 127422, Russian Federation; Assistant of Statistics and Cybernetics Department. Tel. +7 499 976-12-53. E-mail: aeulianckin@rgau-msha.ru

Karataeva Oksana Grigoriyevna, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, 49, Timiryazevskaya Street, Moscow, 127422, Russian Federation; Associate Professor of Pedagogy and Psychology Department, PhD in Economics. Tel. +7 499 976-12-53. E-mail: okarataeva@rgau-msha.ru

Lumpova Polina Sergeyevna, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, 49, Timiryazevskaya Street, Moscow, 127422, Russian Federation; Master's Degree Student of Statistics and Cybernetics Department. Tel. +7 499 976-12-53. E-mail: polinalumpova@gmail.com

Titov Artyom Denisovich, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, 49, Timiryazevskaya Street, Moscow, 127422, Russian Federation; Assistant of Statistics and Cybernetics Department. Tel. +7 499 976-12-53. E-mail: a.titov@rgau-msha.ru

References

1. Sklyar M.A., Kudryavtseva K.V. Digitization: trends, benefits and risks. *Ekonomicheskoe vozrozhdenie Rossii*, 2019, no. 3 (61), pp. 103–114. (In Russ.)
2. Ermakov S.A., Ermakova L.M. Overview of sentiment analysis methods. *Vestnik Permskogo universiteta. Matematika. Mekhanika. Informatika,* 2012, no. 1 (9), pp. 85–90. (In Russ.)
3. Bogdanova D.R., Rakhimov A.M. Review of methods for assessing the emotional coloring of texts. *Modelirovanie sistem i processov*, vol. 14, no. 3, pp. 11–16. (In Russ.)
4. Rubtsova Yu.V. Building a corpus of texts for tuning a tone classifier. *Programmnye produkty i sistemy*, 2015, no. 1, pp. 72–78. (In Russ.)
5. Mankov S.A., Detkov A.A. Sentiment analysis of text. Review article. *Vesennie dni nauki: materialy Mezhdunarodnoj konferencii studentov i molodyh uchenyh*. Ekaterinburg: OOO Izdatel'skij Dom «Azhur», 2023, pp. 66–68. (In Russ.)
6. Adaskina Y.V., Panicheva P.V., Popov A.M. Syntax-based sentiment analysis of tweets in Russian. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: materialy ezhegodnoj Mezhdunarodnoj konferencii «Dialog».* Moscow: RGGU, 2015, vol. 2, pp. 1–11. URL: https://dialogue-conf.org/media/2778/dialogue-2015_vol2.pdf (accessed: 01.11.2024).
7. Baymyradov G., Shohradova J. Natural language Processing (NLP): language modeling, text analysis, and voice response systems. *Nauchnyj elektronnyj zhurnal «Matrica nauchnogo poznaniya»*, 2024, no. 10-2, pp. 20–23. (In Russ.)
8. Semina T.A. Sentiment analysis: modern approaches and existing problems. Social'nye i gumanitarnye nauki. *Otechestvennaya i zarubezhnaya literatura. Seriya 6: YAzykoznanie*, 2020, no. 4, pp. 47–63. (In Russ.)
9. Pavlov Yu.N., Maistruk K.A. Comparison of methods for assessing the tonality of a text. *Molodoj uchenyj*, 2016, no. 12 (116), pp. 59–64. (In Russ.)
10. MegaFon has implemented artificial intelligence in its work with staff. URL: https://www.cnews.ru/news/line/2024-08-06_megafon_vnedril_iskusstvennyj (accessed: 01.11.2024). (In Russ.)
11. Boyarsky K.K. *Introduction to computational linguistics*: Textbook. Saint Petersburg: Universitet ITMO, 2013, 73 p. (In Russ.)
12. Grechachin V.A The issue of text tokenization. *Mezhdunarodnyj nauchno-issledovatel'skij zhurnal*, 2016, no. 6-4 (48), pp. 25–27. (In Russ.)
13. Smetanin S.I. Analysis of the tonality of texts from social networks based on machine learning methods for monitoring public sentiment: diss. … cand. techn. nauk. Moscow, 2022. 162 p. (In Russ.)

Received 13.11.2024
Accepted 16.12.2024