УДК 004.934.2
doi:10.31799/1684-8853-2020-4-20-30

Обзоры
Surveys

Аналитический обзор подходов к распознаванию тональности русскоязычных текстовых данных

А. А. Двойникова[a], *программист, orcid.org/0000-0001-8047-6639*
А. А. Карпов[a], *доктор техн. наук, доцент, orcid.org/0000-0003-3424-652X, karpov@iias.spb.su*
[a]*Санкт-Петербургский институт информатики и автоматизации РАН, 14-я линия В. О., 39, Санкт-Петербург, 199178, РФ*

Введение: *в последние годы анализ тональности, или сентимент-анализ, высказываний пользователей находит практическое применение во многих областях: оценка качества товаров и услуг по отзывам покупателей в Интернете, анализ негативных эмоций в сообщениях, прогноз фондовых рынков, политических ситуаций на основе новостных лент и многих других. В связи с этим разрабатываются разнообразные системы и методы для сентимент-анализа русскоязычных текстовых данных.* ***Цель:*** *выполнение подробного обзора подходов и сравнительного анализа существующих баз данных в области сентимент-анализа текстов на русском языке.* ***Результаты:*** *аналитический обзор подходов к анализу тональности русскоязычных текстовых данных показал, что для сентимент-анализа текстов сейчас имеется множество разнообразных методов предобработки текстовых данных, их векторизации и машинной классификации. Из сравнительного анализа существующих баз данных по данной тематике можно сделать вывод, что автоматический сентимент-анализ русскоязычных текстов развит значительно меньше, чем для других основных мировых языков. Исследование программных систем для анализа текстов на русском языке демонстрирует, что пока русскоязычный анализ тональности показывает относительно низкую точность по сравнению с англоязычным, одной из причин этого может являться сложная структура русского языка. В статье описываются основные нерешенные проблемы анализа тональности русскоязычных текстов.* ***Обсуждение:*** *в дальнейших исследованиях планируется реализовать сентимент-анализ разговорной речи дикторов с использованием аудиоданных, для чего необходимо сначала получить орфографическую транскрипцию речи для каждого диктора.*

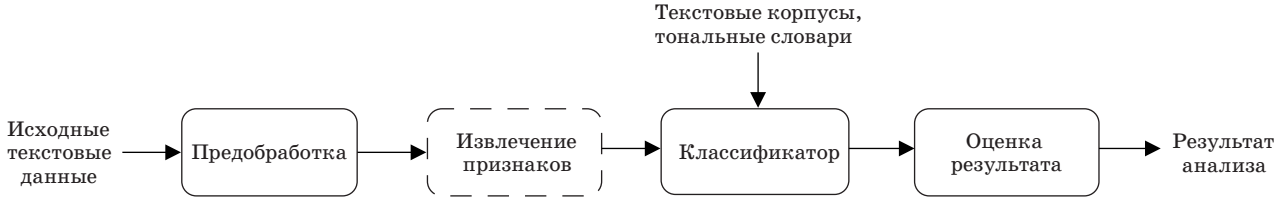
Ключевые слова *— тональность текстовых данных, векторизация текста, сентимент-анализ, компьютерная паралингвистика.*

Для цитирования: Двойникова А. А., Карпов А. А. Аналитический обзор подходов к распознаванию тональности русскоязычных текстовых данных. *Информационно-управляющие системы*, 2020, № 4, с. 20–30. doi:10.31799/1684-8853-2020-4-20-30
For citation: Dvoynikova A. A., Karpov A. A. Analytical review of approaches to Russian text sentiment recognition. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 4, pp. 20–30 (In Russian). doi:10.31799/1684-8853-2020-4-20-30

Введение

Анализ тональности текста, или сентимент-анализ (sentiment analysis), — область компьютерной лингвистики и интеллектуального анализа текста, ориентированная на извлечение из него субъективных мнений и эмоций человека. Анализ тональности находит практическое применение во многих областях: оценка качества товаров и услуг по отзывам покупателей в Интернете, анализ негативных эмоций в сообщениях, прогноз фондовых рынков, политических ситуаций на основе новостных лент [1]. Также сентимент-анализ необходим в автоматизированных системах, в которых человек общается с машиной на естественном языке. Чтобы проанализировать такой объем информации, в последние годы были предложены многочисленные методы автоматического сентимент-анализа, которые рассмотрены в данной статье.

Анализ тональности текстов происходит в несколько этапов (рис. 1). На первом этапе выполняется предобработка исходного текста, далее извлекаются информативные признаки (векто-



■ ***Рис. 1.*** Этапы анализа тональности текста
■ ***Fig. 1.*** The stages of sentiment analysis of text

ризация текста), на их основе строится классификатор (распознаватель) тональности, и последним этапом является оценка результата работы. Этап векторизации текста для лингвистических методов классификации не является обязательным, так как такие классификаторы работают непосредственно с текстами, а не с их векторами.

Предобработка текста

Предобработка текста является первым этапом в его анализе. Она необходима для того, чтобы выделить из «зашумленного» текста релевантную информацию. Предобработка текста включает в себя приведение всех слов к единому регистру, удаление знаков пунктуации, удаление стоп-слов, токенизацию, нормализацию слов и при необходимости иные операции.

При приведении всех слов к единому регистру, как правило, все прописные символы преобразуются в их строчные формы, поскольку предполагается, что прописные или строчные формы слов не имеют различий. Во всех текстах присутствуют знаки пунктуации, выполняющие чаще всего синтаксическую функцию, поэтому при анализе эмоций в тексте нет необходимости сохранять их. Также при обработке текста удаляются стоп-слова — слова, не содержащие смысловой нагрузки, например предлоги, союзы, частицы и т. п. Необходимым этапом предобработки для последующего компьютерного анализа текста является токенизация слов — разбиение текста на отдельные значимые единицы (токены) [2]. Самый простой способ токенизировать русскоязычный текст — разделить его на слова по пробелам. Парадигмы слов в русском языке имеют большое количество словоформ, передающих одинаковый смысл. Форма слова не всегда несет в себе полезную информацию, поэтому при анализе текста рекомендуется производить нормализацию всех слов, т. е. представление слова в его начальной форме. Нормализация может осуществляться двумя способами: лемматизацией и стеммингом. *Лемматизация* — преобразование слова к его начальной форме (лемме). Лемматизация основана на морфологическом словаре. Если слово не присутствует в словаре, то строится гипотеза о способах изменения слова и получения для него леммы. *Стемминг* — получение основы слова, при этом у слов отбрасываются окончания, суффиксы, приставки. Тем самым все слова в тексте приводятся к единой форме. Стемминг основан на морфологических правилах и не требует наличия словаря.

Каждый из этапов предобработки текста позволяет снизить размерность признакового пространства. В зависимости от исходного текста предобработка может включать в себя только несколько операций, а каждая операция может дорабатываться вручную с учетом всех исключений.

Извлечение признаков из текста

Перед тем как использовать машинный классификатор, необходимо представить текст в числовом виде (признаковое описание), т. е. векторизовать текст. Рассмотрим несколько современных способов векторизации текста.

BoW (Bag of Words — «мешок слов») — метод, представляющий текст в виде неупорядоченного набора слов [3]. Каждому слову присваивается свой вес, часто используются веса TF-IDF, отражающие отношение частоты слова в документе к частоте слова во всех документах.

One-hot encoding (прямое кодирование) — метод, преобразующий слова в бинарные векторы [4]. Размер каждого вектора равен объему всех слов в тексте. Перед кодированием все слова, присутствующие в тексте, располагаются по алфавиту.

SVD (Singular Value Decomposition) — метод, преобразующий текст в разреженную матрицу $\mathbf{A}_{n\times m} = \{\mathbf{a}_{i1}, \mathbf{a}_{i2}, \mathbf{a}_{i3}, ..., \mathbf{a}_{im}\}$, где $\mathbf{a}_{ij}$ — взвешенный вектор-столбец частоты терминов предложения i в рассматриваемом документе [5]. Если в документе содержится всего n терминов и m предложений, то на выходе будет матрица размерностью $n \times m$.

Word2Vec (инструментарий, разработанный компанией Google) — нейронная сеть, которая генерирует векторы слов [6]. Она обучена на двух алгоритмах: BoW (предсказывает слово с учетом контекста) и Skip-gram (предсказывает контекст с учетом слова). Word2Vec сначала строит словарь из обучающего текстового корпуса и анализирует векторные представления каждого слова. Кроме того, Word2Vec имеет возможность рассчитать косинусное расстояние между словами.

Glove — метод, разработанный в Стэнфордском университете (США) [7]. В его основе лежит способ подсчета частоты появления слов в текстовом корпусе. Фактически он состоит из двух основных этапов: на первом происходит построение матрицы смежности из обучающего корпуса, а на втором — факторизация матрицы для получения векторов.

FastText — метод, преобразующий в векторы не только слова, но и символьные n-граммы, из которых составлены слова [8].

BERT (Bidirectional Encoder Representations from Transformers) — нейронная сеть, разработанная компанией Google [9]. BERT обучали на

корпусе текстов из Wikipedia и сборнике книг BookCorpus. Идея векторизации в BERT заключается в том, что каждому слову из текста присваивается число, обозначающее порядковый номер слова в словаре, далее это число преобразуется в вектор из 512 символов. Словарь, который использует данная нейросеть, построен таким образом, что слова, близкие по смыслу, располагаются рядом. Тем самым нейронная сеть BERT векторизует текст, учитывая близость слов. Существуют также модификации BERT, например DistilBERT [10]. Это более легкая и быстрая версия BERT, которая примерно соответствует его производительности. Авторы работы [11] показали, что перевод обучения с многоязычной модели BERT на одноязычную модель для русского языка приводит к значительному росту производительности при выполнении анализа эмоций в тексте.

ELMo (Embeddings from Language Models) — нейронная сеть, которая генерирует контекстное представление слов [12]. Модели ELMo обучены на корпусе объемом 1 млрд слов, собранных из новостных лент сети Walmart. Для обучения ELMo применяется минимальная предобработка текстовых данных, выполняется только токенизация и лемматизация. ELMo позволяет векторизовать слова, учитывая контекст до и после этого слова. Идея модели состоит в том, чтобы сначала построить для каждого слова в тексте посимвольный эмбеддинг (embedding) слова, а потом для них применить нейросеть LSTM (Long Short-Term Memory) таким образом, что получатся эмбеддинги, учитывающие контекст, в котором встретилось слово.

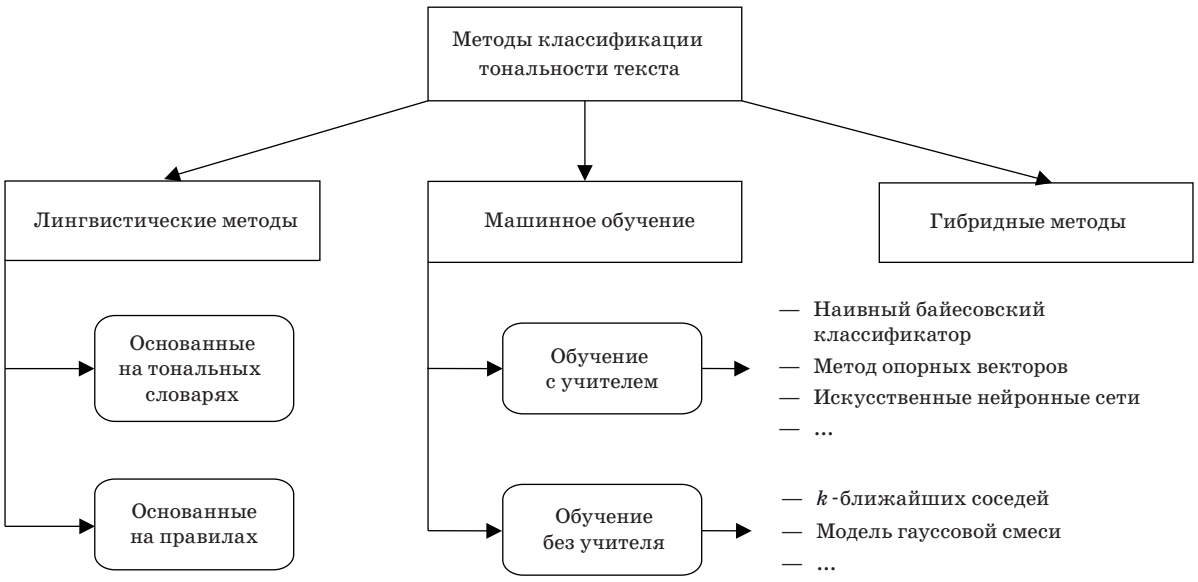
Классификация тональности текстовых данных

В настоящее время существует несколько основных методов для определения (классификации) тональности текста [13]. Все их можно разделить на несколько основных типов (рис. 2), в том числе лингвистические методы, методы на основе машинного обучения и гибридные методы. Рассмотрим все эти методы более подробно.

Первый лингвистический метод основан на *тональных словарях*. Тональный словарь представляет собой набор слов или биграмм, которым задается определенный вес принадлежности к позитивному или негативному классу. При анализе текста каждое слово ищется в этом словаре, и его вес записывается. Если слова нет в словаре, то его класс считается нейтральным, и вес равняется нулю. После того как все веса получены, высчитывается принадлежность данного текста к определенному классу тональности. Данный метод был использован для сентимент-анализа в работе [14].

Второй лингвистический метод основан на *правилах*. Для работы этого метода необходим большой набор продукционных правил конструкции «если –> то». Этот метод также подразумевает использование тональных словарей, в которых слова принадлежат определенному классу. Задача сентимент-анализа решается при помощи метода, основанного на правилах, например в работе [15].

Методы на основе машинного обучения подразделяются на обучение *с учителем (*supervised learning) и *без учителя* (unsupervised learning). Метод обучения с учителем основан на том,



■ ***Рис. 2.*** Систематизация методов классификации тональности текста
■ ***Fig. 2.*** Systematization of methods of classifying the tonal of text

чтобы обучить классификатор на заранее размеченных обучающих текстовых данных [16]. Наиболее распространенные методы в области тонального анализа: наивный байесовский классификатор, метод опорных векторов, логистическая регрессия и искусственные нейронные сети такие, как сверточные (Convolutional Neural Network — CNN), рекуррентные нейронные сети (Recurrent Neural Network — RNN), нейросетевые модели с длинной краткосрочной памятью (Long Short-Term Memory — LSTM) и управляемым рекуррентным блоком (Gated Recurrent Unit — GRU). Авторы статьи [17] сравнивают работу различных традиционных методов обучения с учителем при анализе тональности текстов в социальных сетях. В статье [18] использовались различные нейронные сети для анализа негативных текстовых сообщений. В отличие от метода обучения с учителем, метод обучения без учителя определяет взаимосвязь и закономерности между объектами без размеченных обучающих данных [19]. К таким методам можно отнести модель гауссовой смеси и k-ближайших соседей.

Существуют также гибридные методы, объединяющие в себе несколько различных методов. В работе [20] для задачи классификации текста использовался гибридный метод, объединяющий тональные словари и метод опорных векторов. В статье [21] авторы для решения задачи сентимент-анализа объединяли CNN и k-ближайших соседей.

После этапа классификации сентимент-анализа текстов следует количественная оценка результатов, которая может быть проведена с использованием набора следующих статистических показателей: точности (accuracy или precision), полноты (recall) и F-меры (F-score) [13].

Корпусы для анализа тональности текстов

Несмотря на актуальность анализа тональности русскоязычных текстов, количество аннотированных корпусов для русского языка невелико [22]. На начало 2020 г. в свободном доступе нам удалось найти четыре тональных словаря и семь текстовых корпусов, предназначенных для задачи сентимент-анализа русскоязычных текстов.

Русскоязычные тональные словари в свободном доступе

При использовании метода, основанного на тональных словарях, для автоматической классификации текстов необходимо опираться на словарь, в котором содержатся слова с разметкой принадлежности их к определенному сентименту. Разметка может быть бинарная (2 класса), тернарная (3 класса) и многоклассовая (больше трех классов). Известны несколько тональных словарей для русского языка. Базовые сведения об этих словарях, включая количество содержащихся в них слов, а также количество рассматриваемых классов, представлены в табл. 1.

Тональный словарь RuSentiLex [23] может содержать как отдельные слова, так и словосочетания, для которых указаны их характеристики, обозначающие часть речи или синтаксический тип группы, их лемматизированную форму, тональность, источник информации. В зависимости от контекста одно и то же слово может принимать разное значение тональности. Поэтому авторы словаря ввели отдельный класс тональности, обозначающий смешанную оценку слова. Также авторы отчасти решили проблему со словами, имеющими несколько значений. Они перечисляют все значения слова по тезаурусу РуТез [24] и дают отсылку на соответствующее понятие,

■ ***Таблица 1***. Тональные словари для русскоязычного анализа тональности текста
■ ***Table 1***. Tonal dictionaries for Russian-language text tonality analysis

Название словаря (ссылка для доступа)	Число слов	Число классов	Классы с количеством слов
RuSentiLex (https://www.labinform.ru/pub/rusentilex/index.htm)	16 057	4	Положительные (3785), отрицательные (10 234), нейтральные (1747), смешанная оценка (291)
LinisCrowd (http://linis-crowd.org/)	7545	5	Сильно отрицательные (228), отрицательные (1598), нейтральные (4864), положительные (806), сильно положительные (49)
WordNetAffect (http://lilu.fcim.utm.md/resourcesRoRuWNA_ru.html)	2401	6	Радость (749), страх (617), гнев (398), печаль (445), отвращение (74), удивление (118)
Словарь Белякова [27]	690	2	Положительные (300), отрицательные (390)

имя понятия прописывают в кавычках. В таких случаях каждому значению слова присваивается свое значение тональности.

LinisCrowd [25] — тональный словарь на основе пользовательского интернет-контента социально-политической тематики. Изначально словарь составлялся по размеченным текстам, полученным из социальной сети Facebook. Впоследствии словарь расширился за счет добавления к нему других словоформ, а также слов из других словарей.

WordNetAffect [26] является лексическим ресурсом, который содержит слова, описывающие эмоции. Он был создан на базе онтологии WordNet — семантического лексикона английского языка — путем выбора и разметки наборов синонимов (синсетов) эмоциональными концепциями. Наборы синонимов были вручную размечены эмоциональными метками, далее они были дополнительно переразмечены на шесть эмоциональных категорий. Для русского языка авторы словаря вручную перевели синсеты WordNetAffect с английского языка.

Тональный словарь из работы Белякова [27] содержит 690 основ эмоциональных слов. Словарь разбит на два класса: основы русскоязычных слов с положительной и отрицательной эмоциональной окраской.

Русскоязычные эмоционально окрашенные текстовые корпусы в свободном доступе

Существует несколько эмоционально окрашенных текстовых корпусов для русского языка, их основные характеристики представлены в табл. 2.

Крупнейшая российская конференция по компьютерной лингвистике «Диалог» ежегодно проводит соревнования по компьютерному анализу русского языка (http://www.dialog-21.ru/evaluation/), одним из основных направлений соревнований является анализ тональности текстов. Так, в 2015 и 2016 гг. организаторы предоставили текстовые корпусы SentiRuEval. SentiRuEval-2015 [28] содержит в себе отзывы, собранные из сети Twitter, о ресторанах и автомобилях. Помимо общей тональности отзыва, SentiRuEval-2015 содержит различные целевые аспекты оцениваемого объекта. Каждый из этих аспектов также может иметь тональную оценку. SentiRuEval-2016 [29] включает отзывы о банках и мобильных операторах, собранные из Twitter. Разметка отзывов показывает объект отзыва и отношение субъекта к этому объекту.

LinisCrowd [25] — коллекция документов, посвященных социально-политической тематике. В качестве источника данных использовались записи блог-платформы «Живой Журнал». RuSentiment [22] — текстовый корпус, имеет в своем составе посты, собранные из социальной сети «ВКонтакте», на разные тематики. Некоторые посты могут не иметь разметку по тональности, но они могут относиться к определенному классу высказывания (шаблонные приветствия, благодарственные и поздравительные сообщения). RuTweetCorp [30] — корпус русскоязычных twitter-постов, автоматически размеченных на два класса. Корпусы РОМИП 2012 [31] и Auto_reviews [32] также находятся в свободном доступе.

■ ***Таблица 2***. Текстовые корпусы для исследований русскоязычного сентимент-анализа
■ ***Table 2***. Text corpora for research of Russian-language sentimental analysis

Название корпуса (ссылка для доступа)	Тематика текстов	Число фраз	Число классов
RuTweetCorp (https://study.mokoron.com/#download)	Широкая	226 914	2
РОМИП 2012 (http://romip.ru/ru/2012/tracks.html)	Книги, фильмы и фотокамеры	50 247	2, 3, 5
RuSentiment (https://github.com/strawberrypie/rusentiment)	Широкая	31 185	3
LinisCrowd (http://linis-crowd.org/)	Социально-политическая	26 873	5
SentiRuEval-2016 (http://www.dialog-21.ru/evaluation/2016/sentiment/)	Банки и мобильные операторы	23 595	3
SentiRuEval-2015 (http://www.dialog-21.ru/evaluation/2015/sentiment/)	Рестораны и автомобили	17 628	4
Auto_reviews (https://github.com/oldaandozerskaya/auto_reviews)	Автомобили	6152	5

Сентимент-анализ может также применяться при анализе разговорной речи дикторов. Для решения такой задачи можно использовать мультимодальный корпус RAMAS [33]. Он содержит около семи часов аудио- и видеозаписей интерактивных диалогов, разыгранных несколькими актерами. Перед анализом текстовой составляющей высказываний дикторов необходимо для начала получить орфографическую транскрипцию аудиофайлов, которая не предоставляется разработчиками.

Программные системы для сентимент-анализа

Экспериментальные системы для русскоязычного сентимент-анализа

На соревнованиях в рамках конференции «Диалог» в 2012 г. был предоставлен корпус РОМИП [31]. Авторы работы [34] показали наилучший результат классификации на 5 классов, применив *n*-граммный метод опорных векторов, использовав двоичные веса вместо традиционного TF-IDF, а также обучив модель на комбинированных корпусах. С применением данного подхода на корпусе РОМИП получено среднее значение F-меры = 30,63 %.

На соревнованиях в 2013 г. использовался тот же текстовый корпус, что и в 2012-м, дополнительно к нему организаторы включили корпус из новостных лент, который содержит прямую и косвенную речь с оценкой тональности высказывания (http://romip.ru/ru/collections/sentiment-news-collection-2012.html). Метка тональности текста может принимать одно из четырех значений: положительная, отрицательная, смешанная или нет оценки. В этом корпусе содержится около 5 тыс. новостных фрагментов. Авторы работы [35] достигли наилучшего значения F-меры = 65,9 % для бинарной классификации и 35,36 % для 5-классовой задачи, использовав метод максимальной энтропии и опорных векторов соответственно.

В 2015 г. на «Диалоге» была поставлена более широкая задача. Участникам был предоставлен корпус SentiRuEval-2015, и им необходимо было выделить аспектные термины, определить их тональность и тональность отзыва в целом. Лучший результат решения данной задачи описан в работе [36]. Автор работы применял рекуррентные нейронные сети и получил результат F-меры, равный 61,9 и 64,7 % для отзывов о ресторанах и автомобилях соответственно.

В 2016 г. соревнования проходили на текстовом корпусе SentiRuEval-2016. Задача участников состояла в том, чтобы определить репутационное отношение твита по отношению к конкретной компании. Авторы работы [37] использовали двухслойную нейронную сеть GRU (управляемый рекуррентный блок) и подавали входной вектор в обратной последовательности. При помощи этого метода достигнута F-мера = 55,17 и 55,94 % для отзывов о банках и мобильных операторах соответственно.

Программные продукты для анализа тональности русскоязычных текстов

Задача определения тональности текста является коммерчески востребованной, в связи с этим разрабатываются различные ориентированные компьютерные системы, анализирующие тональность текстов. На начало 2020 г. нам удалось найти пять программных систем в свободном доступе, предназначенных для сентимент-анализа русскоязычных текстов.

SentiFinder [38] — программный модуль высокоскоростной системы лингвистического анализа текстов Eureka Engine. Он определяет тональность текстов на русском, английском и армянском языках. Особенностью данного модуля является то, что он позволяет оценить степень эмоциональности высказывания. Он предназначен для определения тональности отзывов о различных продуктах, а также новостных лент и блогов.

Semantria [39] — программный модуль сентимент-анализа на базе платформы Lexalytics. Система позволяет классифицировать тональность сообщений на нескольких европейских языках, в том числе и на русском. Semantria предназначена для анализа текстов в области маркетинга.

SentiScan — технология распознавания тональности текста на базе платформы YouScan [40]. Классификатор SentiScan обучался на данных, которые содержали в себе отзывы о товарах из различных отраслей. YouScan является коммерческим продуктом, но у него есть бесплатный пробный период, который предоставляется по запросу.

SentiStrength — программный продукт для анализа настроений пользователя [41]. Он предназначен для анализа коротких социальных интернет-текстов. Результатом анализа текста являются две оценки, которые принимают значения от −5 (крайне отрицательно) до 1 (не отрицательно) и от 1 (не положительно) до 5 (крайне положительно). Изначально SentiStrength разрабатывался для анализа английского языка, но впоследствии адаптирован для других языков, в том числе для русского.

Texterra — приложение для анализа тональности новостных сообщений [42]. Анализируемые тексты могут быть из определенных областей: политики, финансов, Интернета, здоровья и постов Twitter. Демоверсия Texterra находится в сво-

■ ***Таблица 3***. Программные продукты для анализа тональности текстов
■ ***Table 3***. Software for analyzing the tonality of texts

Название системы (ссылка для доступа)	Число классов	Используемые методы	Ограничения демоверсии
SentiFinder (http://eurekaengine.ru/ru/demo/)	3	Случайный лес и градиентный бустинг	Анализ текстов объемом не более 10 тыс. символов
Semantria (https://www.lexalytics.com/demo)	3	Нет данных	Анализ текстов объемом до 16 384 символов
SentiScan (https://youscan.io/ru/demo)	3	Метод, основанный на правилах и машинном обучении	Не известны
SentiStrength (http://sentistrength.wlv.ac.uk/)	2	Метод, основанный на тональных словарях и правилах	Анализ сообщений до 100 символов
Texterra (https://texterra.ispras.ru/demo)	3	Метод опорных векторов	Не известны

бодном доступе, ее разработчики предоставляют возможность анализировать фактические новости, собранные с платформы Яндекс.Новости и Twitter, а также пользовательские тексты, введенные вручную.

В свободном доступе можно найти только демоверсии упомянутых систем. Ссылка на демоверсии, их ограничения, а также основные сведения о самих программных продуктах для анализа тональности русскоязычных текстов представлены в табл. 3.

В основе программных продуктов для сентимент-анализа текстов на русском языке лежат, как правило, традиционные методы обучения и не используются нейронные сети. Такой подход может быть обоснован тем, что нейронные сети требуют большого объема обучающих данных, а также большого количества вычислительных и временных ресурсов для их обучения.

Заключение

В статье представлен обзор подходов к анализу тональности русскоязычных текстовых данных. Наличие многочисленных работ на тему анализа тональности текста говорит о том, что данная задача является актуальной и коммерчески востребована во многих сферах, включая рекламу, политику, маркетинг и т. п. Это подтверждается увеличением с каждым годом количества конференций в области анализа текста, а также количества публикаций по анализу как русскоязычных данных, так и текстов на других языках. Однако системы сентимент-анализа русскоязычных текстов развиты меньше, чем для основных мировых языков. По данным академии Google за 2019 г., было опубликовано всего около 28 000 работ по сентимент-анализу русскоязычных текстов, тогда как по англоязычным текстам вышло около 43 000 публикаций. Также русскоязычный сентимент-анализ показывает довольно низкую точность по сравнению с англоязычным, что связано со сложной структурой русского языка. Чтобы подтвердить это утверждение, можно рассмотреть работы по сентимент-анализу чешского языка, так как грамматики русского и чешского языков схожи. В работах [43–45] проводится анализ тональности текстов на английском и чешском языках, по результатам исследования видно, что точность распознавания сентимента в английском языке выше, чем в чешском.

В дальнейших исследованиях мы планируем реализовать сентимент-анализ разговорной речи дикторов с использованием корпуса RAMAS [33]. Для этого необходимо будет получить орфографическую транскрипцию речи для каждого диктора. На основе полученных данных планируется построить классификатор, используя для начала метод на основе тональных словарей, а в последующем и другие методы классификации, описанные в статье.

Финансовая поддержка

Исследование проведено при поддержке РФФИ (проект № 18-07-01407), РНФ (проект № 18-11-00145, раздел 3) и бюджетной темы № 0073-2019-0005.

Литература

1. **Ениколопов С. Н., Кузнецова Ю. М., Смирнов И. В., Станкевич М. А., Чудова Н. В.** Создание инструмента автоматического анализа текста в интересах социогуманитарных исследований. Часть 1. Методические и методологические аспекты. *Искусственный интеллект и принятие решений*, 2019, № 2, с. 28–38. doi:10.14357/20718594190203
2. **Поляков Е. В., Восков Л. С., Абрамов П. С., Поляков С. В.** Исследование обобщенного подхода к решению задач анализа настроений коротких текстовых сообщений в задачах обработки естественного языка. *Информационно-управляющие системы*, 2020, № 1, с. 2–14. doi:10.31799/1684-8853- 2020-1-2-14
3. **Soumya G. K., Joseph S.** Text classification by augmenting bag of words (BOW) representation with co-occurrence feature. *IOSR Journal of Computer Engineering*, 2014, vol. 16(1), pp. 34–38.
4. **Potdar K., Pardawala T. S., Pai C. D.** A comparative study of categorical variable encoding techniques for neural network classifiers. *International Journal of Computer Applications*, 2017, vol. 175, no. 4, pp. 7–9.
5. **Steinberger J., Jezek K.** Text summarization and singular value decomposition *Proceedings of International Conference on Advances in Information Systems*, Springer, Berlin, Heidelberg, 2004, pp. 245–254.
6. **Mikolov T., Chen K., Corrado G., Dean J.** Efficient estimation of word representations in vector space. *Proceedings of the International Conference on Learning Representations (ICLR 2013)*, 2013. https://openreview.net/forum?id=idpCdOWtqXd60#7b076554-87ba-4e1e-b7cc-2ac107ce8e4d (дата обращения: 02.05.2020).
7. **Pennington J., Socher R., Manning C. D.** Glove: Global vectors for word representation. *Proceedings of International Conference on Empirical Methods in Natural Language Processing (EMNLP-2014)*, 2014, pp. 1532–1543.
8. **Pylieva H., Chernodub A., Grabar N., Hamon T.** Improving automatic categorization of technical vs. Laymen medical words using fasttext word embeddings. *Proceedings of the 1st International Workshop on Informatics and Data-Driven Medicine*, *IDDM 2018*, 2018, pp. 93–102.
9. **Devlin J., Chang M., Lee K., Toutanova K.** Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 2019, vol. 1 (Long and Short Papers), pp. 4171–4186. doi:10.18653/v1/N19-1423
10. **Sanh V., Debut L., Chaumond J., Wolf T.** DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108* (дата обращения: 05.04.2020).
11. **Куратов Ю., Архипов М.** Адаптация глубоких двунаправленных многоязычных моделей на основе архитектуры transformer для русского языка. *Компьютерная лингвистика и интеллектуальные технологии*, 2019, вып. 18, с. 333–339.
12. **Peters M., Neumann M., Iyyer M., Gardner M., Clark C., Lee K., Zettlemoyer L.** Deep contextualized word representations. *NAACL-HLT*, 2018, vol. 1 (Long Papers), pp. 2227–2237. doi:10.18653/v1/N18-1202
13. **Dvoynikova A., Verkholyak O., Karpov A.** Analytical review of methods for identifying emotions in text data. *CEUR-WS*, 2020, vol. 2552, pp. 8–21.
14. **Тутубалина Е. В., Иванов В. В., Загулова М., Мингазов Н., Алимова И., Малых В.** Тестирование методов анализа тональности текста, основанных на словарях. *Электронные библиотеки*, 2015, т. 18, № 3-4, с. 138–162.
15. **Паничева П. В.** Система сентиментного анализа ATEX, основанная на правилах, при обработке текстов различных тематик. *Компьютерная лингвистика и интеллектуальные технологии*, 2013, вып. 12, т. 2, с. 101–113.
16. **Котельников Е. В., Клековкина М. В.** Автоматический анализ тональности текстов на основе методов машинного обучения. *Компьютерная лингвистика и интеллектуальные технологии*, 2012, вып. 11, т. 2, с. 27–36.
17. **Maltseva A. V., Makhnytkina O. V., Shilkina N. E., Lizunova I. A.** Social media sentiment analysis with context space model. *Communications in Computer and Information Science*, 2020, vol. 1135, pp. 399–412. doi:10.1007/978-3-030-39296-3_29
18. **Aken B., Risch J., Krestel R., Loser A.** Challenges for toxic comment classification: An in-depth error analysis. *EMNLP*, 2018, pp. 33–42.
19. **Воронина И. Е., Гончаров В. А.** Анализ эмоциональной окраски сообщений в социальных сетях (на примере сети «ВКонтакте»). *Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии*, 2015, № 4, с. 151–158.
20. **Konig A. C., Brill E.** Reducing the human overhead in text categorization. *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006, pp. 598–603.
21. **Lakshmi B. S., Raj P. S., Vikram R. R.** Sentiment analysis using deep learning technique CNN with KMeans. *International Journal of Pure and Applied Mathematics*, 2017, vol. 114, no. 11, pp. 47–57.
22. **Rogers A., Romanov A., Rumshisky A., Volkova S., Gronas M., Gribov A.** Rusentiment: An enriched sentiment analysis dataset for social media in Russian. *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 755–763.
23. **Loukachevitch N., Levchik A.** Creating a general Russian sentiment lexicon. *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC'16)*, 2016, pp. 1171–1176.

24. **Loukachevitch N., Dobrov B.** RuThes linguistic ontology vs. Russian wordnets. *Proceedings of the 7th Global Wordnet Conference*, 2014, pp. 154–162.
25. **Алексеева С. В., Кольцов С. Н., Кольцова О. Ю.** Linis-crowd. org: лексический ресурс для анализа тональности социально-политических текстов на русском языке. *Труды XVIII объединенной конференции «Интернет и современное общество» (IMS-2015)*, 2015, с. 25–34.
26. **Sokolova M., Bobicev V.** Classification of emotion words in Russian and Romanian languages. *Proceedings of the International Conference RANLP-2009*, 2009, pp. 416–420.
27. **Беляков М. В.** Анализ новостных сообщений сайта МИД РФ методом сентимент-анализа (статья 2). *Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика*, 2016, № 4, с. 115–124.
28. **Loukachevitch N., Blinov P., Kotelnikov E., Rubtsova Y., Ivanov V., Tutubalina E.** SentiRuEval: testing object-oriented sentiment analysis systems in Russian. *Компьютерная лингвистика и интеллектуальные технологии*, 2015, вып. 14, т. 2, с. 3–13.
29. **Lukashevich N. V., Rubtsova Y. V.** SentiRuEval-2016: overcoming time gap and data sparsity in tweet sentiment analysis. *Компьютерная лингвистика и интеллектуальные технологии*, 2016, вып. 15, с. 416–426.
30. **Рубцова Ю. В.** Построение корпуса текстов для настройки тонового классификатора. *Программные продукты и системы*, 2015, № 1(109), с. 72–78.
31. **Chetviorkin I., Braslavskiy P., Loukachevich N.** Sentiment analysis track at ROMIP 2011. *Компьютерная лингвистика и интеллектуальные технологии*, 2012, вып. 11, т. 2, с. 1–14.
32. **Глазкова А. В.** Оценка степени близости категорий текстов при решении задач классификации электронных документов. *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*, 2015, № 2 (31), с. 18–25. doi:10.17223/19988605/31/2
33. **Perepelkina O., Kazimirova E., Konstantinova M.** RAMAS: Russian multimodal corpus of dyadic interaction for affective computing. *Proceedings of 20th International Conference on Speech and Computer SPECOM-2018*, Springer, Cham, 2018, pp. 501–510.
34. **Pak A., Paroubek P.** Language independent approach to sentiment analysis (LIMSI participation in ROMIP'11). *Компьютерная лингвистика и интеллектуальные технологии*, 2012, вып. 11, т. 2, с. 37–50.
35. **Blinov P. D., Klelovkina M. V., Ktelnikov E. V., Pestov O. A.** Research of lexical approach and machine learning methods for sentiment analysis. *Компьютерная лингвистика и интеллектуальные технологии*, 2013, вып. 12, т. 2, с. 51–61.
36. **Тарасов Д.** Глубокие рекуррентные нейронные сети для аспектно-ориентированного анализа тональности отзывов пользователей на различных языках. *Компьютерная лингвистика и интеллектуальные технологии*, 2015, вып. 14, т. 2, с. 53–64.
37. **Trofimovich J.** Comparison of neural network architectures for sentiment analysis of Russian tweets. *Компьютерная лингвистика и интеллектуальные технологии*, 2016, вып. 15, с. 50–59.
38. **Zafar L., Afzal M. T., Ahmed U.** Exploiting polarity features for developing sentiment analysis tool. *CEUR-WS*, 2017, vol. 1874, no. 4. http://ceur-ws.org/Vol-1874/paper_4.pdf (дата обращения: 02.05.2020).
39. **Зверева П. П.** Сентимент-анализ текста (на материале печатных текстов газеты "The New York Times" о России и россиянах). *Вестник Московского государственного областного университета. Серия: Лингвистика*, 2014, № 5, с. 32–37.
40. **Кривоногова С. А.** Психоэмоциональная окрашенность текста: теория и методы исследования. *Материалы 68-й научной конференции «Наука ЮУРГУ»*, 2016, т. 100, с. 368–375.
41. **Thelwall M.** The heart and soul of the web? Sentiment strength detection in the social web with SentiStrength. *Cyberemotions, Springer, Cham*, 2017, pp. 119–134.
42. **Mayorov V., Andrianov I.** MayAnd at SemEval-2016 Task 5: Syntactic and word2vec-based approach to aspect-based polarity detection in Russian. *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 325–329.
43. **Hercig T., Brychcin T., Svoboda L., Konkol M., Steinberger J.** Unsupervised methods to improve aspect-based sentiment analysis in Czech. *Computacion y Sistemas*, 2016, vol. 20 (3), pp. 365–375. doi:10.13053/cys-20-3-2469
44. **Hercig T., Brychcin T., Svoboda L., Konkol M.** Uwb at semeval-2016 task 5: Aspect based sentiment analysis. *SemEval-2016*, 2016, pp. 342–349.
45. **Prikrylova K., Kubon V., Veselovska K.** The role of conjunctions in adjective polarity analysis in Czech. *Computacion y Sistemas*, 2016, vol. 20 (3), pp. 377–386. doi:10.13053/cys-20-3-2460

UDC 004.934.2
doi:10.31799/1684-8853-2020-4-20-30

Analytical review of approaches to Russian text sentiment recognition

A. A. Dvoynikova[a], Programmer, orcid.org/0000-0001-8047-6639
A. A. Karpov[a], Dr. Sc., Tech., Associate Professor, orcid.org/0000-0003-3424-652X, karpov@iias.spb.su
[a]Saint-Petersburg Institute for Informatics and Automation of the RAS, 39, 14 Line, V. O., 199178, Saint-Petersburg, Russian Federation

Introduction: In recent years, sentiment analysis has found practical application in many areas, such as evaluating the quality of products and services based on customers' online reviews, analyzing negative emotions in messages, forecasting stock markets or political situations based on news data. In this regard, a large number of systems and methods for Russian text sentiment analysis are being developed. **Purpose:** A detailed review of approaches, and comparative analysis of available databases in the field of Russian text sentiment analysis. **Results:** Our analytical review of the approaches to Russian text data sentiment analysis has shown that there are a large number of ways for preprocessing, vectorization and machine classification of the text data. Studying the available databases shows that the Russian text sentiment analysis is less developed than that for other major world languages. Studying the existing software systems for Russian text analysis reveals their low accuracy compared to English, which can be caused by the sophisticated structure of Russian. **Discussion:** In our further research, we plan to implement sentiment analysis of spoken speech using audio data. To do this, we will need to obtain a spelling transcription of speech for each speaker.
Keywords — text tonality, text vectorization, sentiment analysis, computational paralinguistic.

For citation: Dvoynikova A. A., Karpov A. A. Analytical review of approaches to Russian text sentiment recognition. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 4, pp. 20–30 (In Russian). doi:10.31799/1684-8853-2020-4-20-30

References

1. Enikolopov S. N., Kuznetsova Y. M., Smirnov I. V., Stankevich M. A., Chudova N. V. Creating a text analysis tool for socio-humanitarian research. Part 1. Methodical and methodological aspects. *Artificial Intelligence and Decision Making*, 2019, no. 2, pp. 28–38 (In Russian). doi:10.14357/20718594190203
2. Polyakov E. V., Voskov L. S., Abramov P. S., Polyakov S. V. Generalized approach to sentiment analysis of short text messages in natural language processing. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 1, pp. 2–14 (In Russian). doi:10.31799/1684-8853- 2020-1-2-14
3. Soumya G. K., Joseph S. Text classification by augmenting bag of words (BOW) representation with co-occurrence feature. *IOSR Journal of Computer Engineering*, 2014, vol. 16(1), pp. 34–38.
4. Potdar K., Pardawala T. S., Pai C. D. A comparative study of categorical variable encoding techniques for neural network classifiers. *International Journal of Computer Applications*, 2017, vol. 175, no. 4, pp. 7–9.
5. Steinberger J., Jezek K. Text summarization and singular value decomposition *Proceedings of International Conference on Advances in Information Systems*, Springer, Berlin, Heidelberg, 2004, pp. 245–254.
6. Mikolov T., Chen K., Corrado G., Dean J. Efficient estimation of word representations in vector space. *Proceedings of the International Conference on Learning Representations (ICLR 2013)*, 2013. Available at: https://openreview.net/forum?id=idpCdOWtqXd60#7b076554-87ba-4e1e-b7cc-2ac107ce8e4d (accessed 2 May 2020).
7. Pennington J., Socher R., Manning C. D. Glove: Global vectors for word representation. *Proceedings of International Conference on Empirical Methods in Natural Language Processing (EMNLP-2014)*, 2014, pp. 1532–1543.
8. Pylieva H., Chernodub A., Grabar N., Hamon T. Improving automatic categorization of technical vs. Laymen medical words using fasttext word embeddings. *Proceedings of the 1st International Workshop on Informatics and Data-Driven Medicine*, *IDDM 2018*, 2018, pp. 93–102.
9. Devlin J., Chang M., Lee K., Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 2019, vol. 1 (Long and Short Papers), pp. 4171–4186. doi:10.18653/v1/N19-1423
10. Sanh V., Debut L., Chaumond J., Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108* (accessed 05 April 2020).
11. Kuratov Yu., Arkhipov M. Adaptation of deep bidirectional multilingual transformers for russian language. *Computational Linguistics and Intellectual Technologies*, 2019, iss. 18, pp. 333–339 (In Russian).
12. Peters M., Neumann M., Iyyer M., Gardner M., Clark C., Lee K., Zettlemoyer L. Deep contextualized word representations. *NAACL-HLT*, 2018, vol. 1 (Long Papers), pp. 2227–2237. doi:10.18653/v1/N18-1202
13. Dvoynikova A., Verkholyak O., Karpov A. Analytical review of methods for identifying emotions in text data. *CEUR-WS*, 2020, vol. 2552, pp. 8–21.
14. Tutubalina E. V., Ivanov V. V., Zagulova M. A., Mingazov N. R., Alimova I. S., Malykh V. A. Sentiment classification of reviews and twitter posts based on dictionaries. *Russian Digital Libraries Journal*, 2015, vol. 18, no. 3-4, pp. 138–162 (In Russian).
15. Panicheva P. V. ATEX: a rule-based sentiment analysis system processing texts in various topics. *Computational Linguistics and Intellectual Technologies*, 2013, iss. 12, vol. 2, pp. 101–113 (In Russian).
16. Kotelnikov E. V., Klekovkina M. V. Automatic text tonality analysis based on machine learning methods. *Computational Linguistics and Intellectual Technologies*, 2012, iss. 11, vol. 2, pp. 27–36 (In Russian).
17. Maltseva A. V., Makhnytkina O. V., Shilkina N. E., Lizunova I. A. Social media sentiment analysis with context space model. *Communications in Computer and Information Science*, 2020, vol. 1135, pp. 399–412. doi:10.1007/978-3-030-39296-3_29
18. Aken B., Risch J., Krestel R., Loser A. Challenges for toxic comment classification: An in-depth error analysis. *EMNLP*, 2018, pp. 33–42.
19. Voronina I. E., Goncharov V. A. Analysis of the emotional color of messages in social networks (for example, the "Vkontakte network"). *Bulletin of the Voronezh State University. Series: System Analysis and Information Technologies*, 2015, no. 4, pp. 151–158 (In Russian).
20. Konig A. C., Brill E. Reducing the human overhead in text categorization. *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006, pp. 598–603.
21. Lakshmi B. S., Raj P. S., Vikram R. R. Sentiment analysis using deep learning technique CNN with KMeans. *International Journal of Pure and Applied Mathematics*, 2017, vol. 114, no. 11, pp. 47–57.
22. Rogers A., Romanov A., Rumshisky A., Volkova S., Gronas M., Gribov A. Rusentiment: An enriched sentiment analysis dataset for social media in Russian. *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 755–763.
23. Loukachevitch N., Levchik A. Creating a general Russian sentiment lexicon. *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC'16)*, 2016, pp. 1171–1176.
24. Loukachevitch N., Dobrov B. RuThes linguistic ontology vs. Russian wordnets. *Proceedings of the 7th Global Wordnet Conference*, 2014, pp. 154–162.

25. Alexeeva S., Kolcov S., Koltsova O. Linis-crowd.org: A lexical resource for Russian sentiment analysis of social media. *Trudy XVIII ob"edinennoj konferencii «Internet i sovremennoe obshchestvo» (IMS-2015)* [Proceedings of the XVIII Joint Conference "Internet and Modern Society" (IMS-2015)], 2015, pp. 25–34 (In Russian).
26. Sokolova M., Bobicev V. Classification of emotion words in Russian and Romanian languages. *Proceedings of the International Conference RANLP-2009*, 2009, pp. 416–420.
27. Belyakov M. V. The analysis of news messages on the RF ministry of foreign affairs affairs website by the sentimental analysis (article 2). *Bulletin of the Peoples' Friendship University of Russia. Series: Theory of Language. Semiotics. Semantics*, 2016, no. 4, pp. 115–124 (In Russian).
28. Loukachevitch N., Blinov P., Kotelnikov E., Rubtsova Y., Ivanov V., Tutubalina E. SentiRuEval: testing object-oriented sentiment analysis systems in Russian. *Computational Linguistics and Intellectual Technologies*, 2015, iss. 14, vol. 2, pp. 3–13.
29. Lukashevich N. V., Rubtsova Y. V. SentiRuEval-2016: overcoming time gap and data sparsity in tweet sentiment analysis. *Computational Linguistics and Intellectual Technologies*, 2016, iss. 15, pp. 416–426.
30. Rubcova U. V. Building a text corpus for setting up a tone classifier. *Software & Systems*, 2015, no. 1(109), pp. 72–78 (In Russian).
31. Chetviorkin I., Braslavskiy P., Loukachevich N. Sentiment analysis track at ROMIP 2011. *Computational Linguistics and Intellectual Technologies*, 2012, iss. 11, vol. 2, pp. 1–14.
32. Glazkova A. V. The evaluation of the proximity of text categories for solving electronic documents classification tasks. *Bulletin of Tomsk State University. Management, Computer Engineering and Informatics*, 2015, no. 2(31), pp. 18–25 (In Russian). doi:10.17223/19988605/31/2
33. Perepelkina O., Kazimirova E., Konstantinova M. RAMAS: Russian multimodal corpus of dyadic interaction for affective computing. *Proceedings of 20th International Conference on Speech and Computer SPECOM-2018*, Springer, Cham, 2018, pp. 501–510.
34. Pak A., Paroubek P. Language independent approach to sentiment analysis (LIMSI participation in ROMIP'11). *Computational Linguistics and Intellectual Technologies*, 2012, iss. 11, vol. 2, pp. 37–50.
35. Blinov P. D., Klelovkina M. V., Ktelnikov E. V., Pestov O. A. Research of lexical approach and machine learning methods for sentiment analysis. *Computational Linguistics and Intellectual Technologies*, 2013, iss. 12, vol. 2, pp. 51–61.
36. Tarasov D. S. Deep recurrent neural networks for multiple language aspect-based sentiment analysis of user reviews. *Computational Linguistics and Intellectual Technologies*, 2015, iss. 14, vol. 2, pp. 53–64 (In Russian).
37. Trofimovich J. Comparison of neural network architectures for sentiment analysis of russian tweets. *Computational Linguistics and Intellectual Technologies*, 2016, iss. 15, pp. 50–59.
38. Zafar L., Afzal M. T., Ahmed U. Exploiting polarity features for developing sentiment analysis tool. *CEUR-WS*, 2017, vol. 1874, no. 4. Available at: http://ceur-ws.org/Vol-1874/paper_4.pdf (accessed 2 May 2020).
39. Zvereva P. Sentiment-analysis of text (texts about Russia and the Russians from The New York Times). *Bulletin of the Moscow State Regional University. Series: Linguistics*, 2014, no. 5, pp. 32–37 (In Russian).
40. Krivonogova S. A. Psychoemotional color of the text: theory and research methods. *Materialy 68-j nauchnoj konferencii «Nauka YUURGU»* [Materials of the 68th Scientific Conference "Science of the South Ural State University"], 2016, vol. 100, pp. 368–375 (In Russian).
41. Thelwall M. The heart and soul of the web? Sentiment strength detection in the social web with SentiStrength. *Cyberemotions, Springer, Cham*, 2017, pp. 119–134.
42. Mayorov V., Andrianov I. MayAnd at SemEval-2016 Task 5: Syntactic and word2vec-based approach to aspect-based polarity detection in Russian. *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 325–329.
43. Hercig T., Brychcin T., Svoboda L., Konkol M., Steinberger J. Unsupervised methods to improve aspect-based sentiment analysis in Czech. *Computacion y Sistemas*, 2016, vol. 20 (3), pp. 365–375. doi:10.13053/cys-20-3-2469
44. Hercig T., Brychcin T., Svoboda L., Konkol M. Uwb at semeval-2016 task 5: Aspect based sentiment analysis. *SemEval-2016*, 2016, pp. 342–349.
45. Prikrylova K., Kubon V., Veselovska K. The role of conjunctions in adjective polarity analysis in Czech. *Computacion y Sistemas*, 2016, vol. 20 (3), pp. 377–386. doi:10.13053/cys-20-3-2460