

оригинальная статья

Сентимент-анализ: лингвистический потенциал регламентации предобработки

Баркович Александр Аркадьевич

Минский государственный лингвистический университет, Минск, Беларусь
<https://orcid.org/0000-0001-8469-8431>
Scopus Author ID: 57208124708
albark@tut.by

Поступила в редакцию 03.04.2023. Принята после рецензирования 28.04.2023. Принята в печать 28.04.2023.

Аннотация: Рассматривается специфика регламентации сентимент-анализа как актуального направления автоматизированной обработки естественного языка. Цель – охарактеризовать лингвистический потенциал регламентации предобработки языкового материала. Связанная с этим деятельность характеризуется значимыми достижениями практического характера, однако ее теоретическое обоснование недостаточно упорядоченно и дискуссионно. Несмотря на динамичное развитие множества направлений сферы информационных технологий, принципиальные основы такой деятельности по-прежнему тесно коррелируют с лингвистической системой знаний. Практически безальтернативен методологический приоритет прикладной традиции языкознания с учетом междисциплинарной специфики сферы современной коммуникации. Комплексный характер исследования во многом обеспечен разносторонним инструментарием компьютерной лингвистики. Сложность проблемной области обуславливает ориентацию метаописания на алгоритмизацию и моделирование процедуры оценки тональности текста. Результативность процедуры существенно детерминирована ее оптимальной конфигурацией. Обоснованной представляется регламентация процедуры предобработки материала с последовательным выявлением метаструктуры, определением референтности, уровневой ориентацией и выбором модели анализа. Описаны основные шаги алгоритма предобработки и их особенности, выявлена и охарактеризована специфика соответствующей практики. Исследование будет способствовать продуктивной теоретической рефлексии и оптимизации практической деятельности по оценке тональности текста, или сентимент-анализу. В широком контексте целесообразное раскрытие лингвистического потенциала актуально для всей сферы автоматизированной обработки естественного языка.

Ключевые слова: сентимент-анализ, автоматизированная обработка, естественный язык, лингвистическая специфика, регламентация, предобработка, модель, алгоритм

Цитирование: Баркович А. А. Сентимент-анализ: лингвистический потенциал регламентации предобработки. *Виртуальная коммуникация и социальные сети*. 2023. Т. 2. № 3. С. 116–123. <https://doi.org/10.21603/2782-4799-2023-2-3-116-123>

full article

Sentiment Analysis: Linguistic Potential of Preprocessing Regimentation

Aleksandr A. Barkovich

Minsk State Linguistic University, Minsk, Belarus
<https://orcid.org/0000-0001-8469-8431>
Scopus Author ID: 57208124708
albark@tut.by

Received 3 Apr 2023. Accepted after peer review 28 Apr 2023. Accepted for publication 28 Apr 2023.

Abstract: The article deals with the sentiment analysis regimentation as a relevant direction in automated natural language processing and its linguistic potential. Despite its impressive practical significance, the sentiment analysis still lacks reliable theoretical foundation. Although information technologies develop very fast, their fundamental foundations correlate with the linguistic system of knowledge. In fact, the methodological priority of the applied linguistics has no alternative with regard to the interdisciplinary specificity of the modern communication.

The complex nature of this research made the authors appeal to the computer linguistics in order to provide a meta-description on the algorithmization and modeling of sentiment evaluation. The effectiveness of the relevant practice was conditioned by the optimal configuration of the procedure and an appropriate material evaluation. The preprocessing included identifying the meta-structure, defining its referentiality and level orientation, and choosing the analysis model. The authors described these main steps of the preprocessing algorithm, as well as the relevant practice. The study contributes to productive theoretical optimization of text sentiment analysis. In a broad context, the expedient disclosure of linguistic potential is relevant to the whole sphere of automated natural language processing.

Keywords: sentiment analysis, automated processing, natural language, linguistic specificity, regimentation, preprocessing, model, algorithm

Citation: Barkovich A. A. Sentiment Analysis: Linguistic Potential of Preprocessing Regimentation. *Virtual Communication and Social Networks*, 2023, 2(3): 116–123. (In Russ.) <https://doi.org/10.21603/2782-4799-2023-2-3-116-123>

Введение

Сентимент-анализ – инновационная методика выявления и обобщения неочевидных лингво-прагматических особенностей речевой практики. Возможности компьютерного сопровождения коммуникации естественным образом распространились на аспекты ее научного осмысления, что в числе прочих достижений прикладного характера позволило реализовать продуктивный и высокоидентичный инструментарий изучения тональности текста, или сентимент-анализ. Соответствующая проблемная область оказалась в фокусе исследований по автоматизированной обработке естественного языка (NLP) в конце прошлого столетия. Уже в 1980-х гг. был осознан потенциал формализации «представлений не только о неодушевленном мире, но и о других людях и их убеждениях»¹ [Wilks, Bien 1983: 96]. Однако настоящее становление сентимент-анализа пришлось на начало XXI в. Сам термин был зафиксирован в статье японских исследователей [Nasukawa, Yi 2003]. В настоящее время динамично растущая востребованность сентимент-анализа обусловлена экспансией сети Интернет и лавинообразным расширением информационного континуума [Баркович 2015]. Очевидно, что традиционные лингвистические средства уже малоэффективны для осмысления колоссальных объемов новейшей речевой практики. Ежесекундный прирост терабайтов коммуникационных данных уже не охватывается специалистами вручну за год. Времена, когда отдельные тексты изучались поколениями исследователей, безвозвратно уходят в прошлое. Сегодня нужна максимальная автоматизация лингвистического анализа, прагматическим аспектам такой практики и призван помочь сентимент-анализ.

Методы и материалы

Сентимент-анализ предоставляет возможность анализировать субъективную специфику больших объемов текстов за максимально короткое время, и это находит свое применение во многих областях науки и производства. Данная технология уже используется для мониторинга общественного мнения, сопровождения маркетинга, анализа новостных потоков и разного рода прогнозирования. В данной связи анализируются мнения, суждения и умозаключения с целью понять истинное отношение их автора к фактам и аргументам, субъектам и объектам, сути и атрибутам явлений.

Сентимент-анализ, или оценка тональности текста – оценка эмоциональной составляющей речевой продукции. В английском языке синонимия данной номинации более вариативна: sentiment analysis, sentiment detection, sentiment classification, opinion mining, subjectivity analysis, polarity classification, affect analysis и т. д. (см., напр., [Beigi et al. 2016: 315]). Однако пока рано говорить о совершенстве технологий сентимент-анализа. Возможно, определенный нераскрытый потенциал соответствующей практики связан как раз с упрощенными представлениями о специфике естественных языков. Именно лингвистический потенциал традиционно обогащает теорию и практику информационно-зависимых технологий со времен изобретения компьютера. Сегодня, как никогда ранее, в сферу информационной разработки активно включаются культурные, контекстные, содержательные и эмоциональные компоненты речевой практики [Barkovich 2020: 37].

Наше исследование основывается на прикладной традиции языкознания и методологии компьютерной лингвистики. В качестве методик актуальны

¹ Здесь и далее по тексту перевод выполнен автором статьи.

описание, лингвистический и статистический анализ, сравнение, моделирование, машинное обучение и другие инструменты научной рефлексии. Данные средства использовались комплексно, соотносясь с лингвистической направленностью исследования.

Для выявления и описания потенциала и специфики регламентации предобработки речевой продукции, в которой предполагается оценить тональность (эмоциональность и экспрессию), необходим широкий охват референтной проблематики и ее материального воплощения. Поэтому необходимо не только исследовать эмпирический материал, но и проанализировать логику и результативность релевантной процедуры. Подобная сбалансированность представляется рациональной для достижения заявленной цели – характеристики лингвистического потенциала регламентации предобработки при сентимент-анализе.

Анализ и результаты

С процедурной точки зрения типичная задача сентимент-анализа может быть представлена примерно следующим образом: оценка тональности текста положительна или отрицательна? (при бинарном шкалировании). Факультативная задача: насколько эта оценка выражена? При этом не исключается и нейтральная тональность, или полюсность, объекта анализа (предложения или текста), которая в подобной ситуации будет третьим параметром шкалы оценок. Шкалирование оценок по многополосной шкале может быть и более сложным, состоять, например, из 4, 5 и более параметров. Решению подобных задач подчинена процедурная логика сентимент-анализа, или его регламент. Ключевыми элементами такого регламента являются предобработка (предварительная оценка референтного потенциала материала), оценка тональности и постобработка (интерпретация результатов анализа).

Качественная подготовка к проведению сентимент-анализа позволяет не только избежать необязательных ресурсных и временных затрат в процессе соответствующей деятельности, но и своевременно компетентно оценить возможные сложности и перспективы: сложностей можно избежать, а перспективы реализовать. Именно подготовка является критерием продуктивности предварительной лингвистической обработки материала исследования. Рассмотрим 4 основных компонента предобработки материала для оценки его тональности.

1. Выявление метаструктуры подразумевает идентификацию ключевых компонентов исследуемого образца речевой продукции. Отражение в речи

действительности разного рода описывается разными лингвистическими категориями: субъект, действие, объект, атрибут, продукт, инструмент, место, время, адресат и др. Однако для решения прикладных задач такой континуум является слишком абстрактным и явно избыточным. Конечно, для определения тональности текста не все категории равнозначны, и имеются все основания сосредоточиться на основных элементах. Список таких элементов не настолько обширен. По мнению одних ученых, он может включать 5 элементов: объект, субъект, атрибут, оценка, время (см., напр., [Pang, Lee 2008]). По мнению других, это должен быть «кортеж из 4 элементов» (entity, sentiment value, holder, time). Где entity – объект, о котором автор (holder) высказал мнение в момент времени (time) [Харламов и др. 2013]. Есть и другие взгляды на состав ключевых элементов. Критерием оптимальности можно считать наличие минимально достаточного круга метаязыковых маркеров для проведения предварительной оценки языкового материала. И такими маркерами, безусловно, будут *субъект* (источник мнения), *объект*, по отношению к которому высказывается суждение, и *действие* (процессуальность, состояние), подтверждающее предикативную состоятельность ситуации. Именно эти три элемента метаструктуры являются ключевыми для идентификации коммуникационной самодостаточности предполагаемого к анализу речевого фрагмента.

Субъектная специфика при анализе дискурса неперсонализированных участников не является определяющей: зачастую высказывания, например, комментарии к новостной статье, принадлежат разным субъектам. Установление их личностных особенностей, конечно, возможно, но обычно неприципиально для решения поставленных задач. *Объект* в такого рода исследованиях описывается предварительно и в процессе сентимент-анализа рассматривается статично, будучи фактически обезличенным. *Действие* актуально как показатель динамики коммуникационной ситуации. Принципиальным является установление комплекса данных элементов в структуре анализируемой единицы материала.

2. Определение референтности направлено на решение вопроса о целесообразности оценки тональности исследуемого материала. Здесь учитывается главным образом наличие его оценочной атрибутивности. Если материал содержит данные оценочного характера, или мнения (выражаются как непосредственное мнение или сравнение), то он может быть источником релевантной

информации о настроенности или намерениях коммуниканта (автора, реципиента, актора и т. д.). Соответственно, соблюдается условие обоснованности дальнейшего сентимент-анализа. При отсутствии оценочности материал считается содержащим сугубо фактическую информацию и признается не представляющим интереса для оценки тональности текста. Существенным потенциалом при решении вопроса обладает корпусный подход [Liu 2012]. Нередко в процессе предобработки решается вопрос об удалении из анализируемого материала т. н. фактических фрагментов. Однако, как правило, наличие таких фрагментов никак не влияет на качество автоматизированного анализа. Принципиальным критерием является наличие в материале выраженной тем или иным образом оценочности.

3. Уровневая ориентация предполагает выбор единицы исследования: по сложившимся в компьютерной лингвистике стереотипам это предложение либо текст. В компьютерной лингвистике вместо термина *текст* часто используется термин *документ* (см., напр., [Большакова и др. 2011]). Особой целесообразности в этом нет, однако во избежание интерференции наименования самой деятельности по оценке тональности текста и процедуры изучения конкретного текста логика использования во втором случае термина *документ* имеется. Разумеется, можно дополнительно выделять в качестве уровней и аспекты, и сущности, и события [Семина 2020; Liu 2012]. Но на практике тональность отдельных слов и словосочетаний при отсутствии предикативной составляющей оценивается редко и с лингвистической точки зрения не характеризуется очевидной продуктивностью.

Безусловно, общепринятым эталоном завершённой синтактико-семантической структуры является предложение. Сентимент-анализ отдельных предложений в последнее время актуален на фоне распространённости коротких сообщений и комментариев в социальных сетях. Например, в сети Twitter ограничением является объём в 280 знаков. Исходя из статистики, «средняя длина русского слова составляет 5,28 символа. Средняя длина предложения в русском языке составляет 10,38 слов»². Таким образом, в Twitter русскоязычный текст может в среднем состоять из приблизительно пяти предложений. Это свидетельствует о существенной значимости каждого из них. Однако, как правило, определение тональности отдельных предложений

требуется для решения более масштабных задач и уточнения характера мнения в тексте в целом. Поэтому анализ подобного рода менее востребован, чем анализ документа. При этом задача определения тональности предложения представляется более простой по сравнению с процедурой анализа документа. Нередко ориентация на уровень предложения напрямую обусловлена недостаточной предварительной структуризацией материала исследования.

Под уровнем документа в практическом представлении понимается анализ всего анализируемого фрагмента материала. Большинство исследований в рамках сентимент-анализа нацелено на задачи совокупной оценки документа [Майорова 2020; Чернышев 2018]. В реальности он может одновременно содержать предложения как с негативной, так и с позитивной оценкой, мнения в одном документе могут быть и взаимоисключающими. В таком случае средняя арифметическая оценка документа оказывается менее информативной, чем оценка его отдельных частей – предложений. Преимуществом анализа документа является жанровая прозрачность материала. Собственно, сентимент-анализ и формировался за счёт специфики как направление социальных сетей и сайтов с отзывами. Оценочная составляющая (комментирование) отвечает критериям спонтанности, диалогичности, экспрессивности, а значит, и аутентичности материала. Так, уровневая ориентация исследуемого материала во многом предопределяет стратегию и тактику анализа тональности речевой практики.

4. Выбор модели анализа – существенный компонент процедуры предварительного планирования оценки тональности текста. На сегодняшний день реализуются 3 основных модели сентимент-анализа:

- основанная на правилах,
- статистическая,
- гибридная.

Классификаций моделей сентимент-анализа существует множество. Типична следующая позиция: «Мы идентифицируем методы обнаружения настроений как принадлежащие к одной из трех категорий, каждая из которых имеет свои достоинства и недостатки: словарные методы, методы машинного обучения с учителем и методы машинного обучения без учителя» [Reagan et al. 2017]. Достаточно популярна точка зрения, согласно которой выделяют модели машинного обучения, словарную и гибридную [Poria et al. 2020; Rana et al. 2018]. Здесь не упомянуты

² Шаров С. А. Статистика слов в русском языке. *Lingvisto*. URL: http://lingvisto.org/artikoloj/ru_stat.html (дата обращения: 30.03.2023).

основанная на правилах модель (rule-based model). Однако основанное на правилах моделирование так или иначе признается актуальным, о чем все чаще говорится прямо [Пазельская, Соловьев 2011; Taboada et al. 2011]. И это целесообразно, поскольку «система анализа тональности ищет в рассматриваемом тексте слова, имеющие эмоционально-оценочный заряд, и, применяя заложенные в ней правила, учитывающие отрицание и слова-усилители, вычисляет тональность всего текста» [Кулагин 2021: 1108]. В любом случае в словарях отражаются человеческие знания в виде правил. Правилom в данном контексте является система индексов, согласно которой за каждой языковой единицей (лексемой, словосочетанием, N-граммой) закрепляется определенное цифровое значение оценки. Тональность всего документа определяется по совокупности индексов выявленной оценочной (эмоционально окрашенной) лексики, представленной в специализированном словаре.

Словари оценочной лексики, отражающие правила, имеют следующие номинации: *тональные словари, лексиконы, семантические тезаурусы* и др. Наиболее прозрачным и мотивированным в данной связи является термин *тональный словарь*. Такие словари разработаны для разных языков [Котельников и др. 2020]. Больше всего их разработано для английского языка³. В достаточном количестве они представлены и для русского языка⁴. Тональные словари могут содержать, например, «список слов», «список слов только с положительными и отрицательными аннотациями» или «список слов со скалярными числовыми значениями» [Agaque et al. 2019: 346]. Таким образом решаются задачи репрезентации сведений не только о номенклатуре языковых единиц, но и об их полярной ориентации и выраженности оценки. Особенно информативны сопоставления датасетов, способствующие объективации сведений (табл.⁵). Итак, чтобы проанализировать документ посредством основанной на правилах модели, нужен предварительно структурированный набор метаданных, что позволяет реализовать заложенные в них правила.

Статистическая модель (statistical model) сегодня практически безальтернативно базируется на методике машинного обучения. Обучение в данном контексте – обычная метафора. Смысл такого

Табл. Разница между данными из тональных словарей LabMT и ANEW

Tab. Sentiment dictionaries: LabMT vs. ANEW

Слово	h_{LabMT}	h_{ANEW}	h_{diff}
lust	4,64	7,12	1,72
bees	5,60	3,20	1.66
silly	5,30	7,41	1.43
engaged	6,16	8,00	1,20
book	7,24	5,72	1,15
hospital	3,50	5,04	1,15
evil	1,90	3,23	1,09
gloom	3,56	1,88	1,05
anxious	3,42	4,81	1,05
flower	7,88	6,64	1,00

обучения – в загрузке однотипных решений, на базе чего строится алгоритм решения следующих задач такого же типа. По сложившейся практике в машинном обучении выделяют два основных типа: без учителя (unsupervised learning) и с учителем (supervised learning) ([см., например, Полетаева 2020]). Несмотря на дифференциацию этих видов как самостоятельных моделей, они, по сути, являются вариантами одной хорошо известной статистической модели. О. Жерон называет другие виды машинного обучения: пакетное и динамическое обучение, обучение на основе образцов или на основе моделей, обучение с подкреплением и др. [Géron 2019]. Машинное обучение – популярная методика компьютерной лингвистики, используемая не только для оценки тональности текста. Процесс задействования статистической модели для сентимент-анализа похож на процесс ее использования при решении других задач автоматизированной обработки естественного языка.

В *гибридной модели* используется та или иная пропорция основанной на правилах и статистической моделей. Как правило, тот или иной выбранный приоритет хорошо комбинируется с другими подходами к анализу тональности [Paramonov, Poletaev 2021]. В свою очередь, приоритет использования какой-либо модели зависит от целой совокупности факторов: характера данных, возможностей их обработки, наличия программных ресурсов, компетенции исследователя и др.

Выбор оптимальной модели сентимент-анализа во многом определяется характером и степе-

³ LIWC. URL: <http://liwc.wpengine.com> (accessed 30 Mar 2023); Opinion Lexicon. URL: <https://www.kaggle.com/datasets/nltkdata/opinion-lexicon> (accessed 30 Mar 2023).

⁴ КартаСловСент. URL: <https://github.com/dkulagin/kartaslov> (дата обращения: 30.03.2023); PyСентиЛекс. URL: http://www.labinform.ru/pub/rusentilex/rusentilex_2017.txt (дата обращения: 30.03.2023).

⁵ Подробнее о словарях см. [Bradley, Lang 1999; Dodds et al. 2015; Reagan et al. 2017].

нью аннотирования материала. Теоретически есть все основания предполагать, что при должной подготовке инструментария модель, основанная на правилах, обладает большим потенциалом, чем статистическая модель, однако практика не так категорична. Усредненные показатели эффективности данных моделей вполне сопоставимы. При этом на практике результативность статистической модели сентимент-анализа зачастую оказывается более высокой: например, при изучении тональности текстов медиадискурса была достигнута точность 78,7 % [Kim, Novu 2006]; при изучении отзывов покупателей результативность составила 72,5 % [Bancken et al. 2014]. Общий разброс показателей эффективности сентимент-анализа (60–90 %) свидетельствует об их высокой зависимости от корректности самой процедуры. Существенна в данной связи роль качественной предобработки материала и каждого ее компонента в отдельности, что в информационно-компьютерной сфере во многом зависит от регламентации.

Данная последовательность действий (шагов), являясь, по сути, алгоритмом, рассматривается и как регламент. В процедуре предобработки материала могут быть и иные компоненты, их исчерпывающий перечень всегда индивидуален и зависит от задач при реализации того или иного проекта. Однако последовательность упомянутых шагов является логически и лингвистически обоснованной и целесообразной.

Заключение

Современная практика автоматизированной обработки естественного языка является сложной и многоаспектной. Несмотря на динамичное развитие сферы информационных технологий в прикладном русле, принципиальные основы такой деятельности

по-прежнему тесно коррелируют с лингвистической системой знаний: в методологическом контексте оценка тональности текста основывается на ярко выраженной лингвистической парадигматике. Вместе с тем вполне очевидна интердисциплинарная специфика используемых методик и моделей, выражающаяся в активном привлечении инструментария математической логики и статистики. Сентимент-анализ имеет комплексную методологическую базу, формирование которой обусловлено не только информационно-лингвистической, но и психолингвистической и прагмалингвистической спецификой. Это предопределяет поиск методологического баланса и оптимальной регламентации соответствующей деятельности. Особой целесообразностью использования лингвистического потенциала в процедуре оценки тональности текста характеризуется этап предобработки материала. Обоснованной представляется алгоритмическая организация процедуры предобработки материала с последовательным выявлением метаструктуры, определением референтности, уровневой ориентацией и выбором модели анализа. Из четырех перечисленных компонентов предобработки материала существенными являются все. Данная регламентация может быть использована для совершенствования процедуры сентимент-анализа и развития сферы автоматизированной обработки естественного языка в целом.

Конфликт интересов: Автор заявил об отсутствии потенциальных конфликтов интересов в отношении исследования, авторства и / или публикации данной статьи.

Conflict of interests: The author declared no potential conflict of interests regarding the research, authorship, and / or publication of this article.

Литература / References

- Баркович А. А. Интернет-дискурс: метаязыковые модели практики. *Вестник Волгоградского государственного университета. Серия 2, Языкознание*. 2015. № 5. С. 171–183. [Barkovich A. A. Internet discourse: metalanguage models of practice. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie*, 2015, (5): 171–183. (In Russ.)] <https://doi.org/10.15688/jvolsu2.2015.5.21>
- Большакова Е. И., Клышинский Э. С., Ландэ Д. В., Носков А. А., Пескова О. В., Ягунова Е. В. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика. М.: МИЭМ, 2011. 272 с. [Bolshakova E. I., Klyshinsky E. S., Lande D. V., Noskov A. A., Peskova O. V., Yagunova E. V. *Automatic processing of natural language texts and computational linguistics*. Moscow: MIEM, 2011. 272. (In Russ.)] <https://elibrary.ru/tdhfwd>
- Котельников Е. В., Разова Е. В., Котельникова А. В., Вычегжанин С. В. Современные словари оценочной лексики для анализа мнений на русском и английском языках (аналитический обзор). *Научно-техническая информация. Сер. 2. Информационные процессы и системы*. 2020. № 12. С. 16–33. [Kotelnikov E. V.,

- Razova E. V., Kotelnikova A. V., Vychezhzhanin S. V. Modern sentiment lexicons for opinion mining in English and Russian (analytical survey). *Nauchno-tekhnicheskaya informatsiya. Ser. 2. Informatsionnye protsessy i sistemy*, 2020, (12): 16–33. (In Russ.)] <https://doi.org/10.36535/0548-0027-2020-12-3>
- Кулагин Д. И. Открытый тональный словарь русского языка КартаСловСент. *Компьютерная лингвистика и интеллектуальные технологии: ежегодная Междунар. конф. «Диалог»*. (Москва, 16–19 июня 2021 г.) М.: РГГУ, 2021. Вып. 20. С. 1106–1119. [Kulagin D. I. Publicly available sentiment dictionary for the Russian language KartaSlovSent. *Computational linguistics and intellectual technologies: Annual Intern. Conf. "Dialogue"*, Moscow, 16–19 Jun 2021. Moscow: RSUH, 2021, iss. 20, 1106–1119. (In Russ.)] <https://doi.org/10.28995/2075-7182-2021-20-1106-1119>
- Майорова Е. В. О сентимент-анализе и перспективах его применения. *Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 6: Языкознание. Реферативный журнал*. 2020. № 4. С. 78–87. [Mayorova E. V. On sentiment analysis and prospects for its application. *Sotsialnye i gumanitarnye nauki. Otechestvennaya i zarubezhnaya literatura. Seriya 6: Yazykoznanie. Referativnyy zhurnal*, 2020, (4): 78–87. (In Russ.)] <https://www.elibrary.ru/tagobd>
- Пазельская А. Г., Соловьев А. Н. Метод определения эмоций в текстах на русском языке. *Компьютерная лингвистика и интеллектуальные технологии: ежегодная Междунар. конф. «Диалог»*. (Бекасово, 25–29 мая 2011 г.) М.: РГГУ, 2011. Вып. 10. С. 510–522. [Pazelskaya A. G., Solovyev A. N. A method of sentiment analysis in Russian texts. *Computational linguistics and intellectual technologies: Annual Intern. Conf. "Dialogue"*, Bekasovo, 25–29 May 2011. Moscow: RSUH, 2011, iss. 10, 510–522. (In Russ.)] <https://www.elibrary.ru/pjsrlj>
- Полетаева Н. Г. Классификация систем машинного обучения. *Вестник Балтийского федерального университета им. И. Канта. Сер.: Физико-математические и технические науки*. 2020. № 1. С. 5–22. [Poletaeva N. G. Classification of systems machine learning. *Vestnik of Immanuel Kant Baltic Federal University. Series: Physical-mathematical and technical sciences*, 2020, (1): 5–22. (In Russ.)] <https://www.elibrary.ru/rchveu>
- Семина Т. А. Анализ тональности текста: современные подходы и существующие проблемы. *Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 6: Языкознание. Реферативный журнал*. 2020. № 4. С. 47–64. [Semina T. A. Sentiment analysis: modern approaches and existing problems. *Sotsialnye i gumanitarnye nauki. Otechestvennaya i zarubezhnaya literatura. Seriya 6: Yazykoznanie. Referativnyy zhurnal*, 2020, (4): 47–64. (In Russ.)] <https://www.elibrary.ru/icgxzf>
- Харламов А. А., Ермоленко Т. В., Жонин А. А. Моделирование динамики процессов на основе анализа последовательности текстовых выборок. *Инженерный вестник Дона*. 2013. № 4. [Kharlamov A. A., Yermolenko T. V., Zhonin A. A. Processes dynamics modeling on the base of text corpus sequence analysis. *Inzhenernyy vestnik Dona*, 2013, (4). (In Russ.)] URL: <http://ivdon.ru/ru/magazine/archive/n4y2013/2047> (accessed 30 Mar 2023). <https://www.elibrary.ru/sblkbn>
- Чернышев М. В. Классификация тональности мнений для задачи автоматического сентимент-анализа текста. *Ученые записки УО «ВГУ им. П. М. Машерова»*. 2018. Т. 28. С. 136–140. [Chernyshevich M. V. Opinion classification for automatic sentiment analysis of the text. *Scientific notes of the Higher Educational Institution "VSU named after P. M. Masharov"*, 2018, 28: 136–140. (In Russ.)] <https://www.elibrary.ru/vxagrm>
- Araque O., Zhu G., Iglesias C. A. A semantic similarity-based perspective of affect lexicons for sentiment analysis. *Knowledge-Based Systems*, 2019, (165): 346–359. <https://doi.org/10.1016/j.knosys.2018.12.005>
- Bancken W., Alfarone D., Davis J. Automatically detecting and rating product aspects from textual customer reviews. *Interactions between Data Mining and Natural Language Processing: Proc. 1st Intern. Workshop (DMNLP 2014)*, Nancy, 15 Sep 2014. CEUR-WS, 2014, 1–16.
- Barkovich A. A. *Informational linguistics: the new communicational reality*. Cambridge: Cambridge Scholars, 2020, 271. <https://www.elibrary.ru/sjdauu>
- Beigi G., Hu X., Maciejewski R., Liu H. An overview of sentiment analysis in social media and its applications in disaster relief. *Sentiment analysis and ontology engineering: an environment of computational intelligence*, eds. Pedrycz W., Chen S.-M. Cham: Springer, 2016, 313–340. https://doi.org/10.1007/978-3-319-30319-2_13
- Bradley M. M., Lang P. J. *Affective Norms for English Words (ANEW): instruction manual and affective ratings*. Technical Report C-1. The Center for Research in Psychophysiology, University of Florida, 1999, 48.

- Dodds P. S., Clark E. M., Desu S., Frank M. R., Reagan A. J., Williams J. R., Mitchell L., Harris K. D., Kloumann I. M., Bagrow J. P., Megerdooian K., McMahon M. T., Tivnan B. F., Danforth C. M. Human language reveals a universal positivity bias. *PNAS*, 2015, 112(8): 2389–2394. <https://doi.org/10.1073/pnas.1411678112>
- Géron A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. 2nd ed. Sebastopol: O'Reilly Media, 2019, 856.
- Kim S.-M., Hovy E. Extracting opinions, opinion holders, and topics expressed in online news media text. *Proceedings of the Workshop on Sentiment and Subjectivity in Text (SST '06)*, Sidney, 22 Jul 2006. Stroudsburg: ACL, 1–8. <http://dx.doi.org/10.3115/1654641.1654642>
- Liu B. Corpus-based approach. *Sentiment analysis and opinion mining*. San Rafael: Morgan & Claypool, 2012, 95–99.
- Nasukawa T., Yi J. Sentiment analysis: capturing favorability using natural language processing. *Proceedings of the 2nd International Conference on Knowledge Capture (K-CAP'03)*, Sanibel, 23–25 Oct 2003. NY: ACM, 2003, 70–77. <https://doi.org/10.1145/945645.945658>
- Pang B., Lee L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2008, 2(1-2): 1–135. <http://dx.doi.org/10.1561/15000000011>
- Paramonov I. V., Poletaev A. Yu. Adaptation of semantic rule-based sentiment analysis approach for Russian language. *Proceedings of the 30th Conference of Open Innovations Association FRUCT*, Oulu, 27–29 Oct 2021. IEEE, 2021, 155–164. <http://dx.doi.org/10.23919/FRUCT53335.2021.9599992>
- Poria S., Hazarika D., Majumder N., Mihalcea R. Beneath the tip of the iceberg: current challenges and new directions in sentiment analysis research. *IEEE Transactions on Affective Computing*, 2020, 14(1): 108–132. <https://doi.org/10.1109/TAFFC.2020.3038167>
- Rana M. R. R., Nawaz A., Iqbal J. A survey on sentiment classification algorithms, challenges and applications. *Acta Universitatis Sapientiae, Informatica*, 2018, 10(1): 58–72. <http://dx.doi.org/10.2478/ausi-2018-0004>
- Reagan A. J., Danforth C. M., Tivnan B., Williams J. R., Dodds P. S. Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs. *EPJ Data Science*, 2017, 6. <https://doi.org/10.1140/epjds/s13688-017-0121-9>
- Taboada M., Brooke J., Tofiloski M., Voll K., Stede M. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 2011, 37(2): 267–307. http://dx.doi.org/10.1162/COLI_a_00049
- Wilks Y., Bien J. Beliefs, points of view, and multiple environments. *Cognitive science*, 1983, 7(2): 95–119. https://doi.org/10.1207/s15516709cog0702_1