

Оригинальная статья

УДК 81`322.2

<http://doi.org/10.32603/2412-8562-2025-11-4-121-138>

Семиотический анализ текстов и интерпретация знаковых систем в цифровую эпоху: Sentiment-анализ с использованием платформы KNIME

**Екатерина Владимировна Исаева¹✉, Сергей Владимирович Семенов²,
Денис Львович Черных³, Алексей Викторович Гудовщиков⁴**

^{1, 2, 3, 4}Пермский государственный национальный исследовательский университет, Пермь, Россия

¹✉ekaterinainsae@gmail.com, <https://orcid.org/0000-0003-1048-7492>

²ssemenov2002@gmail.com, <https://orcid.org/0009-0001-0610-5511>

³denis.1291.chernykh@gmail.com, <https://orcid.org/0009-0001-0704-2364>

⁴revandarh375@gmail.com, <https://orcid.org/0009-0004-4440-4303>

Введение. Целью статьи является изучение возможности интеграции семиотических подходов и методов машинного обучения для автоматизированного анализа тональности текстов (Sentiment-анализа). Sentiment-анализ текста является популярным направлением лингвистики на стыке с компьютерными науками и анализом данных. Новизна работы заключается в попытке интерпретации результатов машинного обучения с опорой на содержание текстов отзывов как знаковых систем, выявляя их лексические, синтаксические и прагматические характеристики.

Методология и источники. Исследование опирается как на фундаментальные основы семантики, синтактики и прагматики, так и на современные подходы к автоматизации обработки текстовой информации и применению математических методов для обоснования речевых явлений. Материалом исследования послужил свободно распространяемый набор данных отзывов на кинофильмы с платформы IMDB. В качестве инструмента автоматизации применяется система KNIME для анализа данных в парадигме «No-coding» (без кодирования). В статье представлен рабочий поток, включающий этапы предобработки данных, построения моделей классификации, а также оценки их эффективности, предложена лингвистическая интерпретация ошибок автоматической классификации отзывов.

Результаты и обсуждение. Результаты демонстрируют высокую точность классификации (до 92,0 %) и способность алгоритмов выявлять ключевые лексические и синтаксические маркеры, формирующие эмоциональную окраску текста. Исследование расширяет границы традиционной семиотики, интегрируя методы машинного обучения и анализа больших данных, а также подчеркивает практическую ценность использования KNIME в задачах обработки естественного языка.

Заключение. В статье дается детализированное описание алгоритма автоматизации Sentiment-анализа отзывов на кинофильмы с учетом преимуществ и потенциальных сложностей такого подхода для интерпретации текста. Перспективы дальнейших исследований включают применение предложенных методов к многоязычным корпусам и анализу мультимодальных данных, что открывает новые возможности для изучения знаковых систем в условиях цифровой коммуникации. Предложенная методика

© Исаева Е. В., Семенов С. В., Черных Д. Л., Гудовщиков А. В., 2025



Контент доступен по лицензии Creative Commons Attribution 4.0 License.

This work is licensed under a Creative Commons Attribution 4.0 License.

может найти применение в коммерческой сфере для выявления настроений пользователей товаров, услуг, приложений, книг, фильмов и т.д., что повышает интерес к лингвистической науке, а именно к автоматическому анализу тональности или Sentiment-анализу.

Ключевые слова: семиотика, тональность, анализ настроений, интерпретация, знаковые системы, лексические маркеры, машинное обучение, KNIME

Для цитирования: Семиотический анализ текстов и интерпретация знаковых систем в цифровую эпоху: Sentiment-анализ с использованием платформы KNIME / Е. В. Исаева, С. В. Семенов, Д. Л. Черных, А. В. Гудовщиков // ДИСКУРС. 2025. Т. 11, № 4. С. 121–138. DOI: 10.32603/2412-8562-2025-11-4-121-138.

Original paper

Semiotic Analysis of Texts and Interpretation of Sign Systems in the Digital Era: Sentiment-analysis Using the KNIME Platform

*Ekaterina V. Isaeva*¹✉, *Sergey V. Semenov*², *Denis L. Chernykh*³,
*Aleksey V. Gudovshikov*⁴

^{1, 2, 3, 4}Perm State University, Perm, Russia

¹✉ekaterinaisae@gmail.com, <https://orcid.org/0000-0003-1048-7492>

²ssemenov2002@gmail.com, <https://orcid.org/0009-0001-0610-5511>

³denis.1291.chernykh@gmail.com, <https://orcid.org/0009-0001-0704-2364>

⁴revandarh375@gmail.com, <https://orcid.org/0009-0004-4440-4303>

Introduction. The aim of the article is to study the feasibility of integrating semiotic approaches and machine learning methods for Sentiment-analysis. Sentiment-analysis is a popular area of linguistics at the interface with computer science and data analysis. The novelty of the paper lies in the attempt to interpret the results of machine learning based on the text of reviews as sign systems, revealing their lexical, syntactic, and pragmatic characteristics.

Methodology and sources. The research is based on the fundamental principles of semantics, syntactics, and pragmatics, as well as on modern approaches to the automation of textual information processing and the application of mathematical methods to substantiate speech phenomena. The research material is a freely distributed data set of film reviews from the IMDB platform. The KNIME system for data analysis in the 'No-coding' paradigm is used as an automation tool. The paper presents a workflow including the stages of data preprocessing, construction of classification models, and evaluation of their effectiveness, and proposes a linguistic interpretation of automatic review classification errors.

Results and discussion. The results demonstrate high classification accuracy (up to 92,0 %) and the ability of the algorithms to identify key lexical and syntactic markers that form the emotional colouring of the text. The study extends the boundaries of traditional semiotics by integrating methods of machine learning and big data analysis, and emphasises the practical value of using KNIME in natural language processing tasks.

Conclusion. This paper provides a detailed description of an algorithm for automating Sentiment analysis of film reviews, taking into account the advantages and potential challenges of this approach for text interpretation. Prospects for further research include applying the proposed methods to multilingual corpora and analysing multimodal data, which opens up new opportunities for studying sign systems in digital communication. The proposed methodology can be applied in the commercial sphere to identify the attitudes of

users to goods, services, applications, books, films, etc., which increases the interest in linguistic services, namely Sentiment analysis.

Keywords: semiotics, sentiment, sentiment analysis, interpretation, sign systems, lexical markers, machine learning, KNIME

For citation: Isaeva, E.V., Semenov, S.V., Chernykh, D.L. and Gudovshikov, A.V. (2025), "Semiotic Analysis of Texts and Interpretation of Sign Systems in the Digital Era: Sentiment-analysis Using the KNIME Platform", *DISCOURSE*, vol. 11, no. 4, pp. 121–138. DOI: 10.32603/2412-8562-2025-11-4-121-138 (Russia).

Введение. Современное общество, характеризующееся высокой интенсивностью интернет-коммуникаций, производит большие объемы текстовой информации. Каждый день миллионы пользователей выражают свои эмоции, мнения и оценки в цифровой среде – социальных сетях, каналах, форумах, блогах, платформах онлайн-обзоров. Эти тексты представляют собой сложные знаковые системы, отражающие социальные и культурные процессы. Для извлечения смысла и интерпретации эмоций, в них заложенных, одним из ключевых инструментов становится анализ тональности текста.

Этот метод исследования текста находит широкое применение в различных контекстах, например, для выявления экстремистских настроений с помощью лингвистического профилирования, что подчеркивает его значимость в обеспечении безопасности и профилактике деструктивного поведения [1]. Кроме того, алгоритмы машинного обучения позволяют эффективно анализировать большие текстовые данные, обеспечивая высокую скорость обработки и точность классификации эмоциональной окраски текстов [2].

С точки зрения семиотики, текст представляет собой знаковую структуру, где каждый элемент – от лексемы до синтаксической конструкции – участвует в создании общего смысла. Тональность текста, выражающая субъективное отношение автора к обсуждаемому объекту, формирует важный слой прагматической информации. Например, положительная или отрицательная эмоциональная окраска может влиять на восприятие продукта, бренда или общественного события.

Технологический прогресс, в частности в области обработки естественного языка (NLP), позволил автоматизировать процессы анализа текстов. Инструменты машинного обучения и платформы, такие как KNIME, предоставляют возможность эффективно обрабатывать большие массивы данных, выделяя ключевые признаки текста и определяя его эмоциональную окраску. Это особенно важно в условиях, когда объемы данных превышают возможности ручной обработки, а скорость принятия решений требует мгновенной интерпретации результатов, что обуславливает актуальность нашего исследования.

Тем не менее, не смотря на неоспоримые преимущества автоматизации в анализе эмоциональной окраски текстов, пользователи таких систем сталкиваются с рядом сложностей, вызванных спецификой анализируемых текстов и ограничениями применяемых методов. Основные вызовы включают трудности лингвистической интерпретации результатов автоматического анализа, адаптацию к особенностям различных языков и предметных областей, а также сложность обработки специфических текстовых форматов, таких как посты в социальных сетях.

Современные подходы стремятся преодолеть эти вызовы. Например, В. Себестьен и другие предложили многослойную нейросетевую модель для анализа документов, которая позволяет кластеризовать темы и оценивать их значимость, а также повышать точность ин-

терпретации. Такие методы, в частности их реализация в инструментах, подобных KNIME, предлагают перспективные решения благодаря интеграции визуализации, кластеризации и машинного обучения [3].

Таким образом, автоматический анализ тональности (будем называть его Sentiment-анализом, термином, включающим в себя признаки автоматизации и лингвистического анализа тональности), несмотря на значительные достижения, остаётся областью активных исследований. Он требует дальнейшей адаптации методов к специфике текстов, улучшения алгоритмов обработки контекста и разработки гибких моделей, способных учитывать культурные, языковые и жанровые особенности анализируемых данных.

Сложности интерпретации эмоциональной окраски текстов через автоматизированные методы, рассмотренные ранее, подчеркивают необходимость интеграции лингвистических, семиотических и вычислительных подходов для повышения точности и глубины анализа. Традиционные методы обработки текстов, основанные исключительно на машинном обучении, часто страдают от ограниченности контекстной интерпретации, что делает интеграцию междисциплинарных знаний особенно актуальной.

Лингвистический компонент играет ключевую роль в создании более интеллектуальных моделей анализа настроений. М. Табоада подчеркивает, что лексические, контекстуальные и структурные аспекты текста, такие как синтаксические конструкции и прагматические маркеры, являются важными элементами интерпретации настроений. Без учета этих факторов алгоритмы машинного обучения рискуют неверно классифицировать сложные высказывания, особенно содержащие сарказм, иронию или неоднозначные выражения [4].

В исследованиях Ф. Бенамара и других предложен подход, включающий динамическое использование контекстуальной информации для анализа настроений. Это особенно важно для анализа больших текстов, где эмоциональная окраска может меняться в зависимости от развития сюжета или дискурсивных целей автора. Такой интегративный подход позволяет моделям учитывать переходы между полярностями и субъективностью текста [5].

Современные исследования также акцентируют внимание на важности учета когнитивных и психологических процессов при анализе текста. В статье Р. Бэйли и других авторов предложен мета-фреймворк, объединяющий методы компьютерной лингвистики и машинного обучения с моделированием процессов человеческого чтения. Этот подход не только автоматизирует обработку текстов на поверхностном уровне (например, идентификация словесных паттернов), но и обеспечивает анализ на глубинном уровне, включая семантические и дискурсивные связи [6].

Интеграция лингвистических знаний также помогает решать проблемы, связанные с нехваткой маркированных данных. В своей диссертации на соискание степени доктора философских наук Б. Лу предлагает использовать методы полуконтролируемого обучения, которые комбинируют немаркированные данные и лингвистическую информацию для улучшения моделей. Это позволяет анализировать такие аспекты, как субъективность текста, полярность и идентификация носителя или цели мнения, что особенно полезно для языков с ограниченными размеченными данными [7].

Семиотический подход в данной области может стать основой для интерпретации текстов как знаковых систем, где каждое слово, синтаксическая конструкция или даже стили-

стический выбор автора имеют значение в определенном культурном и социальном контексте. Этот подход, дополненный лингвистическими знаниями, создает условия для более точного анализа текстов, которые интерпретируются не только с точки зрения их лексического наполнения, но и как части сложной знаковой структуры.

Таким образом, интеграция лингвистических, семиотических и компьютерных подходов позволяет не только преодолевать существующие ограничения автоматизированного анализа, но и создает возможности для расширения области применения Sentiment-анализа. Этот синтез знаний особенно актуален для текстов, насыщенных эмоциональной информацией, где междисциплинарность становится ключом к точной и контекстно обоснованной интерпретации.

Из представленного нами описания актуальности и проблем, связанных с Sentiment-анализом, вытекает и цель представляемого исследования – изучение потенциала аналитической платформы KNIME для автоматизированного анализа настроений текста с точки зрения семиотики. Это предполагает изучение возможностей платформы не только для обработки текстов и классификации их эмоциональной окраски, но и для интерпретации текстов как сложных знаковых систем.

Для достижения поставленной цели необходимо решить ряд задач, включающих следующие аспекты:

1) демонстрация алгоритмов и их реализации в контексте анализа знаков и их интерпретации. Наше исследование направлено на описание алгоритмического подхода к анализу настроений, реализуемого с помощью инструментов KNIME. Особое внимание уделяется тому, как платформа позволяет интерпретировать текстовые данные, подчеркивая связь между элементами текста и их эмоциональной семантикой;

2) оценка возможностей платформы для обработки больших текстовых массивов. Анализируется эффективность KNIME для работы с крупными наборами данных, включая такие аспекты, как скорость обработки, точность классификации и возможности визуализации. Это включает проверку алгоритмов на корпусе данных, например, отобранных отзывов с платформы IMDB для выявления сильных и слабых сторон платформы в контексте анализа текстов как знаковых систем.

Исследование основывается на междисциплинарном подходе, объединяющем семиотику, лингвистику и цифровые методы анализа данных, что позволяет рассматривать Sentiment-анализ как инструмент для глубокой интерпретации текстов в цифровую эпоху.

1. Теоретические основы семиотического Sentiment-анализа

1.1. Семиотическая природа текстов и Sentiment-анализа

Тексты, как и знаковые системы могут быть рассмотрены через призму семантики, синтактики и прагматики – подход, впервые введенный Ч. Моррисом, составляющий основу семиотики [8] и позволяющий установить семиотическую природу текстов. Так, семантика исследует связь между знаками и их значениями или референтами, синтактика изучает формальные отношения между знаками, а прагматика фокусируется на том, как знаки используются и интерпретируются людьми [9; 10]. Как отмечает О. А. Гриневич, эти три аспекта применимы к различным типам знаковых систем, в том числе естественные языки, научные тексты, литературные произведения и др. Например, в художественных текстах семантика отражает парадигматический уровень значений, синтактика – синтагматические связи, а прагматика – коммуникативные аспекты [11].

Анализ тональности текста как один из методов его автоматизированного изучения также может быть рассмотрен в перспективе этих трех измерений. С точки зрения семантики, Sentiment-анализ направлен на изучение субъективного содержания текста. В данном контексте в задачи исследователя входит извлечение значений, связанных с положительной, отрицательной или нейтральной эмоциональной окраской [12]. Для выявления сложных паттернов эмоций, таких как сарказм, потребуется изучение специальных словарей и глоссариев настроений, контекстуальный и дефиниционный анализ и исследование полисемии [13].

Синтаксический уровень позволяет выявить структурные отношения в тексте, такие как порядок слов и грамматические зависимости. Применение таких инструментов, как POS-тегирование (разметка частей речи), лемматизация (приведение к начальной форме) и автоматический синтаксический анализ позволяют выявить содержательные элементы текста, особенно в коротких и неформальных сообщениях, характерных для социальных сетей [14]. Например, автоматический синтаксический (парсинг) помогает выявить сложные связи между словами, что особенно важно при анализе контента, богатого идиомами или жаргонизмами.

Анализируя тональность текста с позиции прагматики, мы определяем его коммуникативную цель и углубляемся в социальный контекст сообщения, определяем намерения автора и интерпретируем неконвенциональные знаки, такие как эмодзи и другие символы, которые играют важную роль в передаче эмоций [15]. Следует также отметить важность учета таких прагматических особенностей, как сарказм, метафорический перенос или культурные различия в интерпретации знаков.

Таким образом, подход к текстам как к сложным многомерным знаковым системам, требующим учета их семантических, синтаксических и прагматических особенностей, позволяет проводить качественный Sentiment-анализ и добиваться высокой точности категоризации и интерпретации текстов по тональности в различных текстовых жанрах и доменах в их социальной и когнитивной динамике.

Анализ тональности находит свое применение в различных сферах: электронные рынки, бизнес-аналитика, исследование и обеспечение безопасности социальных сетей – везде, где существует взаимодействие участников коммуникации. Например, с помощью Sentiment-анализа можно выявить различия в восприятии текстовой информации разными целевыми аудиториями, что позволяет определить, как различные эмоциональные паттерны влияют на принятие решений [16]. Этот метод позволяет интерпретировать не только содержание текста, но и знаковые элементы, выражающие субъективность участников дискурса.

В бизнес-аналитике Sentiment-анализ используется в связке с инструментами сбора и репрезентации мнений, что делает его эффективным средством обработки больших массивов отзывов и комментариев [17]. Он помогает интерпретировать эмоциональные аспекты текстов, определять тематические направления и тенденции, что особенно важно для понимания поведения потребителей и оценки товаров.

Таким образом, мы рассмотрели Sentiment-анализ как процесс интерпретации текстовых знаков, выражающих субъективность и эмоциональность через определение семантических связей, особенностей построения текста и соотнесенность с контекстом и целью сообщения, что делает анализ тональности неотъемлемой частью современной теоретической и прикладной семиотики.

1.2. Платформа KNIME как инструмент семиотического анализа

Аналитическая платформа KNIME с открытым программным кодом позволяет «по кирпичикам» создавать и запускать рабочие процессы обработки данных [18], решать широкий спектр задач, включая анализ текстовых и графических данных. Это делает KNIME универсальным инструментом современного исследователя [19].

Для нас как аналитиков текстов особое значение имеет расширение KNIME «Text Processing Extension», которое предоставляет инструменты выполнения сложных задач в области обработки естественного языка и интеллектуального анализа текста. Это расширение необходимо для извлечения ключевой информации из текстов, построения списков частотности слов, анализа контекстов и выделения тем [20].

Большим преимуществом платформы является ее формат «No-coding» (без кодирования), содержащих библиотеки готовых и шаблонных решений и поддержку форума, что делает ее доступной для междисциплинарных исследователей, не владеющих навыками программирования и анализа данных. При этом KNIME не теряет своей привлекательности для профессионалов в области информационных технологий, которые могут добавлять собственные расширения, а ее кроссплатформенность обеспечивает стабильную работу в различных операционных системах [20].

При реализации семиотического анализа могут быть использованы инструменты автоматизированной текстовой аналитики, позволяющие не только быстро обрабатывать большие объемы данных, но и качественно их интерпретировать.

Одним из примеров использования KNIME для понимания и интерпретации данных является разработка онтологий представления оборудования и ключевых показателей эффективности (KPI) в компании Siemens. Платформа KNIME в этом случае использовалась для обработки и визуализации данных, что демонстрирует её пригодность для интерпретации сложных знаковых структур [21].

Семиотический подход к проектированию и управлению знаниями также может быть реализован на платформе KNIME. В частности, платформа используется для тематического моделирования на основе знаний. В качестве примера приведем наши более ранние работы по тематическому моделированию в области компьютерной безопасности. Мы продемонстрировали способность платформы к суммаризации больших объемов текста, кластеризации и тематическому моделированию для определения жанровой специфики текстов [22]. В другой работе был показан рабочий поток изучения новостных сообщений (RSS) на сайте Лаборатории Касперского с целью тематического моделирования новостных потоков. Однако, к нашему удивлению, в процессе эксперимента с помощью семантического анализа были выявлены признаки инфодемии [23].

В подобных исследованиях важна одновременная работа лингвистов и экспертов предметной области, направленная на порождение совместных знаний, повышение интерпретируемости результатов исследования и интеграции данных в кросс-доменных системах [24].

Методология и источники. *Методы автоматизированного семиотического исследования.* Рассмотрим наш эксперимент по реализации семиотического Sentiment-анализа на платформе KNIME.

2. Данные для семиотического Sentiment-анализа

В нашем эксперименте для анализа тональности используются текстовые данные из набора Large Movie Review Dataset v1.0, включающего 50 000 рецензий на фильмы на английском языке. Текстовый материал размечен тегами настроения двух типов: положительные и отрицательные. Этот набор получен из открытой базы датасетов, размещенных на платформе Keggler, и является эталонной выборкой для изучения методов обработки естественного языка (NLP) и Sentiment-анализа и дает хорошие результаты при разработке моделей классификации и регрессии. В нашем случае под классификацией понимается распределение текстов отзывов на положительные и отрицательные, а под регрессией – предсказание пользовательской оценки.

В нашем эксперименте была использована только часть описанного ранее набора данных, а именно 2000 текстовых документов с нормальным распределением: 1000 положительных и 1000 отрицательных отзывов. Каждая строка данных (будем называть это, как принято в цифровой лингвистике, вхождение) включает следующие параметры: сам текст отзыва, тег тональности (положительная или отрицательная) и ссылка на фильм на платформе IMDb.

Для реализации семиотического Sentiment-анализа мы сфокусировались на лексике, грамматических структурах и контекстах. Полный цикл NLP, в том числе обработка текстов отзывов о кинофильмах и визуализация результатов классификации, выполняется с помощью платформы KNIME. Отобранная нами подвыборка из 2000 вхождений загружается в рабочий поток KNIME с использованием узла Reader CSV, который считывает текстовые документы и метки тональности. Далее данные проходят предобработку, извлекаются семиотические признаки из полей «тексты отзывов» и «теги настроений». Эти этапы предобработки позволяют подготовить данные к дальнейшему анализу и классификации.

Данные IMDb хорошо подходят для Sentiment-анализа, так как рецензии представляют собой примеры текстов, где знаки активно используются для выражения субъективных оценок. Это позволяет рассматривать Sentiment-анализ как интерпретацию знаков, выражающих субъективные смыслы в цифровом пространстве. Эти элементы текста играют ключевую роль в интерпретации знаков и требуют анализа их значения в контексте.

Примеры из нашего датасета демонстрируют, как один и тот же знак может выражать различные эмоции в зависимости от окружающего контекста:

– *Unbelievable* (Невероятно):

Положительный отзыв: “*Her performance was unbelievable, such a joy to watch*” (Ее игра была невероятной, такое удовольствие смотреть).

Отрицательный отзыв: “*The level of incompetence in this movie is unbelievable*” (Уровень некомпетентности в этом фильме просто невероятен).

– *Fantastic* (Фантастический):

Положительный отзыв: “*The visuals in this film were absolutely fantastic, a real masterpiece*” (Визуальный ряд в этом фильме был просто фантастическим, настоящий шедевр).

Отрицательный отзыв с элементами сарказма: “*The script was fantastic... if you enjoy cliches and poor dialogue*” (Сценарий был фантастическим... если вам нравятся клише и плохие диалоги).

– *Dark* (Темный):

Положительный отзыв: “*The dark tone of the movie added depth and made it more intriguing*” (Мрачный тон фильма добавил глубины и сделал его более захватывающим).

Отрицательный отзыв: “*The picture was so dark in some scenes that it was hard to understand what was happening*” (Картинка была настолько темной в некоторых сценах, что было трудно понять, что происходит).

– *Classic* (Классический):

Положительный отзыв с выражением восхищения: “*This movie is a classic example of brilliant storytelling*” (Этот фильм – классический пример великолепного повествования).

Отрицательный отзыв с выражением критики: “*Another classic example of how not to make a movie*” (Очередной классический пример того, как не надо снимать фильмы).

Эти примеры иллюстрируют часто встречающиеся ситуации в работе с отзывами, в которых одни и те же слова могут иметь диаметрально противоположные значения в разных контекстах. В связи с этим для точной интерпретации таких знаков необходимо учитывать их окружение, грамматические особенности и жанровые характеристики.

3. Описание рабочего потока KNIME для Sentiment-анализа

Для реализации нашего эксперимента по семиотическому Sentiment-анализу мы используем рабочий поток «03_Sentiment_Classification rev 1» (рис. 1), доступный для скачивания в библиотеке открытых проектов на платформе KNIME [25]. Рассмотрим основные этапы рабочего потока для обеспечения воспроизводимости эксперимента другими лингвистами.

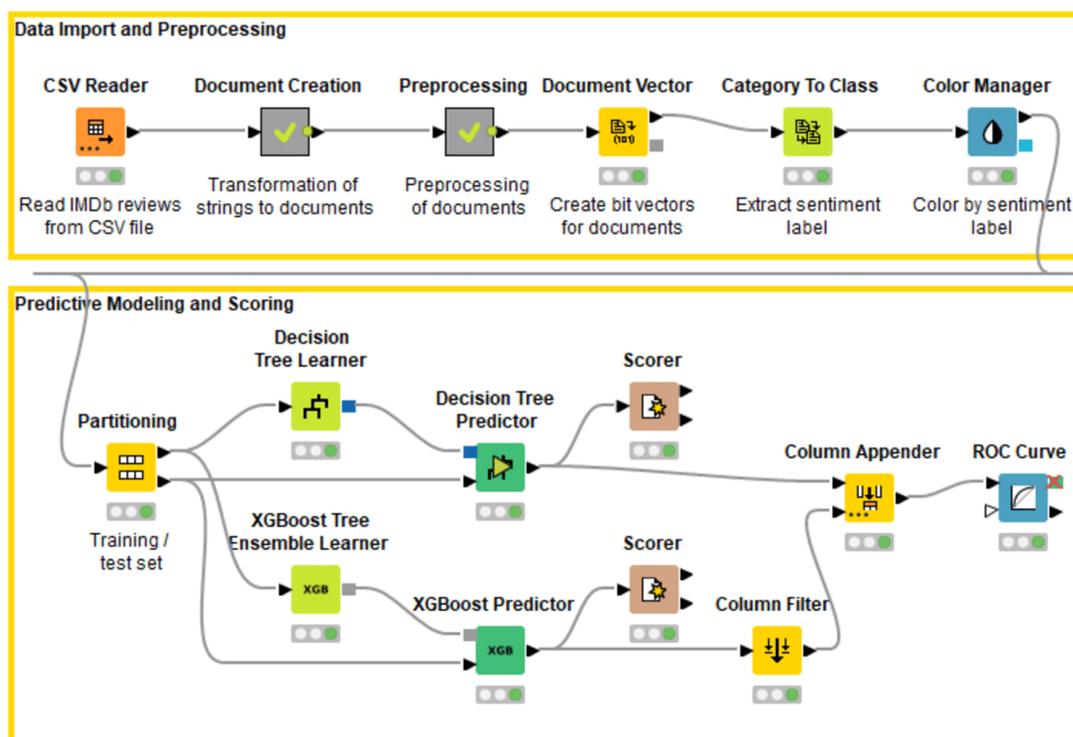


Рис. 1. Схема рабочего потока классификации текстов отзывов по тональности на платформе KNIME
Fig. 1. Workflow of classifying review texts by sentiment on the KNIME platform

На первом этапе текстовые данные, представленные в формате CSV, считываются с помощью узла CSV Reader и преобразовываются в текстовые документы, пригодные для семиотического анализа с помощью метаузла узла «Document Creation». Этот узел преобразует текстовые строки – отзывы IMDb – в специализированный формат, поддерживаемый KNIME, для дальнейшей обработки.

Таким образом, текст становится объектом семиотического анализа, где каждое слово, фраза или символ рассматривается как знак, обладающий семантическими, синтаксическими и прагматическими свойствами. Тексты приобретают матричную структуру, удобную для Sentiment-анализа.

Следующий метаузел «Preprocessing» отвечает за предобработку данных, направленную на устранение шумов, нормализацию текста и выделение значимых признаков. Этот метаузел включает в себя узлы (рис. 2):

- Punctuation Erasure (удаление пунктуации);
- Number Filter (удаление чисел);
- N Chars Filter (удаление коротких слов, длина которых меньше заданного числа n символов);
- Stop Word Filter (удаление стоп-слов, таких как предлоги, артикли, союзы, и другие высокочастотные слова);
- Case Converter (приведение всего текста к нижнему регистру);
- Snowball Stemmer (преобразование всех слов к их начальной форме).

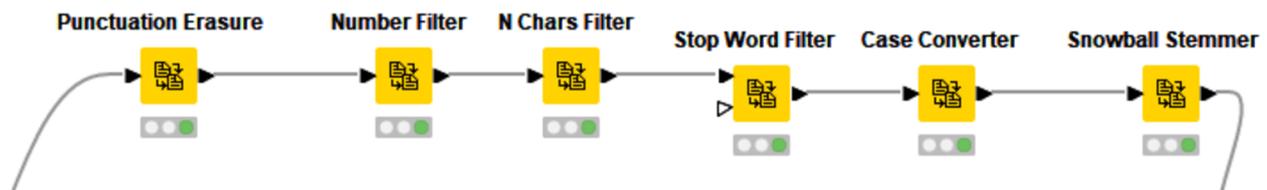


Рис. 2. Схема этапов предварительной обработки на платформе KNIME
Fig. 2. Stages of preprocessing on the Knime platform

Эти шаги предобработки позволяют сфокусироваться только на наиболее значимых смысловых элементах текста.

После этапа очистки следует узел «Document Vector», который преобразует текст в числовой формат – векторы (bag-of-words и TF-IDF), которые позволяют выделить ключевые знаки на основе частоты их появления и значимости в корпусе текстов.

Векторизация подразумевает составление своеобразного словаря всех слов корпуса текстов и последующее представление документов в виде нулей и единиц, которые обозначают наличие/отсутствие или степень значимости слова в документе (рис. 3).

#	RowID	Document	love	hook	time	wish	act	touch	realiti	think	spend	rest	life	event	meet
1	Row0	"	1	1	1	1	1	1	1	1	1	1	1	1	1
2	Row1	"	0	0	1	0	0	0	0	0	0	0	0	0	0
3	Row2	"	0	0	1	0	1	0	0	0	0	0	1	0	0
4	Row3	"	0	0	1	0	0	0	0	0	0	0	1	0	0
5	Row4	"	0	0	0	0	0	0	0	0	0	0	0	0	0

Рис. 3. Векторное представление данных на платформе KNIME
Fig. 3. Vector representation of data on the Knime platform

Таким образом, пройдя этапы предобработки, мы подготовили тексты для семиотического Sentiment-анализа. Подготовленные данные становятся удобным материалом для дальнейшей интерпретации и автоматизированного анализа тональности.

4. Построение моделей машинного обучения

В представленном рабочем потоке KNIME для категоризации отзывов по их эмоциональной окрашенности используются алгоритмы: Decision Trees и XGBoost.

Рассмотрим первый алгоритм классификации – Decision Trees. В представленном рабочем потоке узел Decision Tree Learner формирует древовидную структуру, где каждый узел (в терминах машинного обучения – лист), представляет собой одно из решений, связанных с характеристиками текста (например, наличие значимых для анализа слов или выражений). Затем на вход узлу Decision Tree Predictor подается подготовленная выборка и созданная модель, а на выходе мы получаем классифицированные на положительные и отрицательные наборы текстов.

Для сравнения используется еще один алгоритм – XGBoost, реализуемый с помощью узлов XGBoost Tree Ensemble Learner и XGBoost Predictor. Аналогично предыдущему алгоритму в потоке используются узел, ответственный за построение модели машинного обучения (XGBoost Tree Ensemble Learner), и узел, предсказывающий эмоциональную окраску отзыва (XGBoost Predictor). Этот алгоритм особенно эффективен для работы с большими текстовыми данными.

Оценивание точности обеих моделей выполняется с помощью узла Scorer по метрикам качества: точность (accuracy) и область под кривой ROC (Receiver Operating Characteristic, график для визуальной оценки качества бинарной классификации). Результаты обеих моделей сравниваются для выбора наиболее эффективного подхода.

Визуализация результатов классификации выполняет не только задачу распределения текстов на положительные и отрицательные, но и способствуют интерпретации знаков исследуемых текстов. Алгоритм Decision Trees позволяет интерпретировать текст, опираясь на таксономии ключевых слов и фраз, которые оказываются наиболее значимыми при классификации. Алгоритм XGBoost является менее интерпретируемым, но он демонстрирует высокую точность, позволяя учитывать сложные взаимосвязи между знаками, включая их сочетания и контекстуальность. Это обосновывает удачность сочетания выбранных алгоритмов для достижения баланса точности и интерпретируемости результатов Sentiment-анализа.

Результаты и обсуждение. Эффективность KNIME для Sentiment-анализа. Результаты классификации показывают, что оба алгоритма демонстрируют высокий уровень точности:

- алгоритм Decision Trees достиг точности в 91,3 %, что подтверждает способность модели выделять ключевые знаки и их значения для предсказания тональности;
- XGBoost Tree Ensemble обеспечил ещё более высокую точность – 92,0 %, благодаря использованию ансамблевого подхода, который учитывает сложные взаимосвязи между знаками.

Таким образом, использование платформы KNIME совместно с современными алгоритмами классификации обеспечивает точный анализ тональности текста.

Использованный в настоящем эксперименте рабочий поток на платформе KNIME позволил не только классифицировать тексты, но и продемонстрировать примеры успешной интерпретации знаков в тексте. Например, в положительных отзывах алгоритмы выявили

знаки *wonderful* (*прекрасный*) и *fantastic* (*фантастичный*), которые чаще всего встречались в контексте восхищения фильмами. В отрицательных отзывах ключевыми знаками оказались слова *boring* (*скучный*) и *disappointing* (*разочаровывающий*), которые отражают негативное отношение зрителей.

Взаимосвязь между знаковыми системами и эмоциональной окраской. Определение эмоциональной окрашенности текстов опирается на изучение сочетаний лексических и синтаксических маркеров, выражающих эмоции. Сфокусируемся прежде всего на лексических маркерах эмоций – отдельных словах и устойчивых выражениях, через которые автор отзыва чаще всего выражает свое субъективное отношение к объекту отзыва. Например, в положительно окрашенных отзывах часто встречаются слова *wonderful* (*замечательный*), *amazing* (*удивительный*), *fantastic* (*фантастический*), которые являются носителями семы восхищения или одобрения. Отрицательная тональность текста может быть выражена словами *disappointing* (*разочаровывающий*), *boring* (*скучный*) и *horrible* (*ужасный*), которые несут в себе сему неудовлетворенности или разочаровании. Некоторые слова, например, *unbelievable* (*невероятный*) и *fantastic* (*фантастический*), являются полисемичными, если рассматривать их как носителей тональности, т. е. в зависимости от контекста, они могут выражать как положительные, так и отрицательные эмоции.

Особенно ценна для лингвиста сводная таблица неверно классифицированных отзывов, которая дает материал для лингвистической интерпретации ошибок автоматического Sentiment-анализа (рис. 4).

Rows: 52 | Columns: 6 🔍 🏠

<input checked="" type="checkbox"/>	#	RowID	Ind... Num...	URL String	Text String	Document class String	Prediction (Docu... String	prediction Number (integer)
<input type="checkbox"/>	1	Row...	1574	http://www...	Although the beginning of the movie in New ...	NEG	POS	0
<input checked="" type="checkbox"/>	2	Row...	1212	http://www...	Care Bears Movie 2: A New Generation isn't a...	POS	NEG	0
<input type="checkbox"/>	3	Row...	3933	http://www...	The problem with portraying a real life individ...	NEG	POS	0
<input type="checkbox"/>	4	Row...	530	http://www...	The great talents of Michael Powell and Eme...	NEG	POS	0
<input type="checkbox"/>	5	Row...	10400	http://www...	Probably the finest fantasy film ever made. S...	POS	NEG	0

Рис. 4. Отчет об ошибках классификации
Fig. 4. Classification error report

Например, в выделенной нами строке 2 приводится положительный отзыв о фильме «Care Bears Movie 2: A New Generation», который автоматически был маркирован как отрицательный. Полагаем, что для алгоритма, в большей мере сфокусированном на семиотических признаках, проблемными оказались выражения отзыва, содержащие отрицания: *isn't at all a bad movie* (*совсем неплохой фильм*), *nothing wrong* (*ничего плохого*), *a real tearjerker* (*настоящая слезовыжималка*) и сложные синтаксические конструкции, как *Yes I admit the dialogue is corny and the story is a bit poorly told at times. But Darkheart, while very very dark is a convincing enough shape shifting villain, and Hadley Kay did a superb job voicing him.* (*Да, я признаю, что диалоги банальны, а история временами плоховата. Но Темное Сердце, хотя очень мрачный, но достаточно убедительный злодей, меняющий форму, и Хэдли Кей отлично справилась с его озвучиванием*) требуют глубокого анализа и интерпретации. С такими конструкциями, по нашему мнению, лучше справились бы нейросетевые модели, позволяющие реализовывать глубокое машинное обучение, учитывая контекст и на его основе извлекать смыслы.

В нашем исследовании не были рассмотрены синтаксические маркеры, так как они требуют особой предобработки и векторизации текста и не могут быть интегрированы в описанный нами рабочий процесс. Однако отметим это как перспективу наших дальнейших изысканий по вопросу Sentiment-анализа, так как синтаксические маркеры также играют немаловажную роль в передаче эмоций. Например, дополнительно может быть исследована инверсия и порядок слов в предложении. Такие выражения, как *“What an incredible experience this was!”* (*Что за невероятный опыт это был!*), помогают усилить эмоциональный эффект, а экспрессивные конструкции с повторами, словами-усилителями и дублирующимися восклицательными знаками, например, *“This movie is absolutely amazing!!!”* (*Этот фильм просто потрясающий!!!*), делают эмоциональную окраску отзыва более яркой. Отрицательные конструкции, такие как *“This was not what I expected at all”* (*Это было совсем не то, что я ожидал*), часто используются для передачи сарказма, усиливая отрицательную окраску отзыва.

Представленное исследование демонстрирует важность применения в семиотическом анализе методов машинного обучения, таких как реализуемые на платформе KNIME. Использование платформы позволило не только автоматизировать обработку больших текстовых массивов, но и применить семиотические принципы к интерпретации текстов как знаковых систем. Инструменты KNIME, позволяющие реализовать предобработку текста, векторизацию и классификацию, помогают выявлять ключевые знаки, отражающие эмоциональную окраску, и устанавливать их взаимосвязь с синтаксическими и семантическими элементами.

Кроме того, представленный эксперимент расширяет границы традиционного семиотического исследования, которое, как правило, сосредоточено на качественных методах анализа. Интеграция количественных подходов, таких как алгоритмы классификации, регрессии и кластеризации, позволяет изучать знаки в больших данных, открывая новые перспективы для семиотических исследований. Это сочетание новых технологий и классической теории подчеркивает роль современных платформ в развитии семиотики как научной дисциплины.

KNIME обладает рядом преимуществ по сравнению с другими платформами анализа данных и машинного обучения. Одной из ключевых особенностей является её доступность для пользователей с разным уровнем технической подготовки благодаря визуальному интерфейсу. Это позволяет исследователям, не обладающим глубокими навыками программирования, строить сложные рабочие процессы. Кроме того, интеграция с дополнительными модулями и библиотеками, такими как Text Processing, делает KNIME универсальным инструментом для анализа текстов как знаковых систем.

Тем не менее по сравнению с непосредственным написанием программ для реализации Sentiment-анализа, например, на языке программирования Python (с использованием библиотек NLTK, SpaCy, Hugging Face и др.), KNIME менее гибок в работе с продвинутыми моделями глубокого машинного обучения, которые могут учитывать синтаксические признаки, в частности, инверсию или порядок слов. Например, использование нейронных сетей трансформерного типа, таких как BERT или рекуррентные сети, позволяет более эффективно анализировать сложные семиотические структуры, включая контекстуальные и прагматические аспекты. Эти инструменты также позволяют использовать трансформеры для мультимодального анализа, включающего как текст, так и изображения, что делает их предпочтительными для задач, выходящих за рамки текстовой аналитики. В этом отношении

KNIME уступает платформам, где возможно непосредственное программирование и глубокая настройка моделей.

Результаты исследования показывают потенциал использования KNIME для дальнейших работ в области семиотики. Один из возможных векторов развития – применение платформы к многоязычным корпусам. Это позволит исследовать особенности знаковых систем в разных языках, выявлять универсальные маркеры эмоций и сравнивать их использование в различных культурных контекстах.

Еще одним направлением является анализ других знаковых систем, таких как визуальные или мультимодальные данные. Например, расширение возможностей KNIME для обработки изображений или видео (с использованием модулей для работы с мультимедиа) может открыть перспективы для анализа визуальных знаков и их сочетания с текстовыми элементами. Это особенно важно в условиях, когда мультимодальные знаковые системы играют ключевую роль в цифровой коммуникации.

Таким образом, исследование не только подтверждает эффективность KNIME для анализа текстовых знаков, но и подчеркивает перспективы интеграции платформы в более сложные семиотические задачи. Практические результаты могут быть использованы для оптимизации анализа клиентских отзывов в бизнесе, а также для разработки образовательных инструментов, обучающих навыкам интерпретации текста через знаковые системы.

Заключение. В данном исследовании продемонстрирована значимость интеграции платформы KNIME для анализа тональности текста в семиотическом аспекте. Основные результаты включают разворачивание рабочего процесса, сочетающего этапы предобработки данных, построение моделей классификации и оценку их эффективности. Использование KNIME позволило рассматривать тексты как сложные знаковые системы, в которых лексические, синтаксические и прагматические элементы формируют эмоциональную окраску.

Исследование вносит вклад в развитие методов анализа текста и семиотики, предложив новые подходы к интерпретации знаков. Практическая ценность заключается в том, что платформа KNIME показала себя как эффективный инструмент для обработки больших массивов данных и их дальнейшей интерпретации, включая выявление эмоциональной полисемии.

Для дальнейших исследований данного вопроса возможно рассмотреть перспективы расширения анализа на многоязычные корпуса, чтобы изучить особенности интерпретации знаков в различных культурных и языковых контекстах; использование платформы KNIME для анализа мультимодальных данных, таких как изображения или видео, для изучения визуальных знаков в сочетании с текстовыми.

СПИСОК ЛИТЕРАТУРЫ

1. Popova E. O., Volkova Y. A. Identification of Extremism Signs through the Analysis of the Text Tonality // Functional Aspects of Intercultural Communication. Translation and Interpreting Issues. Vol. 6: XI Int. Sci. Interdisciplinary Conf. on Research and Methodology, Moscow, 24 Nov. 2023 / RUDN Univ. Moscow, 2023. С. 66–76. DOI: 10.22363/2712-7974-2019-6-66-76.
2. Analysis of Tonality of Text Using Machine Learning / D. Gautham Sai, Govind Reddy S, D. Greeshma et al. // IJRASET. 2023. Vol. 11, iss. XII. P. 973–979. DOI: 10.22214/ijraset.2023.57492.
3. Baydogan C., Alatas B. Sentiment analysis using Konstanz Information Miner in social networks // 6th Int. Symposium on Digital Forensic and Security (ISDFS), Antalya, 22–25 March 2018 / IEEE. Antalya, 2018. DOI: 10.1109/ISDFS.2018.8355395.

4. Taboada M. Sentiment Analysis: An Overview from Linguistics // *Annual Review of Linguistics*. 2016. Vol. 2. P. 325–347. DOI: 10.1146/annurev-linguistics-011415-040518.
5. Benamara F., Taboada M., Mathieu Y. Evaluative Language Beyond Bags of Words: Linguistic Insights and Computational Applications // *Computational Linguistics*. 2017. Vol. 43, № 1. С. 201–264. DOI: 10.1162/COLI_a_00278.
6. A Meta-Framework for Modeling the Human Reading Process in Sentiment Analysis / R. Baly, R. Hobeica, H. Hajj et al. // *ACM Transactions on Information Systems*. 2017. Vol. 35, iss. 1: 7. DOI: 10.1145/2950050.
7. Lu B. On computing textual sentiment with linguistic knowledge and semi-supervised learning: Dr. Sci. (Philosophy) Thesis / Hong Kong. City Univ. of Hong Kong, 2013.
8. Stepanov Ju. S. Some Burning Issues of Contemporary Semiotics // *Linguistics*. 1974. Vol. 12, iss. 141. P. 53–66. DOI: 10.1515/ling.1974.12.141.53.
9. Veron E. Ideology and Social Sciences: A Communicational Approach // *Semiotica*. 1971. Т. 3, iss. 1. P. 59–76. DOI: <https://doi.org/10.1515/semi.1971.3.1.59>.
10. Allwood J. A Bird's Eye View of Pragmatics // *Papers from the Fourth Scandinavian Conference of Linguistics*. Odense: Odense Univ. Press, 1978. P. 145–159.
11. Гриневич О. А. Динамика функционирования усадебного сверткста русской литературы: семантика, синтактика, прагматика // *Известия Смолен. гос. ун-та*. 2020. № 1 (49). С. 46–60. DOI: 10.35785/2072-9464-2020-49-1-46-60.
12. Hogenboom A. Sentiment Analysis of Text Guided by Semantics and Structure. Rotterdam: Erasmus Univ. Rotterdam, 2009.
13. A New Approach for Carrying Out Sentiment Analysis of Social Media Comments Using Natural Language Processing / M. Ranjan, S. Tiwari, A. Md Sattar, N. S. Tatkar // *Engineering Proceedings*. 2023. Vol. 59, iss. 1: 181. DOI: 10.3390/engproc2023059181.
14. Vilares D. Sentiment analysis for reviews and microtexts based on lexico-syntactic knowledge. 2013. URL: <https://www.scienceopen.com/hosted-document?doi=10.14236/ewic/FDIA2013.8> (дата обращения: 27.01.2025).
15. Chauhan D., Sutaria K., Doshi R. Impact of Semiotics on Multidimensional Sentiment Analysis on Twitter: A Survey // *Second Int. Conf. on Computing Methodologies and Communication (ICCMC)*, 15–16 Feb. 2018 / *Erode. IEEE*, 2018. P. 671–674. DOI: 10.1109/ICCMC.2018.8487851.
16. Liebmann M., Hagenau M., Neumann D. Information Processing in Electronic Markets: Measuring Subjective Interpretation Using Sentiment Analysis // *ICIS 2012 Proceedings*. 2012. URL: <https://aisel.aisnet.org/icis2012/proceedings/KnowledgeManagement/9> (дата обращения: 27.01.2025).
17. Singh B., Kushwaha N., Vyas O. P. An interpretation of sentiment analysis for enrichment of Business Intelligence // *IEEE Region 10 Conference (TENCON)*, Singapore, 22–25 Nov. 2016 / Singapore, IEEE, 2016. P. 18–23. DOI: 10.1109/TENCON.2016.7847950.
18. Thiel K. Introduction to the KNIME Text Processing Extension // *Text Mining and Visualization: Case Studies Using Open-Source Tools*. 1st ed. / ed. by M. Hofmann, A. Chisholm. London; NY: Chapman and Hall, 2016. P. 63–80. DOI: <https://doi.org/10.1201/b19007>.
19. Meinl T., Jagla B., Berthold M. R. Integrated data analysis with KNIME // *Open Source Software in Life Science Research* / ed by L. Harland, M. Forster. Cambridge: Woodhead Publishing Ltd., 2012. pp. 151–171. DOI: <https://doi.org/10.1533/9781908818249.151>.
20. Dorr R. A., Casal J. J., Toriano R. Text Mining of Biomedical Articles Using the Konstanz Information Miner (KNIME) Platform: Hemolytic Uremic Syndrome as a Case Study // *Healthcare Informatics Research*. 2022. Vol. 28, no. 3. P. 276–283. DOI: 10.4258/hir.2022.28.3.276.
21. Towards Simplification of Analytical Workflows With Semantics at Siemens (Extended Abstract) / E. Kharlamov, G. Mehdi, O. Savkovic et al. // *IEEE Int. Conf. on Big Data (Big Data)*, Seattle, WA, 10–13 Dec. 2018 / IEEE. Seattle, 2018. P. 1951–1954. DOI: 10.1109/BigData.2018.8622652.

22. Smart Technologies for Genre Closeness Evaluation / E. Isaeva, O. Mnazhula, O. Baiburova, R. Crawford // Lecture Notes in Networks and Systems. 2022. Vol. 342, Cham: Springer. DOI: 10.1007/978-3-030-89477-1_60.

23. Isaeva E. Topic Modelling in Computer Security Discourse: a Case Study of Whitepaper Publications and News Feeds // Perm Univ. Herald. Russian and Foreign Philology. 2022. Vol. 14, iss. 2. P. 18–26. DOI: 10.17072/2073-6681-2022-2-18-26.

24. Valtolina S., Barricelli B. R., Dittrich Y. Participatory knowledge-management design: A semiotic approach // J. of Visual Languages & Computing. 2012. Vol. 23, iss. 2. P. 103–115. DOI: <https://doi.org/10.1016/j.jvlc.2011.11.007>.

25. 03_Sentiment_Classification rev 1 – KNIME Community Hub // KNIME Open for Innovation. 23.06.2024. URL: https://hub.knime.com/rfeigel/spaces/Public/03_Sentiment_Classification%20rev%201~4i6l8oqEGQ_ngBU5/current-state (дата обращения: 27.01.2025).

Информация об авторах.

Исаева Екатерина Владимировна – кандидат филологических наук (2013), доцент (2019), заведующая кафедрой английского языка профессиональной коммуникации Пермского государственного национального исследовательского университета, ул. Букирева, д. 15, г. Пермь, 614068, Россия. Автор 86 научных публикаций. Сфера научных интересов: дискурсивная лингвистика, когнитивное терминоведение, интеллектуальный анализ текста, цифровая лингвистика.

Семенов Сергей Владимирович – студент (4-й курс) направления «Лингвистика» Пермского государственного национального исследовательского университета, ул. Букирева, д. 15, г. Пермь, 614068, Россия. Сфера научных интересов: лингвистика, переводоведение, анализ тональности текста, Sentiment-анализ.

Черных Денис Львович – студент (4-й курс) направления «Лингвистика» Пермского государственного национального исследовательского университета, ул. Букирева, д. 15, г. Пермь, 614068, Россия. Сфера научных интересов: цифровая лингвистика, анализ тональности текста, Sentiment-анализ.

Гудовицков Алексей Викторович – студент (4-й курс) направления «Лингвистика» Пермского государственного национального исследовательского университета, ул. Букирева, д. 15, г. Пермь, 614068, Россия. Сфера научных интересов: лингвистика, переводоведение, анализ тональности текста, Sentiment-анализ, интерпретация текста.

Авторский вклад.

Исаева Екатерина Владимировна – замысел, разработка концепции и структуры исследования, общее руководство, подготовка теоретического обоснования семиотического исследования, подготовка текста.

Семенов Сергей Владимирович – подготовка модели машинного обучения для автоматизации анализа данных, анализ данных, подготовка текста.

Черных Денис Львович – подготовка теоретического обоснования использованных в исследовании методов Sentiment-анализа.

Гудовицков Алексей Викторович – семиотическая интерпретация результатов машинного обучения, подготовка текста.

*О конфликте интересов, связанном с данной публикацией, не сообщалось.
Поступила 28.01.2025; принята после рецензирования 05.03.2025; опубликована онлайн 22.09.2025.*

REFERENCES

1. Popova, E.O. and Volkova, Ya.A. (2019), "Identification of Extremism Signs through the Analysis of the Text Tonality", *Functional Aspects of Intercultural Communication. Translation and Interpreting Issues, Vol. 6: XI Int. Sci. Interdisciplinary Conf. on Research and Methodology*, Moscow, RUS, 24 Nov. 2023, pp. 66–76. DOI: 10.22363/2712-7974-2019-6-66-76.
2. Sai, D. Gautham, Reddy S. Govind, Greeshma, D. et al. (2023), "Analysis of Tonality of Text Using Machine Learning", *IJRASET*, vol. 11, iss. XII, pp. 973–979. DOI: 10.22214/ijraset.2023.57492.
3. Baydogan, C. and Alatas, B. (2018), "Sentiment analysis using Konstanz Information Miner in social networks", *6th Int. Symposium on Digital Forensic and Security (ISDFS)*, Antalya, TUR, 22–25 March 2018. DOI: 10.1109/ISDFS.2018.8355395.
4. Taboada, M. (2016), "Sentiment Analysis: An Overview from Linguistics", *Annual Review of Linguistics*, vol. 2, pp. 325–347. DOI: 10.1146/annurev-linguistics-011415-040518.
5. Benamara, F., Taboada, M. and Mathieu, Y. (2017), "Evaluative Language Beyond Bags of Words: Linguistic Insights and Computational Applications", *Computational Linguistics*, vol. 43, no. 1, pp. 201–264. DOI: 10.1162/COLI_a_00278.
6. Baly, R. et al. (2017), "A Meta-Framework for Modeling the Human Reading Process in Sentiment Analysis", *ACM Transactions on Information Systems*, vol. 35, iss. 1: 7. DOI: 10.1145/2950050.
7. Lu, B. (2013), "On computing textual sentiment with linguistic knowledge and semi-supervised learning". Dr. Sci. (Philosophy) Thesis, City Univ. of Hong Kong, Hong Kong, HKG.
8. Stepanov, Ju.S. (1974), "Some Burning Issues of Contemporary Semiotics", *Linguistics*, vol. 12, Iss. 141, pp. 53–66. DOI: 10.1515/ling.1974.12.141.53.
9. Veron, E. (1971), "Ideology and Social Sciences: A Communicational Approach", *Semiotica*, vol. 3, iss. 1, pp. 59–76. DOI: <https://doi.org/10.1515/semi.1971.3.1.59>.
10. Allwood, J. (1978), "A Bird's Eye View of Pragmatics", *Papers from the Fourth Scandinavian Conference of Linguistics*, Odense Univ. Press., Odense, DNK, pp. 145–159.
11. Grinevich, O.A. (2020), "Dynamics of the estate supertext functioning in Russian literature: semantics, syntactics, pragmatics", *Izvestia of Smolensk State Univ.*, no. 1 (49), pp. 46–60. DOI: 10.35785/2072-9464-2020-49-1-46-60.
12. Hogenboom, A. (2009), *Sentiment Analysis of Text Guided by Semantics and Structure*, Erasmus Univ. Rotterdam, Rotterdam, NDL.
13. Ranjan, M., Tiwari, S., Md Sattar, A. and Tatkar, N.S. (2023), "A New Approach for Carrying Out Sentiment Analysis of Social Media Comments Using Natural Language Processing", *Engineering Proceedings*, vol. 59, iss. 1: 181. DOI: 10.3390/engproc2023059181.
14. Vilares, D. (2013), *Sentiment analysis for reviews and microtexts based on lexico-syntactic knowledge*, available at: <https://www.scienceopen.com/hosted-document?doi=10.14236/ewic/FDIA2013.8> (accessed 27.01.2025).
15. Chauhan, D., Sutaria, K. and Doshi, R. (2018), "Impact of Semiotics on Multidimensional Sentiment Analysis on Twitter: A Survey", *Second Int. Conf. on Computing Methodologies and Communication (ICCMC)*, Erode, IND, 15–16 Feb. 2018, pp. 671–674. DOI: 10.1109/ICCMC.2018.8487851.
16. Liebmann, M., Hagenau, M. and Neumann, D. (2012), "Information Processing in Electronic Markets: Measuring Subjective Interpretation Using Sentiment Analysis", *ICIS 2012 Proceedings*, available at: <https://aisel.aisnet.org/icis2012/proceedings/KnowledgeManagement/9> (accessed 27.01.2025).
17. Singh, B., Kushwaha, N. and Vyas, O.P. (2016), "An interpretation of sentiment analysis for enrichment of Business Intelligence", *IEEE Region 10 Conference (TENCON)*, Singapore, SGP, 22–25 Nov. 2016, pp. 18–23. DOI: 10.1109/TENCON.2016.7847950.
18. Thiel, K. (2016), "Introduction to the KNIME Text Processing Extension", *Text Mining and Visualization: Case Studies Using Open-Source Tools*, 1st ed., in Hofmann, M. and Chisholm, A. (eds.), Chapman and Hall, London, NY, UK, pp. 63–80. DOI: <https://doi.org/10.1201/b19007>.

19. Meinl, T., Jagla, B. and Berthold, M.R. (2012), "Integrated data analysis with KNIME", *Open Source Software in Life Science Research*, in Harland, L. and Forster, M. (eds.), Woodhead Publishing Ltd., Cambridge, UK, pp. 151–171. DOI: <https://doi.org/10.1533/9781908818249.151>.

20. Dorr, R.A., Casal, J.J. and Toriano, R. (2022), "Text Mining of Biomedical Articles Using the Konstanz Information Miner (KNIME) Platform: Hemolytic Uremic Syndrome as a Case Study", *Healthcare Informatics Research*, vol. 28, no. 3, pp. 276–283. DOI: 10.4258/hir.2022.28.3.276.

21. Kharlamov, E., Mehdi, G., Savkovic, O. et al. (2018), "Towards Simplification of Analytical Workflows with Semantics at Siemens (Extended Abstract)", *IEEE Int. Conf. on Big Data (Big Data)*, Seattle, WA, USA, 10–13 Dec. 2018, pp. 1951–1954. DOI: 10.1109/BigData.2018.8622652.

22. Isaeva, E., Manzhula, O., Baiburova, O. and Crawford, R. (2022), "Smart Technologies for Genre Closeness Evaluation", *Lecture Notes in Networks and Systems*, vol. 342, Springer, Cham, CHE. DOI: 10.1007/978-3-030-89477-1_60.

23. Isaeva, E. (2022), "Topic Modelling in Computer Security Discourse: a Case Study of Whitepaper Publications and News Feeds", *Perm Univ. Herald. Russian and Foreign Philology*, vol. 14, iss. 2, pp. 18–26. DOI: 10.17072/2073-6681-2022-2-18-26.

24. Valtolina, S., Barricelli, B.R. and Dittrich, Y. (2012), "Participatory knowledge-management design: A semiotic approach", *J. of Visual Languages & Computing*, vol. 23, iss. 2. pp. 103–115. DOI: <https://doi.org/10.1016/j.jvlc.2011.11.007>.

25. "03_Sentiment_Classification rev 1 - KNIME Community Hub" (2024), *KNIME Open for Innovation*, 23.06.2024, available at: https://hub.knime.com/rfeigel/spaces/Public/03_Sentiment_Classification%20rev%201~4i6l8oqEGQ_ngBU5/current-state (accessed 27.01.2025).

Information about the authors.

Ekaterina V. Isaeva – Can. Sci. (Philology, 2013), Docent (2019), Head of Department of English for Professional Communication, Perm State University, 15 Bukireva str., Perm 614068, Russia. The author of 86 scientific publications. Area of expertise: discursive linguistics, cognitive term science, text mining, digital linguistics.

Sergey V. Semenov – Student (4th year, Linguistics), Perm State University, 15 Bukireva str., Perm 614068, Russia. Area of expertise: linguistics, translation studies, Sentiment-analysis.

Denis L. Chernykh – Student (4th year, Linguistics), Perm State University, 15 Bukireva str., Perm 614068, Russia. Area of expertise: digital linguistics, Sentiment-analysis.

Alexei V. Gudovshikov – Student (4th year, Linguistics), Perm State University, 15 Bukireva str., Perm 614068, Russia. Area of expertise: linguistics, translation studies, Sentiment-analysis, text interpretation.

Author's contribution.

Ekaterina V. Isaeva – conception, logic, and design of the study, general management, theoretical justification of semiotic research, text preparation.

Sergey V. Semenov – preparation of machine learning model for automation of data analysis, data analysis, text preparation.

Denis L. Chernykh – preparation of theoretical background of Sentiment-analysis methods used in the study.

Alexei V. Gudovshikov – semiotic interpretation of machine learning results, text preparation.

No conflicts of interest related to this publication were reported.

Received 28.01.2025; adopted after review 05.03.2025; published online 22.09.2025.