

АВТОМАТИЧЕСКАЯ ОБРАБОТКА РУССКОГО ЯЗЫКА В СОЦИАЛЬНЫХ МЕДИА

Антипова Е. А.

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Петрова Н.Е. – канд. филол. наук, доцент

В статье рассматриваются различные направления и аспекты автоматической обработки русского языка в сети Интернет. Основное внимание уделяется понятию социальных медиа, на примере которых описываются особенности машинной обработки русского языка.

Социальные медиа стали неотъемлемой частью нашей современной жизни и играют ключевую роль в коммуникативных привычках и поведении. Они представляют собой различные платформы и приложения, которые позволяют людям обмениваться информацией, создавать и делиться контентом, устанавливать связи с другими пользователями и поддерживать виртуальное общение. Примерами наиболее популярных сегодня социальных медиа являются Facebook, Instagram, Twitter, ВКонтакте, TikTok.

По нашему мнению, одним из важных аспектов социальных медиа является их способность связывать людей вокруг общих интересов, идей и ценностей. Они предоставляют возможность создавать и присоединяться к группам, сообществам и страницам, которые объединяют людей с общими интересами, например, группы по интересам, сообщества по спорту, фан-страницы популярных фильмов и т.д. Это способствует формированию виртуальных сообществ, где люди могут обсуждать, делиться опытом и находить единомышленников. Также стоит отметить, что социальные медиа превратились в важный источник новостей и информации. Многие пользователи получают свежие новости через свои ленты новостей и профили в социальных медиа. Это позволяет людям быть в курсе последних событий, развиваться профессионально, а также выражать свое мнение и участвовать в общественном диалоге [1].

В социальных медиа важную роль имеет автоматическая обработка русского языка. Она представляет собой использование компьютерных алгоритмов и искусственного интеллекта для анализа, классификации и интерпретации текстовой информации, размещённой в социальных медиа. Эта технология позволяет обрабатывать огромные объёмы текстовых данных, собираемых из различных источников, и извлекать из них ценные знания и понимание. Одним из ключевых аспектов автоматизированной аналитики русскоязычных сообщений в социальных медиа является анализ тональности текстов. Этот процесс не только определяет эмоциональную окраску высказываний пользователей (положительную, отрицательную или нейтральную), но и позволяет понять общественное мнение о различных продуктах, услугах, событиях или темах [2].

Автоматический анализ тональности позволяет компаниям и организациям получить ценные инсайты о реакциях пользователей на свои продукты или бренды. Например, одна компания, специализирующаяся на производстве косметических продуктов, использовала автоматическую обработку для анализа тональности отзывов о своих новых продуктах. Благодаря этому анализу, было обнаружено, что многие пользователи выражали положительные эмоции и высокую удовлетворённость новыми кремами для лица. Эта информация позволила компании эффективнее продвигать свою новую линию косметики и улучшить свою репутацию на рынке. Также было выявлено небольшое количество негативных отзывов, связанных с запахом одного из продуктов. Благодаря этому отзыву компания внесла коррективы в формулу продукта, улучшив его качество и удовлетворила требования клиентов. Исходя из этого, можно сделать вывод, что маркетологи могут использовать описанный выше анализ для измерения эффективности своих рекламных кампаний или для выявления негативных отзывов о продукте, чтобы принять меры и улучшить его.

Кроме анализа тональности, машинная обработка русского языка в социальных медиа включает в себя также классификацию тематик текстов. Это означает, что системы могут автоматически определить, о чём идет речь в тексте и какая тема или категория ему соответствует. Например, система может классифицировать текст как спорт, политику, развлечения или технологии. Это позволяет выявить популярные обсуждаемые темы, тренды и интересы пользователей. Компании могут использовать эту информацию для проведения маркетинговых исследований, анализа конкурентов и планирования своих маркетинговых кампаний. Они могут определить, какие темы наиболее популярны среди своей целевой аудитории и создавать контент, соответствующий этим интересам. Рассмотрим более детально: компания, занимающаяся производством

электроники, использовала машинную обработку для классификации тематик текстов, связанных с их продуктами. Анализ показал, что среди текстов, упоминающих их последнюю модель смартфона, наиболее популярными темами были камера, производительность и дизайн. Благодаря этим данным компания осознала, что пользователи активно обсуждают возможности камеры и требуют улучшений в этой области. В результате, они приняли решение усилить маркетинговую активность, сосредоточившись на преимуществах камеры в своих рекламных кампаниях и разработке новых функций, чтобы удовлетворить потребности своей целевой аудитории [3].

Другим важным аспектом обработки русскоязычных текстовых данных в социальных медиа с применением автоматических методов является определение ключевых слов и хэштегов. Алгоритмы автоматической обработки могут анализировать тексты и выделять наиболее часто употребляемые слова и фразы, которые могут быть использованы для классификации и поиска контента. Если речь идет о компании, занимающейся продажей электроники, и нужно провести исследование о популярных запросах и интересах своей целевой аудитории, то теперь можно использовать алгоритмы для анализа текстов, связанных с электроникой, и определения наиболее часто употребляемых слов и фраз. После анализа множества текстовых сообщений пользователей, система может выделить ключевые слова и фразы, например, "смартфон", "наушники", "телевизор", "ноутбук" и т.д. Эти ключевые слова и фразы могут указывать на популярные продукты или интересы пользователей в сфере электроники. Компания может использовать эту информацию для оптимизации своего контента и рекламных кампаний. Если система выявляет, что множество пользователей обсуждают "смартфоны с хорошей камерой", компания может решить создать контент, который подчеркивает качество камеры в их смартфонах и использовать соответствующие хэштеги в своих публикациях. Это поможет привлечь внимание пользователей, интересующихся этой функцией. Таким образом, анализ ключевых слов и хэштегов позволяет компаниям понять предпочтения и интересы своей целевой аудитории и адаптировать свой контент и маркетинговые стратегии соответственно [4].

Важным направлением автоматической обработки является распознавание именованных сущностей. Это означает, что системы могут автоматически распознавать и идентифицировать имена людей, организации, места, даты и другие специфические сущности в тексте. Например, если в тексте упоминается имя, система сможет распознать, что это имя человека, и даже может связать его с другими упоминаниями этого человека в других текстах. Также системы могут распознавать названия компаний, географические места, ключевые даты и другие важные сущности. Системы распознавания именованных сущностей могут помочь отслеживать события и тренды, связанные с какой-либо компанией. Например, компания выпустила новый продукт, система может автоматически распознать упоминания этого продукта в текстах и помочь оценить общественную реакцию на него. Это позволяет более точно измерить эффективность маркетинговых и рекламных кампаний и принять соответствующие меры для улучшения продукта или бренда [5].

Наконец, технология автоматического анализа включает в себя анализ социальных связей и взаимодействий между пользователями. Системы могут анализировать взаимодействия пользователей, такие как комментарии, лайки, репосты и другие действия, чтобы выявить связи и влиятельных пользователей. Один из способов применения этого анализа состоит в определении ключевых мнений линкующих пользователей. Система автоматической обработки русского языка может проанализировать взаимодействия пользователей и определить, кто из них имеет большую аудиторию, активно обсуждает какую-либо тему и имеет влияние на своих подписчиков. Можно использовать эту информацию для установления контакта с такими пользователями и предложения сотрудничества или рекламы продукта. Кроме того, анализ социальных связей и взаимодействий может помочь понять, как распространяется информация и формируются вирусные тренды. Система автоматической обработки может проанализировать взаимодействия пользователей, чтобы определить, как распространяется информация о продукте. Она может выявить, что некоторые пользователи имеют большое количество подписчиков и часто репостят или комментируют связанный с определенным продуктом контент. Можно использовать эту информацию, чтобы предложить этим пользователям дополнительные бонусы или скидки на продукт в обмен на их активную поддержку и рекламу.

В заключение отметим, что автоматическая обработка русского языка в социальных медиа играет важную и все более значимую роль в анализе текстовой информации, размещенной пользователями. С учетом быстрого развития социальных платформ и увеличения объема данных, генерируемых пользователями, становится всё сложнее обрабатывать и анализировать информацию вручную. Используя анализ тональности текстов, классификацию тематик, определение ключевых слов и хэштегов, распознавание именованных сущностей и т.п., автоматическая обработка русского языка способна анализировать огромные объемы текстовых данных, что позволяет повысить эффективность во многих сферах деятельности, в первую очередь в бизнесе и маркетинге. Однако очень важно при этом учитывать этические аспекты и обеспечивать защиту конфиденциальности данных пользователей при использовании автоматической обработки русского языка.

Список использованных источников:

1. Социальные медиа: что это, история, виды, платформы. Интернет [Электронный ресурс]. – Режим доступа: <https://www.mango-office.ru/products/calltracking/for-marketing/osnovy/socialnye-media/>. – Дата доступа: 03.04.2024.
2. Искусственный интеллект в обработке естественного языка: основные задачи, методы, применение. Интернет [Электронный доступ]. – Режим доступа: <https://nauchniestati.ru/spravka/ii-v-obrabotke-estestvennogo-yazyka/>. – Дата доступа: 03.04.2024.

3. Системы автоматической обработки текстов. Интернет [Электронный ресурс]. – Режим доступа: <https://www.osp.ru/os/2014/01/13039687>. – Дата доступа: 03.04.2024.

4. Методы и модели автоматического извлечения ключевых слов / С.О. Шереметьева, П.Г. Осминин [Электронный ресурс]. – Режим доступа: <https://cyberleninka.ru/article/n/metody-i-modeli-avtomaticheskogo-izvlecheniya-klyuchevyh-slov/viewer>. – Дата доступа: 03.04.2024.

5. Автоматическое распознавание именованных сущностей в текстах на русском языке. Интернет [Электронный ресурс]. – Режим доступа: https://revoition.aibest.ru/programming/01181016_0.html. – Дата доступа: 03.04.2024.