

Применение технологии NLP для анализа смыслов в текстах

Ярыгин Олег Витальевич

Кафедра информационной аналитики и политических технологий, Москва

E-mail: oyarygin.oleg@yandex.ru

Соавторы: Канаев А.А.

В современном мире информация окружает нас повсюду, и с каждым днем ее количество увеличивается в геометрической прогрессии. В связи с этим возникает необходимость в разработке новых технологий и методов обработки и анализа данных. Одним из таких методов является технология NLP (Natural Language Processing) – обработка естественного языка, которая представляет собой совокупность инструментов и алгоритмов для работы с текстами и другими формами человеческого общения.

NLP является междисциплинарной областью, объединяющей лингвистику, математику, информатику и другие научные дисциплины. Она позволяет автоматизировать процессы обработки естественного языка, включая извлечение информации из текстов, анализ тональности, определение тематики, категоризацию и кластеризацию документов.

Методы и подходы к анализу смыслов в NLP включают в себя различные алгоритмы и техники, такие как машинное обучение, статистический анализ, графовые модели, методы глубокого обучения и другие [5]. Каждый из них имеет свои преимущества и недостатки, и выбор конкретного подхода зависит от конкретной задачи и требований к точности и эффективности.

Одним из ключевых подходов в NLP является использование векторных представлений слов и предложений. Это позволяет представить текст в виде вектора, который можно использовать для сравнения, классификации и других операций [3]. Кроме того, используются также методы, основанные на правилах, такие как грамматики и словари.

NLP может использоваться для анализа смыслов текстов различными способами. Один из них — это извлечение ключевых слов и фраз из текста. Это может помочь определить основные темы или идеи, содержащиеся в тексте. Другой способ — это анализ тональности текста [4]. Это может быть полезно для определения отношения автора к определенной теме или предмету. Третий способ — это анализ семантических связей между словами в тексте. Это может помочь выявить логические связи между различными идеями в тексте. Наконец, NLP может использоваться для создания моделей языка, которые могут быть использованы для генерации текста или ответов на вопросы.

Одним из примеров использования технологии NLP является анализ отзывов клиентов о продукте или услуге. Алгоритм может проанализировать отзывы и определить, какие аспекты продукта наиболее популярны, а какие нуждаются в улучшении. Это позволяет компании принимать обоснованные решения о том, как улучшить свой продукт или услугу.

Технология NLP является актуальной для распознавания смыслов в тексте, так как она позволяет анализировать большой объем информации и извлекать из нее полезные данные [2]. В настоящее время существует множество приложений, которые используют NLP для обработки естественного языка, таких как чат-боты, системы автоматического ответа на электронные письма и системы рекомендации контента. Кроме того, NLP может быть использована для создания систем машинного перевода, которые позволяют переводить тексты на разные языки.

Изучение и использование NLP имеет важное значение для многих областей, включая обработку естественного языка, машинное обучение, информационный поиск и биоинформатику. Эта технология позволяет компьютерам понимать и анализировать человеческий язык, что делает ее ключевой для разработки систем искусственного интеллекта и машинного обучения [1].

Кроме того, NLP играет важную роль в анализе больших объемов данных, таких как социальные медиа, интернет-трафик и научные публикации. Это позволяет извлекать полезную информацию из этих данных и использовать ее для принятия решений в различных областях, таких как маркетинг, здравоохранение и образование.

Наконец, изучение NLP важно для понимания того, как компьютеры могут обрабатывать и понимать человеческий язык. Это знание может помочь улучшить существующие технологии и разработать новые, более эффективные методы обработки естественного языка.

Работа выполнена в соавторстве с Канаевым Алексем Александровичем

Студентом 1 курса магистратуры СГНЗ Московского государственного технического университета им. Баумана, Факультета социально-гуманитарных наук, Москва, Россия

E-mail: kanaevaa@student.bmstu.ru

Источники и литература

- 1) Рассел С., Норвиг П. Искусственный интеллект: современный подход (AIMA-2) // под. ред. Рассел С., Норвиг П. - М.: Вильямс, 2019. — 1408 с.
- 2) Ремарчук В. Н. Информационная аналитика: теория, методология, технологии : учебник для вузов / В. Н. Ремарчук. — Санкт-Петербург : Лань, 2022. — С. 143.
- 3) Будыльский Дмитрий Викторович, Подвесовский Александр Георгиевич. Исследование применимости моделей глубокого обучения для задачи аспектного анализа тональности текстовых сообщений // Транспортное машиностроение. 2015. №3 (47). URL: <https://cyberleninka.ru/article/n/issledovanie-primenimosti-modeley-glubokogo-obucheniya-dlya-zadachi-aspektnogo-analiza-tonalnosti-tekstovyh-soobscheniy> (дата обращения: 27.11.2023).
- 4) Катермина Татьяна Сергеевна, Тагиров Кадир Межвединович, Тагиров Тагир Межвединович. ЭЛЕМЕНТЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В РЕШЕНИИ ЗАДАЧ АНАЛИЗА ТЕКСТОВ // Computational nanotechnology. 2022. №2. URL: <https://cyberleninka.ru/article/n/elementy-iskusstvennogo-intellekta-v-reshenii-zadach-analiza-tekstov> (дата обращения: 27.11.2023).
- 5) Цитульский Антон Максимович, Иванников Александр Владимирович, Рогов Илья Сергеевич. NLP - обработка естественных языков // StudNet. 2020. №6. URL: <https://cyberleninka.ru/article/n/nlp-obrabotka-estestvennyh-yazykov> (дата обращения: 27.11.2023).