

УДК 004.89

**РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОГО ПОМОЩНИКА ДЛЯ ОЦЕНКИ
СМЫСЛОВОЙ ЦЕЛОСТНОСТИ НАУЧНОГО ТЕКСТА С ИСПОЛЬЗОВАНИЕМ
БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ**

Бондарев Ю.А. (Университет ИТМО)

**Научный руководитель – доктор технических наук, доцент Духанов А.В.
(Университет ИТМО)**

Введение. Оценка качества научных текстов важна как для автора, так и для рецензента. Имеются разные характеристики качества текста, одна из важнейших – связность текста. Связность текста непосредственно влияет на читаемость текста. Один из аспектов связности – смысловая целостность текста (англ. coherence), то есть его логическая завершенность и соответствие заданной тематике как текста в целом, так и каждого его раздела. Чем более целостен текст, тем более он понятен читателю [1]. В настоящее время смысловая целостность текста оценивается чаще всего вручную, реже - с помощью различных языковых моделей и сервисов на их основе. Однако, такие сервисы имеют ограничения в части функционала, например, не позволяют оценивать собственные тексты, не поддерживают иные форматы данных, кроме текстового, а также не дают объяснения своим выводам. В связи с этим задача разработки интеллектуального помощника для оценки смысловой целостности текста является актуальной.

Основная часть. Для оценки смысловой целостности текста могут быть использованы различные методы, включая методы NLP (обработки естественного языка) [2] и большие языковые модели [3], которые показывают несколько большую эффективность в сравнении с прочими методами. Поскольку целостность текста подразумевает и единую тематику текста, необходим анализ тематик всего текста и его разделов; и в этой задаче языковые модели показали свою эффективность [4]. Кроме того, для большей детализации оценки необходим анализ текста путем поиска в нем ответов на вопросы пользователя по содержанию текста. Для такого поиска также широко используются большие языковые модели [5]. Так как для каждой из приведенных задач используются особым образом обученные языковые модели, необходимо комбинирование языковых моделей для более эффективной работы с текстами статей.

Предлагается интеллектуальный помощник в виде чат-бота для Telegram, который является интерфейсом для взаимодействия с набором языковых моделей. Языковые модели используются для поиска ответов на вопросы по тексту, анализа тематик текста и ключевых слов. Кроме того, проводится оценка логической связности предоставленного текста.

Входными данными чат-бота будут являться текст статьи (или файл с текстом) и пользовательские промпты (вопросы). Выходные данные – ответы модели в текстовом виде с подробным описанием и примерами из текста статьи. Такой формат входных данных позволит работать с разными видами текста, а приведенный формат выходных данных даст пользователю детальный и объяснимый ответ на заданный вопрос.

Выводы. Разработан интеллектуальный помощник в виде чат-бота для оценки смысловой связности текста с помощью больших языковых моделей. Разработанный чат-бот позволяет проверять текст на соответствие определенной тематике, а также с помощью ответов на пользовательские вопросы позволяет оценить смысловую целостность текста. Реализованный функционал позволит уменьшить время рецензентов и авторов на оценку текста. В перспективе возможно использование разработанного чат-бота для научно-исследовательской деятельности в различных подразделениях Университета ИТМО. Кроме того, возможна интеграция разработанного чат-бота в проект Edulytica для частичной автоматизации процесса рецензирования научной статьи.

Список использованных источников:

1. Shaaron Ainsworth, Sarah Burcham. The impact of text coherence on learning by self-explanation // Learning and Instruction. - 2007. - №3, т. 17. - С. 286-303.
2. Brito, Jessica Oliveira; Oliveira, Elias de. Essays' Coherence Analysis Via Entity Grid Approach. In: Simpósio Brasileiro De Informática Na Educação (SBIE), 34. , 2023, Passo Fundo/RS. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2023 . p. 1431-1441. URL: <https://doi.org/10.5753/sbie.2023.235218>.
3. Abhishek, Tushar, Rawat Daksh, Gupta Manish, Varma Vasudeva. (2021). Transformer Models for Text Coherence Assessment. URL: <https://doi.org/10.48550/arXiv.2109.02176> (дата обращения 11.02.2025).
4. Wu X., Nguyen T., Luu A. T. A survey on neural topic models: methods, applications, and challenges //Artificial Intelligence Review. – 2024. – Т. 57. – №. 2. – С. 18.
5. Kurnia Muludi, Kaira Milani Fitria, Joko Triloka, Sutedi. Retrieval-Augmented Generation Approach: Document Question Answering using Large Language Model // International Journal of Advanced Computer Science and Applications. - 2023. - №3, т. 15

Автор _____ Бондарев Ю.А.

Научный руководитель _____ Духанов А.В.