

АВТОМАТИЗАЦИЯ КОРРЕКТИРОВКИ ТЕКСТА ПРИ ПОМОЩИ LLM

Нестеренков С. Н., Осадчий О. Э.

Кафедра информатики, Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь

E-mail: s.nesterenkov@bsuir.by, trapper.md@mail.ru

Исследование посвящено созданию автоматизированного сервиса коррективы резюме (CV) с применением технологий искусственного интеллекта и больших языковых моделей. Разработанное решение обеспечивает выполнение ключевых функций: грамматическая проверка, перефразирование текста, перевод с американского на британский английский и перевод на немецкий язык. Особенностью системы является возможность пакетной обработки, что делает её удобной при работе с большими объёмами данных. Практическая значимость исследования заключается в повышении качества подготовки резюме и расширении инструментов автоматизации не только в бизнес, но и образовательной, а также кадровой сфере.

ВВЕДЕНИЕ

Развитие современных технологий, особенно в области искусственного интеллекта и машинного обучения, привело к появлению и активному применению больших языковых моделей (LLM). Эти модели находят широкое использование в задачах обработки текста: от чат-ботов до специализированных сервисов для редактирования документов. Особую актуальность приобретает автоматизация коррективы резюме (CV), поскольку ошибки и несоблюдение языковых норм могут существенно снизить конкурентоспособность соискателя на рынке труда. Использование LLM позволяет решать задачи грамматической проверки, перефразирования и перевода текста, что открывает новые возможности для повышения качества подготовки резюме. Кроме того, внедрение таких технологий способствует экономии времени и ресурсов, особенно в корпоративной среде, где обрабатываются большие объёмы документов одновременно. В перспективе развитие LLM и интеграция с другими системами автоматизации могут позволить создавать более персонализированные и интеллектуальные сервисы для улучшения качества работы пользователей.

I. ПРИНЦИПЫ РАБОТЫ LLM. SELF-ATTENTION

Большие языковые модели (LLM) основаны на архитектуре трансформеров, ключевым элементом которой является механизм self-attention. Он обеспечивает определение взаимосвязей между всеми словами во входной последовательности, независимо от их удалённости друг от друга. В отличие от рекуррентных сетей, работающих последовательно, self-attention [1] позволяет обрабатывать текст параллельно, выделяя наиболее значимые фрагменты для текущего контекста. Это обеспечивает высокую точность при решении задач обработки естественного языка и делает LLM универсальными для задач проверки, перевода и перефразирования текста [1,2]. Схема механизма внимания представлена на рисунке 1.

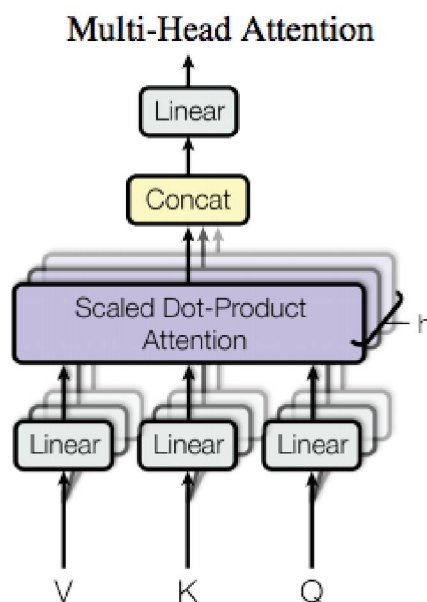


Рис. 1 – Схема механизма внимания

Суть использования данного механизма в содержании сразу нескольких экземпляров механизмов внимания, которые будут принимать входные данные, обрабатывать их, находить взаимосвязи, предоставлять коэффициенты важности слов, а дальше, исходя из слов и их коэффициентов определять их важность и смысл исходного входного текста

II. АВТОМАТИЗАЦИЯ РАБОТЫ С ТЕКСТОМ

Использование LLM в задаче коррективы резюме позволяет автоматизировать ключевые этапы обработки текста [3]. К ним относятся выявление и исправление грамматических ошибок, замена стилистически неудачных конструкций, адаптация формулировок под стандарты академического или делового английского языка. Дополнительно реализуется возможность перевода текста между языками и вариантами английского,

что повышает адаптивность резюме для разных рынков труда. Кроме того дополнительной степенью автоматизации является встраивание сервиса корректировки в сервис создания резюме, либо любой другой, что позволяет пользователю не отрываться в другие места для корректировки текста, а работать в одном удобном месте.

III. ПАКЕТНАЯ ОБРАБОТКА И МАСШТАБИРУЕМОСТЬ

Сервис поддерживает не только работу с отдельными предложениями, но и пакетную обработку больших текстовых массивов [4]. Это особенно важно для корпоративных пользователей и кадровых агентств, где требуется обработка множества документов. Масштабируемость достигается за счёт модульной архитектуры и оптимизации вызовов языковых моделей, что снижает время отклика и нагрузку на вычислительные ресурсы.

Кроме того, отдельным ключевым элементом масштабируемости является добавление любой другой языковой модели, в зависимости от потребностей клиента. Такие изменения не будут ломать систему, а наоборот смогут плавно в неё встроиться.

IV. ОБЕСПЕЧЕНИЕ КОНФИДЕНЦИАЛЬНОСТИ ДАННЫХ

Одним из приоритетов при разработке системы является защита персональной информации. Использование локально развернутых моделей исключает необходимость передачи данных во внешние облачные сервисы, что минимизирует риск утечки конфиденциальных сведений [5]. Дополнительно предусмотрена возможность дообучения модели на закрытых корпоративных датасетах, что повышает качество работы и учитывает специфику конкретных организаций.

V. СКОРОСТЬ РАБОТЫ СИСТЕМЫ

Производительность разработанного сервиса была протестирована на GPU NVIDIA Tesla T4 с 15 ГБ видеопамяти. Тестирование показало: время обработки предложений варьируется от 0,2 до 1,4 секунды в зависимости от длины текста и сложности структуры. Средняя длина предложений составила 60 символов, при этом среднее время обработки одного предложения на стенде с использованием CUDA составило примерно 0,65 секунды. Результаты тестирования представлены на рисунке 2.

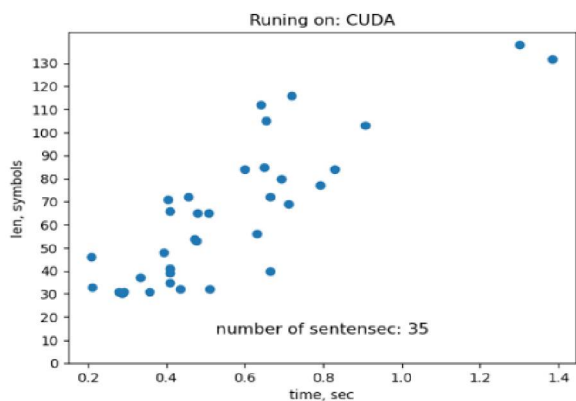


Рис. 2 – Скорость работы системы

VI. ИНТЕРФЕЙС ПОЛЬЗОВАТЕЛЯ И УДОБСТВО РАБОТЫ С СЕРВИСОМ

Сервис не имеет интерфейса в принципе, однако благодаря этому, микросервисной архитектуре и открытому API он может быть интегрирован в абсолютно любое корпоративное (частное) приложение [6], а также в корпоративную инфраструктуру – это увеличивает возможности по переиспользованию сервиса, и его интеграции в экосистему компании. Кроме того, продукт может быть использован отдельно, как облачный сервис [7].

VII. ПРАКТИЧЕСКАЯ ЗНАЧИМОСТЬ И ПЕРСПЕКТИВЫ

Предложенный сервис может использовать индивидуальные соискателями, IT-фирмами, образовательными учреждениями и др.. Его внедрение способствует сокращению времени подготовки резюме, повышению их качества и конкурентоспособности на международном рынке труда. В перспективе возможно расширение языковой базы, интеграция с платформами для трудоустройства и адаптация под другие виды документов (мотивационные письма, академические статьи).

1. Ротман, Д. *Transformers for Natural Language Processing* / D. Rothman. – Packt Publishing, 2021.
2. Голдберг, Й. *Нейросетевые методы в обработке естественного языка* / Й. Голдберг. – О’Рейли, 2017.
3. Батура, Т. В. *Математическая лингвистика и автоматическая обработка текстов* / Т. В. Батура. – РИЦ НГУ, 2016.
4. Рашка, С. *Строим LLM с нуля* / С. Рашка. – Питер, 2025.
5. Акст, Р. *Языковые модели и мир будущего, или Путеводитель по ChatGPT* / Р. Акст. – ЛитРес, 2023.
6. Гоф, Дж., Брайант, Д., Оберн, М. *Проектирование архитектуры API* / Дж. Гоф, Д. Брайант, М. Оберн // 2024.
7. Одиноченко, М. И. Использование облачных сервисов для решения задач, связанных с применением технологии BIG DATA / М. И. Одиноченко, С. Н. Нестеренков // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сб. научных статей VIII Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 2022 год) : / редкол. : В. А. Богуш [и др.]. – Минск : Бестпринт, 2022. – С. 252-256.