

РАЗРАБОТКА ВОПРОСНО-ОТВЕТНОЙ СИСТЕМЫ НА ОСНОВЕ LLM В КОНТЕКСТЕ ПРОЦЕССОВ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ УНИВЕРСИТЕТА

Щавлев К.В. (ИТМО)

**Научный руководитель – кандидат технических наук, доцент Кугаевских А.В.
(ИТМО)**

Введение

В эпоху активной цифровизации образовательных процессов университеты сталкиваются с растущей потребностью в интеллектуальных системах, которые обеспечивают доступ к учебно-методической и нормативной документации образовательных программ [1–2]. Современные достижения в области машинного обучения и обработки естественного языка открывают возможности для автоматизации анализа и интерпретации подобных данных. Однако на практике становится очевидным, что эти технологии требуют существенной адаптации к специфике предметной области, а также к особенностям используемых источников информации [3]. Особого внимания заслуживает перспективное направление — разработка вопросно-ответных систем. Такие системы объединяют мощные языковые модели с механизмами извлечения релевантного контекста как из структурированных, так и из неструктурированных данных. Этот подход не только отвечает актуальным запросам образовательной среды, но и подтверждает значимость поставленной задачи.

Основная часть

Была спроектирована архитектура системы, которая включала в себя:

- LLM-Gateway, сервис, генерирующий ответ на основе имеющихся в системе знаний;
- RAG, сервис с векторной базой данных, хранящий в ней все документы, полезные студентам одной образовательной программы;
- Embedder, сервис, генерирующий векторное представление текста;
- Frontend, точка входа для работы с системой;
- Скраперы текста, извлекающие текст из документов различных форматов.

Система поддерживает два способа использования — «менеджер» и «студент». В рамках студенческого использования можно применять чат-бот как способ получения необходимой информации по документам университета. Со стороны менеджера же — возможность управления базой знаний, добавления документов и просмотра статистики. Архитектура системы была реализована в соответствии с основными идеями, выдвигаемыми ранее [4].

Выводы

Экспериментальная проверка подтвердила эффективность интеграции RAG с языковой моделью, выразившуюся в повышении точности ответов и снижении числа галлюцинаций. Практическая значимость работы заключается в создании прикладного решения для автоматизации работы с документацией образовательных программ и поддержки пользователей при получении достоверной информации.

Литература

1. Jae-Kwang Oh, Soo Kyun Kim: Design of a RAG-based question-answering system for university undergraduate administrative assistance, 2025. DOI: 10.9708/jksci.2025.30.07.079
2. Wayne Matthew A. Dayata, Sabrinah Yonell C. Yap: Carolinian Chatbot: End-to-End Question-Answer Pipeline on the University of San Carlos Policies, Procedures, and Guidelines Via Retrieval Augmented Generation, 2024. DOI: 10.1109/COMNETSAT63286.2024.10862750
3. Zejiang Shen, Ruochen Zhang, Melissa Dell, Benjamin Charles Germain Lee, Jacob Carlson, Weining Li: LayoutParser: A Unified Toolkit for Deep Learning Based Document Image Analysis, 2021. DOI: <https://arxiv.org/abs/2103.15348>
4. Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, Douwe Kiela: Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, 2020,