

РАЗРАБОТКА СИСТЕМЫ АВТОМАТИЧЕСКОЙ ГЕНЕРАЦИИ ВОПРОСОВ НА ОСНОВЕ ГРАФА ЗНАНИЙ WIKIDATA

Пеньков Г. Д.¹

Научный руководитель – доцент ФПИИ Валитова Ю. О.¹

¹Университет ИТМО

penkov.gd@gmail.com

Введение

В условиях роста популярности электронного обучения (e-learning) и самообразования создание качественного контента для проверки знаний занимает важное место. Существующие системы автоматической генерации вопросов сталкиваются с рядом ограничений: генерация по шаблонам выдает однотипные вопросы, а использование исключительно LLM может приводить к фактическим ошибкам (галлюцинациям) [1]. Зарубежный опыт показывает успешное применение графов знаний в генерации вопросов для проверки знаний обучающихся на основе семантических данных Wikidata [2]. Однако, такие решения узкоспециализированы (например, для проверки школьных знаний по физике) и не обладают гибкостью в выборе тематик квизов. Таким образом, актуальной является задача создания системы, способной генерировать достоверные, разнообразные и человекочитаемые вопросы в различных предметных областях.

Основная часть

Разрабатываемое решение представляет собой микросервисную систему для автоматической генерации вопросов и прохождения квизов. Ядром системы выступает подход Graph RAG (Retrieval-Augmented Generation) на основе графа знаний Wikidata. Также используется SKOS (Simple Knowledge Organization System) для гибкой классификации тем квизов. Техническая реализация включает в себя:

- 1) ETL-пайплайн для импорта семантических данных из Wikidata в графовую базу данных Neo4j;
- 2) модуль генерации вопросов (Graph RAG);
- 3) администрирование (управление контентом, классификацией) по принципу “Data as Code”, который состоит в том, что один yaml-файл представляет собой одну тематику (например, “Живопись итальянского Возрождения”), включающую в себя SPARQL-запрос к Wikidata Query Service и SKOS-классификацию этой тематики;
- 4) веб-приложение как средство взаимодействия пользователей с системой (прохождение квизов).

Выводы

Предложенное решение автоматической генерации вопросов демонстрирует ряд преимуществ. Во-первых, использование графа знаний Wikidata в связке с Graph RAG позволяет генерировать разнообразные вопросы, которые основываются на реальных знаниях, что уменьшает галлюцинации. Также вопросы получаются более живыми и человекочитаемыми по сравнению с генерацией по шаблонам. Использование SKOS позволяет сделать системную и при этом гибкую классификацию тематик квизов. В перспективе данное решение можно интегрировать в системы электронного образования и приложений для самообразования, что демонстрирует разработанное веб-приложение.

Литература

1. Zhu Z., Huang T., Wang K. et al. Graph-based Approaches and Functionalities in Retrieval-Augmented Generation: A Comprehensive Survey // arXiv. 2025. DOI: 10.48550/arXiv.2504.10499.
2. Scharpf P., Schubotz M., Spitz A. et al. Collaborative and AI-aided Exam Question Generation using Wikidata in Education // arXiv. 2022. DOI: 10.48550/arXiv.2211.08361.