Hybrid Knowledge-Graph-Enhanced Retrieval-Augmented Generation for Academic Information Systems

Abstract

Academic information hosted on college websites is largely unstructured, heterogeneous, and distributed across multiple webpages, making precise information retrieval difficult using traditional keyword-based search techniques. While retrieval-augmented generation (RAG) has shown effectiveness in grounding large language model outputs using semantic vector retrieval, purely embedding-based approaches often fail to capture explicit relationships among entities such as departments, courses, and faculty, leading to incomplete or contextually weak responses.

This mini project proposes a hybrid knowledge-graph-enhanced retrieval-augmented generation framework for academic information access. Unstructured institutional content is processed through text cleaning, chunking, and semantic embedding, and indexed in a vector database to support similarity-based retrieval. In parallel, a lightweight knowledge graph is constructed to explicitly represent structured academic entities and their relationships. An intent-aware retrieval mechanism dynamically selects semantic retrieval, graph-based reasoning, or a hybrid of both to construct context for generation. The retrieved evidence is supplied to a large language model to generate grounded responses constrained to institutional knowledge. This design is motivated by recent journal studies demonstrating that integrating knowledge graphs into RAG pipelines improves factual grounding, relational reasoning, and interpretability, while domain-specific RAG systems enhance retrieval reliability in academic settings [1, 2].

Sustainable Development Goals (SDGs): This mini project supports SDG 4 (Quality Education) by improving structured access to academic information and SDG 9 (Industry, Innovation and Infrastructure) through the application of hybrid intelligent information retrieval systems.

Keywords: Retrieval-Augmented Generation, Knowledge Graph Integration, Vector Embeddings, Vector Database Indexing, Knowledge-Grounded Generation

Group Members:

- Julia Mariam John (B23CS2137)

- Nirmel B Joseph (B23CS2148)

- Rohith NS (B23CS2156)

Guide: Mr.Praveen J.S, Assistant Professor, Department of Computer Science and Engineering

References

[1] J. Linders and J. M. Tomczak, "Knowledge graph-extended retrieval augmented generation for question answering," *Applied Intelligence*, vol. 55, no. 17, pp. 1102–1118, 2025.

[2] Z. Li, Z. Wang, W. Wang, K. Hung, H. Xie, and F. L. Wang, "Retrieval-augmented generation for educational application: A systematic survey," *Computers and Education: Artificial Intelligence*, vol. 8, p. 100417, 2025.