Hybrid Retrieval-Based Intelligent College Information Assistant
Abstract

Students often face difficulty in accessing accurate and relevant academic information due to the scattered nature of college websites, where details related to departments, courses, faculty, regulations, and notices are distributed across multiple pages and formats. Conventional keyword-based search mechanisms fail to capture semantic meaning and relational structure present in such academic data.

The objective of this mini project is to design and develop a hybrid retrieval-based intelligent college information assistant that provides reliable and context-aware responses grounded in official college website data. Motivated by recent studies on hybrid retrieval and retrieval-augmented generation, the proposed system combines semantic vector-based retrieval with structured knowledge graph reasoning to enhance accuracy and generalization [1, 2].

The methodology involves scraping and preprocessing unstructured college website content, which is indexed using semantic embeddings in a vector database for efficient similarity-based retrieval. In parallel, a lightweight knowledge graph is constructed to model structured academic relationships such as faculty–course–department associations. A retrieval routing mechanism selects relevant information from either the vector store or the knowledge graph, and the retrieved context is supplied to a large language model for response generation. This hybrid approach improves factual grounding, supports relational queries, and reduces hallucination compared to purely semantic retrieval methods.

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

Keywords: Retrieval-Augmented Generation, Hybrid Retrieval, Knowledge Graph, Semantic Search, College Information System

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] M. Luo, S. Jain, A. Gupta, *et al.*, "A study on the efficiency and generalization of light hybrid retrievers," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, 2023.

[2] X. Zhu, Y. Xie, Y. Liu, Y. Li, and W. Hu, "Knowledge graph-guided retrieval augmented generation," in *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the ACL*, 2025.