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

First Review

Group Members

Julia Mariam John (B23CS2137)
Nirmel B Joseph (B23CS2148)
Rohith NS (B23CS2156)

Project Guide: Mr. Praveen J. S.
Assistant Professor, Dept. of CSE, MBCET

Contents

- Abstract
- Introduction
- Literature Review
- Research Gaps
- Problem Statement
- Objectives
- Feasibility Study
- System Architecture
- Data Flow Diagram-0
- Data Flow Diagram-1
- Use Case Diagram
- Class Diagram
- Sequence Diagram
- End User Feedback
- Methodology
- Conclusion
- References

Abstract

Academic data is typically unstructured and fragmented across institutional websites, limiting the effectiveness of keyword-based and embedding-only retrieval systems. Hybrid knowledge-graph-enhanced retrieval-augmented generation improves contextual grounding and relational reasoning by combining semantic vector retrieval with structured knowledge graph inference, supporting SDG 4 and SDG 9 [1, 2].

Introduction

- Growing dependence on online academic information
- Institutional data scattered across multiple web sources
- Limited relational reasoning in conventional retrieval systems
- Need for intelligent, knowledge-driven academic information systems

Literature Review

Sl.No	Title	Methodology	Results	Limitations
1	Knowledge graph-extended RAG for QA [1]	KG-guided retrieval + LLM	Improved multi-hop accuracy	Requires structured knowledge graph
2	RAG for educational applications [2]	Semantic retrieval + LLM	Reduced hallucination	Dependent on retriever quality

Research Gaps Identified

- Limited use of knowledge graphs in academic RAG systems
- Lack of hybrid vector-graph retrieval pipelines
- Insufficient explainability in LLM-based academic QA

Problem Statement

To develop a hybrid knowledge-graph-enhanced retrieval-augmented generation system for accurate and explainable academic information access.

Objectives

1. To preprocess and index institutional text using vector embeddings.
2. To construct a knowledge graph of academic entities for Mar Baselios College of Engineering and Technology.
3. To integrate hybrid retrieval for query answering.
4. To improve factual grounding and relational reasoning.

Feasibility Study

Technical Feasibility

- Uses existing academic data from institutional websites
- Integrates proven RAG models with knowledge graph techniques
- Supports semantic and relationship-based information retrieval
- Can be deployed using standard cloud or local server infrastructure

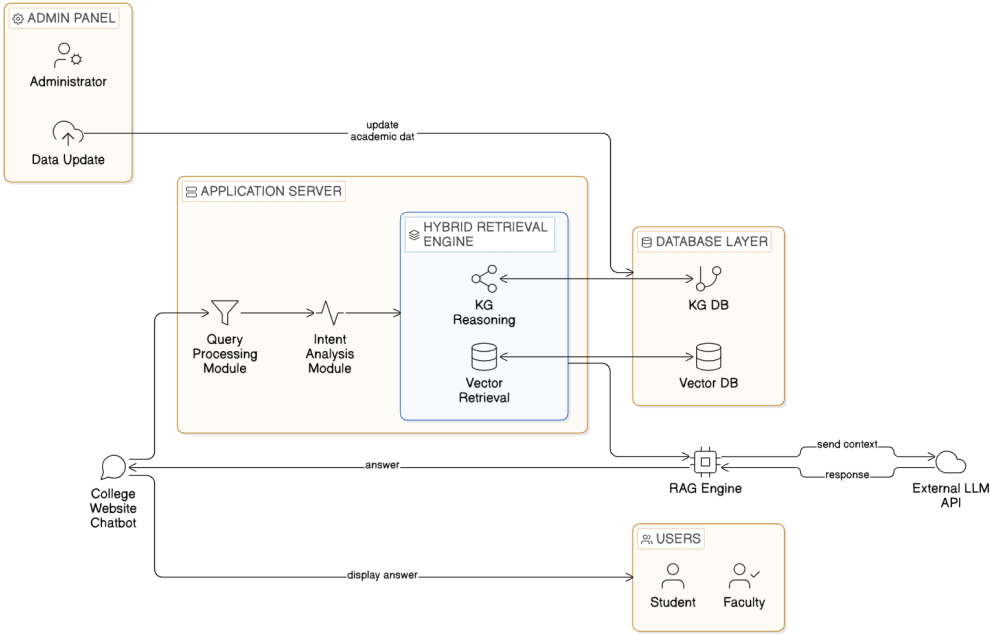
Economic Feasibility

- Relies on open-source NLP, vector databases, and graph frameworks
- No requirement for specialized hardware
- Low development and deployment cost
- Minimal maintenance cost after implementation

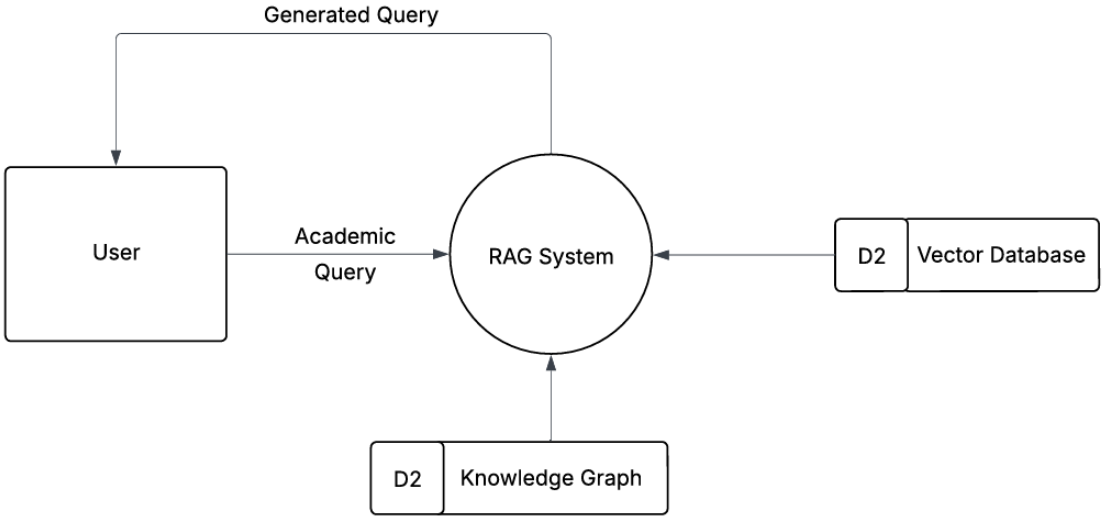
Operational Feasibility

- Provides an easy-to-use chatbot interface for users
- Reduces manual effort for administrative staff
- Requires minimal user training
- Suitable for continuous operation on academic websites

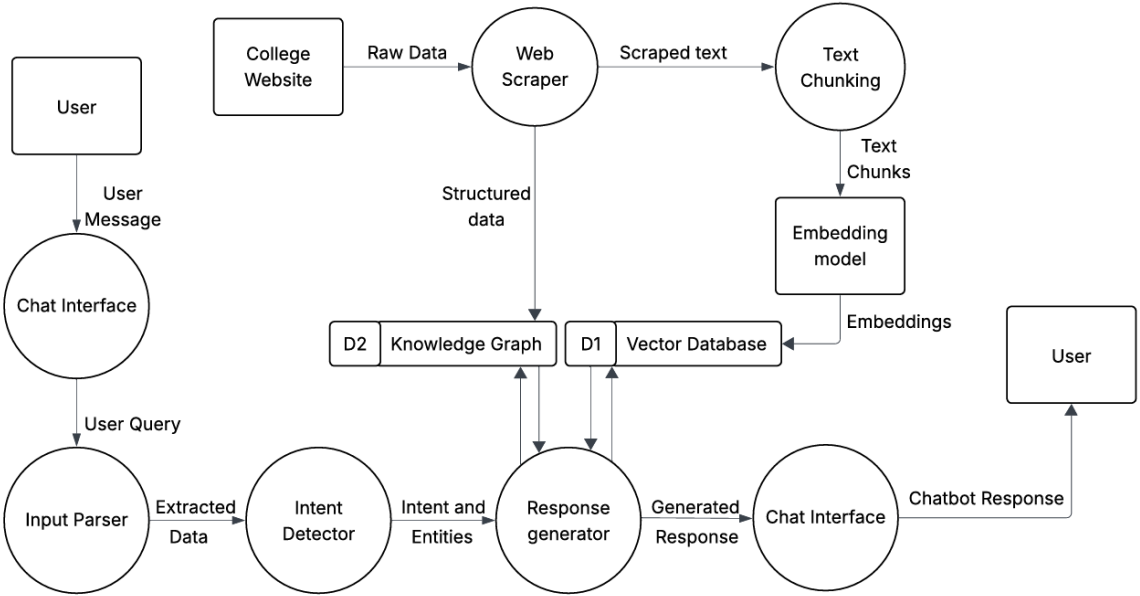
System Architecture



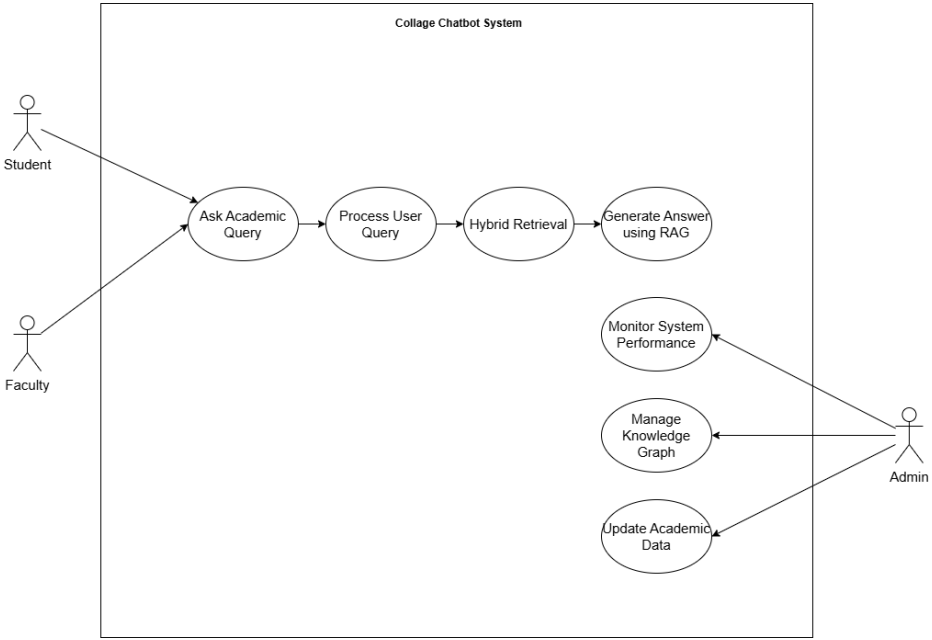
Data Flow Diagram-0



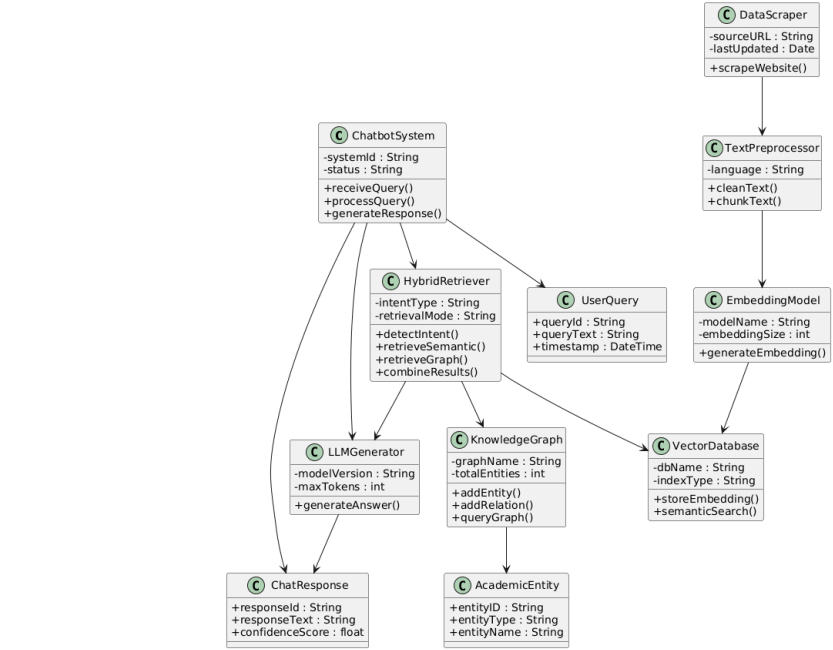
Data Flow Diagram-1



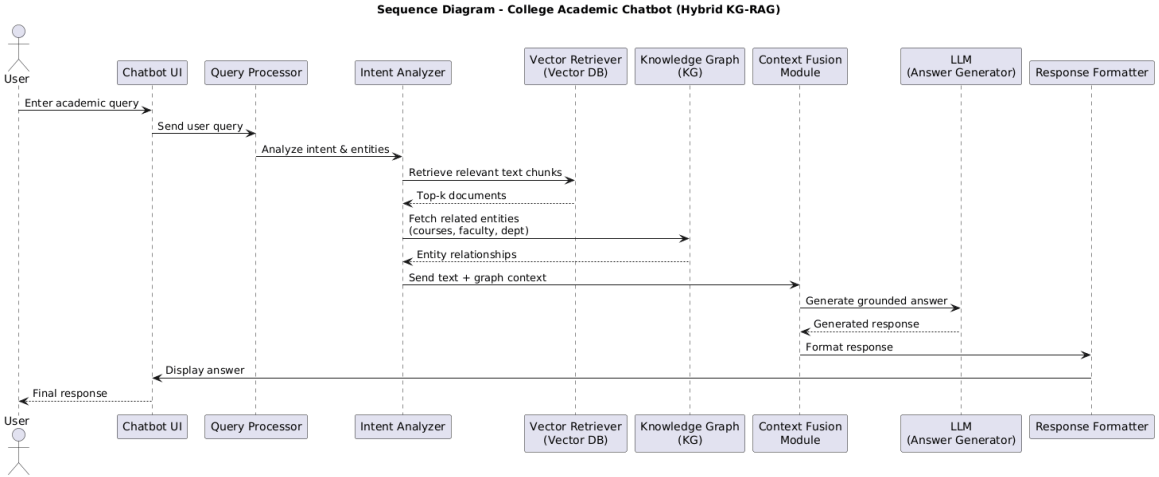
Use Case Diagram



Class Diagram



Sequence Diagram



End User Feedback

- Users mainly visit the college website for exam schedules, syllabus, and academic calendar.
- Many users find it difficult to locate information due to scattered content and complex navigation.
- Common issues include too many clicks and lack of centralized academic information.
- Most users believe a chatbot would help in faster and easier information access.
- Users prefer the chatbot to provide exam-related, academic, and department-specific information.

Methodology

- Data scraping, cleaning, and chunking
- Vector embedding
- Knowledge graph construction
- Hybrid retrieval and LLM-based answer generation

Conclusion

- Hybrid KG-RAG improves academic information retrieval
- Supports both semantic and relational queries
- Provides grounded and explainable responses
- Aligned with SDG 4 and SDG 9

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.

Thank You