Developing a Personalized Course Recommendation System Using RAG Architecture and Domain-Specific Knowledge Bases

CS4991 Capstone Report, 2025

Alex Nguyen
Computer Science
The University of Virginia
School of Engineering and Applied Science
Charlottesville, Virginia USA
hyv3jz@virginia.edu

ABSTRACT

Selecting elective courses is an important yet time-consuming process for university students, but personalized recommendations can significantly improve the decision-making process. This technical project proposes a solution that utilizes Retrieval-Augmented Generation (RAG) architecture to develop a personalized course recommendation for students, specifically at the University of Virginia (UVA). By using domain-specific knowledge bases that represent UVA's available courses and degree requirements, tailored course recommendations can be provided to students given their individual preferences. The system employs a local large language model, such as Llama, for generating responses. Key methods include the creation of a knowledge base from UVA's course catalogs, embedding of documents, and the use of Python and LangChain frameworks to integrate a RAG component. Initial results show up-to-date and directly sourced personalized course recommendations. Future work will focus on evaluating the relevancy of responses to users.

1. INTRODUCTION

Selecting elective courses as a university student involves two primary methods: flipping through online course catalogs with basic filter terms; or relying on word-of-mouth recommendations. Both approaches can be time-consuming, given the number of actions required and also their limited ability to efficiently offer personalized guidance. For instance, browsing through course catalogs typically involves manually sifting through long lists of courses and reading their individual descriptions. On the other hand, relying on peers for advice on what elective courses to take might lead to incomplete information, given that the available courses can change every semester or quarter. Therefore, both methods could leave students unaware of the valuable courses that align with their unique interests and academic goals.

This process is particularly challenging for students at universities like UVA, where courses are added or become unavailable through individual semesters. The diverse courses and frequently updated course offerings can make it difficult for students to actively discover the right courses for them. Existing tools are inadequate because they: 1) have limited search capabilities that do not help students find courses that meet their personal interests; or 2) do not curate personalized course information aligning with the students' interests.

2. RELATED WORKS

Existing works such as theCourseForum, developed by University of Virginia students, provide students with organized information about the offered courses for a given semester. In addition, they provide information such as the professors' ratings and average GPA statistics. This tool is useful for evaluating course difficulty and quality, but does not offer an automated, personalized recommendation system. Instead, students manually search through reviews, which can be time-consuming (theCourseForum, 2005).

Another implementation that has inspired my proposal is an academic advising system that explores the integration of chatbots to enhance student support. Lucien and Park (2024) developed an advising bot that took in university students' current degree progress and grades, in order to provide assistance in answering frequently asked questions (FAQs). While effective at being an interactive FAQ system about academic requirements, it was not necessarily a personalized recommendation tool. The system I propose was inspired by the positive interactions between students and natural language processing systems (NLP), like those the chatbot developed.

3. PROPOSAL DESIGN

Several traditional online course catalog tools currently exist. A website, theCourseForum, was founded in 2005 by volunteer students from UVA. The purpose of the website is to filter classes' primary terms such as name, professor and department. Additionally, the website allows students of the courses to give a rating to the professor/course and provides statistical information such as previous grade point averages (GPA). This application is great for gaining information about professors and course difficulty but does not address the issue of efficiently finding courses that meet personal interests, due to the limited search capabilities (theCourseForum, 2005). Existing tools, such as course catalog websites with basic filtering systems, fail to provide an efficient way to curate personalized course information that aligns with the interests of students.

3.1 Further Background and Proposed Solution

To address these challenges, I propose developing a course recommendation and scheduling system powered by Retrieval-Augmented Generation (RAG) architecture. This system will utilize the capabilities of local Large Language Models (LLMs) to retrieve and process information from UVA's course catalogs and other relevant academic data. The application will have the primary functionality of matching courses with a student's personal interests based on keywords. By integrating an external database or knowledge base with course catalogs, degree requirements and scheduling rules, the system will be able to provide accurate and personalized guidance for students. This solution aims to save university students time by reducing decision-making stress and providing tailored, well-informed course information.

3.2 Data Collection and Course Knowledge Base

The first developmental step in building the RAG-based system is gathering all relevant course data from UVA's course catalogs. UVA publicly provides data on the courses that they provide for a given semester, and for each course, includes information such as course tag, name and description.

The raw data gathered from publicly available sites will then undergo a preprocessing phase to format the text in an organized javascript object notation (JSON). Each section of courses will be chunked into separate, more manageable segments by the department in which the courses are available. Each chunked segment will be vectorized using sentence-transformers to create embeddings that retain the semantic meaning and relationships between the course topics and keywords. The knowledge base will be built from the vectorized course data, using ChromaDB, an open-source vector database designed for fast similarity searching.

3.3 Building the RAG Model with LangChain and Llama

LangChain is an open-source Python framework that supports development of a RAG-enhanced LLM system. Therefore, LangChain will be used to manage the retrieval pipeline with features that include handling query processing, document searching and context influenced responses (Sindhu et. al., 2024).

The retrieval is envisioned as follows: When a user submits a query, LangChain will query ChromaDB to retrieve relevant course descriptions. The retrieved course segments will then be passed the local Llama model, which will generate an appropriate response that is influenced by the context pass through it. This will ensure that the recommendation is fact-based and directly sourced from the knowledge base created.

The proposed design recommends using a locally hosted mode, such as the Llama model, to ensure data privacy and reduced reliance on external APIs. In addition, the cost of development will be generally more cost-efficient by eliminating the recurring cloud costs incurred using API query handling used by other models. Also, since all actions are also handled locally, any sensitive information will be limited to the scope of development.

3.4 Limiting Model Scope and Handling Unknown Queries

To take steps towards ensuring the greatest accuracy and to prevent any hallucinations, the design proposed will limit the type of responses that the system can answer, given the data stored in the ChromaDB knowledge base. When a user queries and there are no relevant results, the system will have a fallback message such as “No relevant courses found,” instead of generating speculative answers with no context provided (Cao, 2024). These steps are crucial in an academic setting where misinformation can negatively impact students’ course planning and possibly their performance. Emphasizing a retrieval-first approach will guarantee that any response or recommendation is based on validated sources.

4. ANTICIPATED RESULTS

One of the expected outcomes of the proposed course recommendation system is that the function will be able to provide students with personalized course recommendations based on their degree requirements and personal interests. In addition, with the overall automatic relevant course retrieval that the RAG architecture uses when accessing the knowledge base of course recommendations, students or users can expect a more streamlined and informed decision-making process when selecting electives, reducing the time and effort traditionally required for selecting future courses.

However, given the proposed design, there are anticipated limitations regarding the system’s capabilities in handling queries beyond the scope of the course recommendations. The scope of queries that the system answers is limited purposefully in the design to mitigate misinformation. Therefore, this constraint will help with the accuracy and reliability of the responses, ensuring that students only receive validated information.

5. CONCLUSION

The proposed development of a personalized course recommendation system was designed to enhance the course selection experience for UVA students. By building from a RAG architecture using UVA's course catalogs as domain-specific knowledge, the system could deliver tailored recommendations based on both individual student preferences and degree requirements. This system not only speeds up the traditionally time-consuming process of elective course selection, but also helps students discover courses that better align with their personal and academic interests.

In addition, the value that this system brings to the academic advising environment consists of reducing the overall workload of academic advisors regarding routine course-related inquiries. This would then allow them to focus on more nuanced, interpersonal advising tasks, such as discussing career goals, academic motivation and an overall support system. Ultimately, the proposed system offers a modern and practical solution that would benefit both students and advisors.

6. FUTURE WORK

After the system is fully developed or a working system can be tested, the future work will involve levels of evaluation to ensure the accuracy and trustworthiness. This would include quantitative measurements of accuracy, such as relevancy and accuracy of the relevant course recommendations based on student queries. In addition, observational testing would be done with various UVA students that would provide valuable, qualitative insights on whether or not the system's recommendations align with student expectations and improve their overall course selection experience. Overall, future steps consist of evaluating the system's capabilities and trustworthiness within an academic advising environment.

REFERENCES

Sindhu, B, Prathamech R., Sameera, M., & KumaraSwamy, S. (2024). The evolution of large language model: Models, applications and challenges. *2024 International Conference on Current Trends in Advanced Computing (ICCTAC)*, 1–8. https://doi.org/10.1109/icctac61556.2024.10581180

Cao, L. (2024). Learn to refuse: Making large language models more controllable and reliable through knowledge scope limitation and refusal mechanism. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 3628–3646. https://doi.org/10.18653/v1/2024.emnlp-main.212

The Course Forum. (2005). *Your guide to courses at UVA*. https://thecourseforum.com/

Lucien, R., & Park, S. (2024). Design and development of an advising chatbot as a student support intervention in a university system. *TechTrends*, *68*(1), 79–90. https://doi.org/10.1007/s11528-023-00898-y