

Statement of Work - CS-5588 Capstone - Group 3
Jeet Das - ad5f2@mail.umkc.edu

Project Title: Enhanced Retrieval and Generation System for References Documents

1. Introduction

The "Enhanced Retrieval and Generation System for References Documents" project seeks to improve the accessibility of critical information within extensive civil engineering reference manuals. Traditional manuals, known for their extensive content spanning thousands of pages, pose a significant challenge in locating specific information promptly. Civil engineers often spend hours sifting through these documents to understand how to apply their skills under the constraints of various forms of governance, ranging from county, city, state, and generally accepted rules and regulations. This project aims to address this challenge by developing an end-to-end system empowered by cutting-edge technology. Although this project seeks to implement this technology focussed on the civil engineering domain, the ideas suggested here should be generally applicable in various domains under similar circumstances.

2. Objectives

The primary objective of this project is to create a comprehensive system that empowers civil engineers and related professionals to efficiently retrieve essential information from vast and intricate reference manuals. To achieve this, we have outlined the following specific objectives:

- *Natural Language Querying:* Enable users to input questions in everyday language, simplifying the search process.
- *Information Retrieval and Generation:* Combine advanced information retrieval and generation techniques to extract accurate answers from dense reference materials.
- *Customizable Responses:* Allow the system to tailor responses based on user-defined parameters, providing personalized outputs.
- *Easy Integration:* Provide the system as an API for seamless integration into existing workflows.

3. Scope

The scope of this project includes the development and implementation of the Enhanced Retrieval and Generation System for Civil Engineering Reference Manuals. It encompasses the creation of the system's core functionality, the integration of advanced data science techniques, and the delivery of the system as a user-friendly API. The project's scope is limited to the domain of civil engineering reference manuals and does not extend to other fields or types of documents.

4. Data Sources

The project will utilize various sources of data, including civil engineering reference manuals, technical documents, and related resources. These data sources will be processed and transformed into a format suitable for use with the system. Data sources will be primarily collected from open-access repositories, libraries, and academic databases.

5. Methodologies

The methodologies to be employed in this project encompass the utilization of large language models (LLMs), such as the GPT family, for text generation. Vector databases will store vectorized documents, and FastAPI will facilitate the delivery of information through a user-friendly REST API. Additionally, the project will explore the applications of encoder-style LLMs, like BERT, for information retrieval, and the integration of fine-tuned LLMs for specific tasks.

6. Timeline

A detailed project timeline will be established as the project progresses. Milestones and estimated completion dates will be defined to ensure steady progress towards achieving the project's objectives. Regular project updates and progress reports will be provided to track and manage the project's timeline effectively.

An estimated schedule for the deliverables is as follows:

- *September 2023*: Research ideas and develop proof of concepts for various technologies in the pipeline.
- *October 2023*: Pick suitable technologies from the proof of concepts and start building an end-to-end system that ingests documents, vectorizes them, stores them in a vector DB for future use, and uses a state-of-the-art LLM for generating answers asked in natural language.
- *November 2023*: Extensive testing with stakeholders to best understand accuracy, and reliability to incorporate feedback. During this timeframe, develop a REST API that allows for integration with third-party platforms.
- *December 2023*: Develop a baseline for future improvements, add automated tests, third-party integrations, improve performance for API, etc.

7. Evaluation Criteria

The success of this project will be evaluated based on the following criteria:

- **Depth and Clarity of answers**: The project will be evaluated based on the intricacies of the answers provided to complex questions which will test its reasoning abilities, extraction of answers from dense text, and clarity to the end user.

- **Accuracy:** The project will be evaluated on its ability to answer the questions and provide accurate answers. This is to additionally ensure it correctly cites sources and checks if answers are hallucinations vs. real.
- **Integration:** This section is to ensure the system does not work in isolation and provides a reliable API for third-party access. This will ensure that the system coheres to the end-users usability expectations and provides a gateway for further enhancements and downstream consumption.

8. Conclusion

The project is poised to bring innovation to the field of civil engineering by offering a rapid and accurate means of retrieving essential information from reference manuals. By adhering to the objectives and scope outlined in this Statement of Work, the project aims to enhance productivity and decision-making for civil engineers and professionals in the domain.