AI Application in Business Use Cases

YIP Ron Yui

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Overview

Digitalization is an inevitable trend for traditional companies. Arup, a leading construction consultancy globally, is currently at the forefront of this revolution. The company is exploring how artificial intelligence (AI) can be applied to improve workflows across its operations. This research describes four projects that utilize various AI technologies to address the needs of different teams.

Objectives

- Convert documents into formatted data that benefits later stage
- Distill non-trivial features from data to reduce the data size
- Extract accurate result from features
- Visualize information to users in a user-friendly approach

Project 1: Document Query Tool

About the project

The Document Query Tool was developed for general use by colleagues at Arup. It aimed to deliver two main functions: (1) querying documents through a chatroom interface and (2) enabling mass queries by uploading an Excel template filled with questions.

Methodology

Data Conversion: We converted the documents into chunks and then obtained the vector database by processing these chunks through embedded models.

Feature Extraction: In the retrieval step of the RAG process, we applied semantic reranking and extracted the top K chunks to serve as augmented data.

Result Generation: We conducted the augmented generation step of the RAG by generating results through a prompt template that includes user query and extracted chunks.

Output Visualization: We provided a frontend website that features (1) a file upload page, (2) an interactive chatroom page, and (3) a mass query Excel download page.

Result

We successfully delivered the basic RAG system with an interactive chatroom, and the mass query function will be implemented by colleagues in future stages.

Project 2: Component Counting Tool

About the project

The Component Counting Tool was developed to meet the needs of the bidding team at Arup. Its aim is to count the frequency of construction elements in reports, which can assist in improving cost estimation.

Methodology

Data Conversion: We converted the documents from PDF format into a list of strings.

Feature Extraction: We mapped the tokens to an ontology database and then applied part-of-speech tagging, semantic analysis to identify construction elements candidate phrases.

Result Generation: We used zero-shot prompting for each element to verify whether they are valid construction element phrases.

Output Visualization: A colleague will integrate the code into the Project Data Management Portal in a future stage so that the results can be displayed on the portal.

Result

We successfully developed the Component Counting Tool. However, the results will need to be verified by the client team, and the functionality will need to be integrated into the existing portal.

Project 3: RDS Workflow

About the project

The RDS Workflow was developed to meet the needs of MEP engineers at Arup. It is designed to extract key-value pairs from the room data sheet (RDS) and store them in a database.

Methodology

Data Conversion: We converted the documents from PDF format into a list of strings.

Feature Extraction: We extracted header-footer for ID identification, then restructured pages from the same document into one long string, removing any header-footer duplicates.

Result Generation: We used few-shot prompting to provide instructions and examples for the LLM, prompting it to return a JSON representation of the documents.

Output Visualization: A colleague will integrate the code into the Project Data Sheet Portal in a future stage so that the results can be displayed on the portal.

Result

We successfully developed the RDS Workflow. However, the functionality will need to be integrated into the existing portal by colleagues.

Project 4: Document Grouping Workflow

About the project

The Document Grouping Workflow was developed to meet the needs of the bidding team at Arup. It is designed to (1) classify documents into directories and (2) identify the construction elements in the drawings.

Methodology

Data Conversion: We converted the as-built drawings from TIFF images into a list of text boxes using OCR technology.

Feature Extraction: We grouped the text boxes using the DBSCAN algorithm.

Result Generation: We will find the best match for the project title by comparing each token in the bottom-right cluster with an existing title list using the Levenshtein Distance. We will also classify the construction elements using a similar approach to Project 2.

Output Visualization: A colleague will integrate the code into the Project Management Sheet Portal in a future stage so that the results can be displayed on the portal. The files will also be rearranged into their respective directories.

Result

We successfully developed the Document Grouping Workflow. However, the functionality will need to be integrated into the existing portal by colleagues.

Conclusion

In conclusion, we have delivered four projects to meet the needs of different teams at Arup. Following the successful delivery and integration into the existing portals commonly used by colleagues, we believe these projects will improve efficiency at work by providing greater convenience.