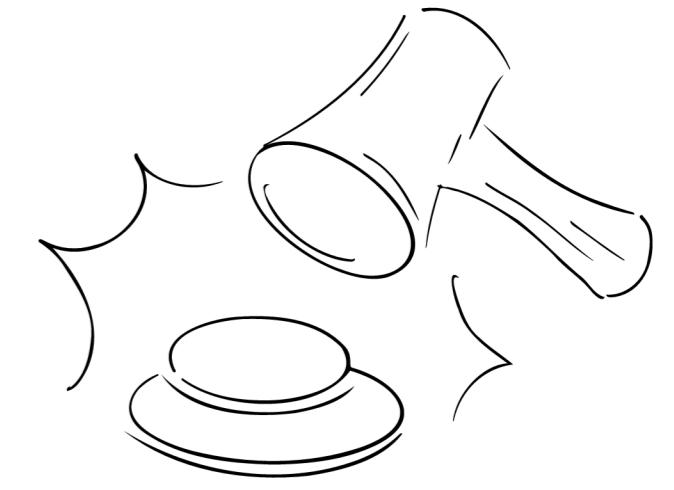


AI-powered Intelligent Legal Retrieval System



資訊三 劉宸均 黃柏淵 李承祐 徐鍊睿

Research motivation

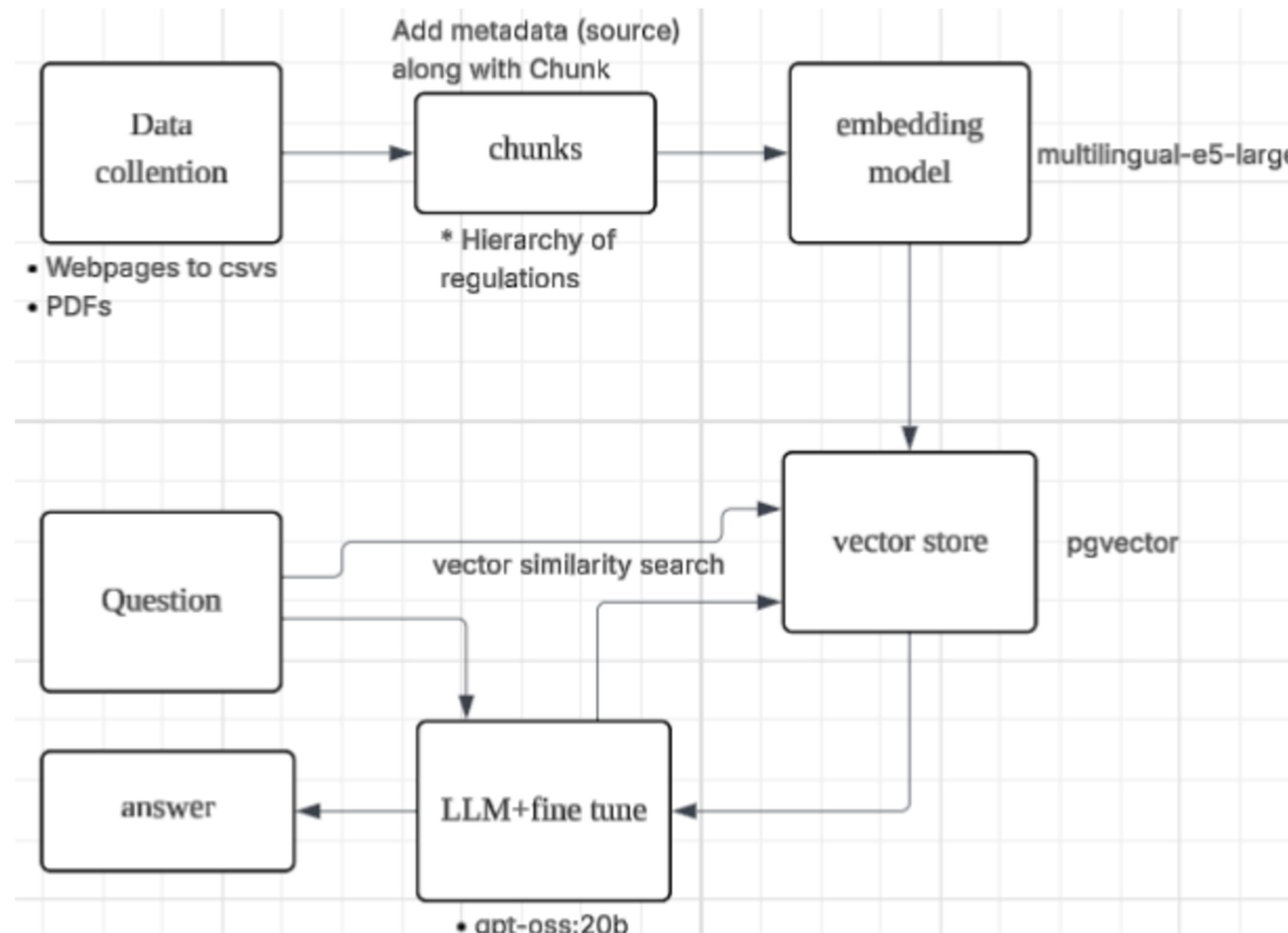
The content of occupational safety regulations is highly specialized and complex, making it difficult for frontline workers to understand. Many are unable to afford costly professional legal consultation, resulting in a high barrier to accessing essential regulatory knowledge.

Therefore, we aim to develop an intelligent query system capable of automatically parsing regulations and providing answers with corresponding legal references, in order to promote information equity for frontline workers.

Research purpose

This study aims to develop a regulation-oriented intelligent query system powered by RAG technology, capable of transforming government regulatory documents into a searchable knowledge base and generating answers with source citations through vector search and large language models (LLMs).

Workflow



The explanation of workflow steps

- Data: The raw data is sourced from the Occupational Safety and Health Section of the Tainan City Government.
- Chunks: 500 characters per chunk with a 200-character overlap.
- Embedding Model: multilingual-e5-large (1024 dimensions), with separate Query and Passage prefixes.
- Vector Store: PostgreSQL with pgvector and HNSW indexing, using L2 distance for Top-K retrieval.
- Similarity Search: Query embedding: Query → 1024-dimensional vector Vector retrieval: pgvector with L2 distance, Top-K search Result formatting: source attribution and numbered citations.

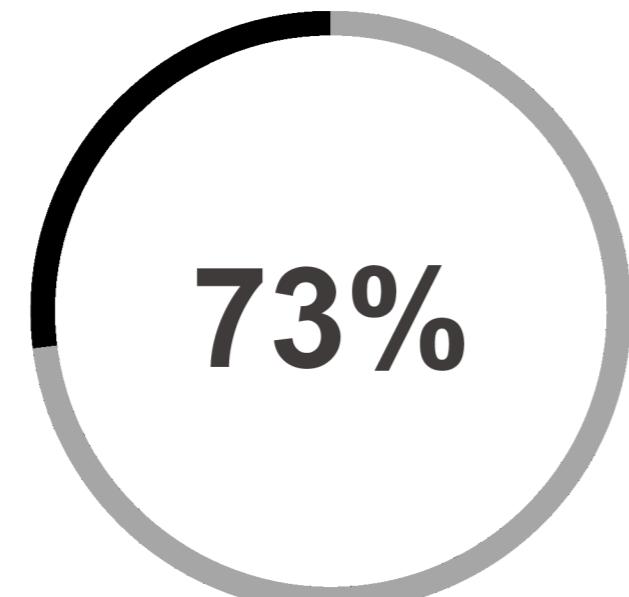
Experiments

Method: Evaluate the model's accuracy by testing it with exam questions.

- Source of Questions : 22200_-職業安全衛生管理學科(乙員)
- The test results are as follows :

file	number	model_answer	correct_answer
0	22200_-職業安全衛生管理學科.csv	1	③
1	22200_-職業安全衛生管理學科.csv	2	3
2	22200_-職業安全衛生管理學科.csv	3	1
3	22200_-職業安全衛生管理學科.csv	4	2
4	22200_-職業安全衛生管理學科.csv	5	2

Results



Model Answer Accuracy



Comparison With and Without RAG

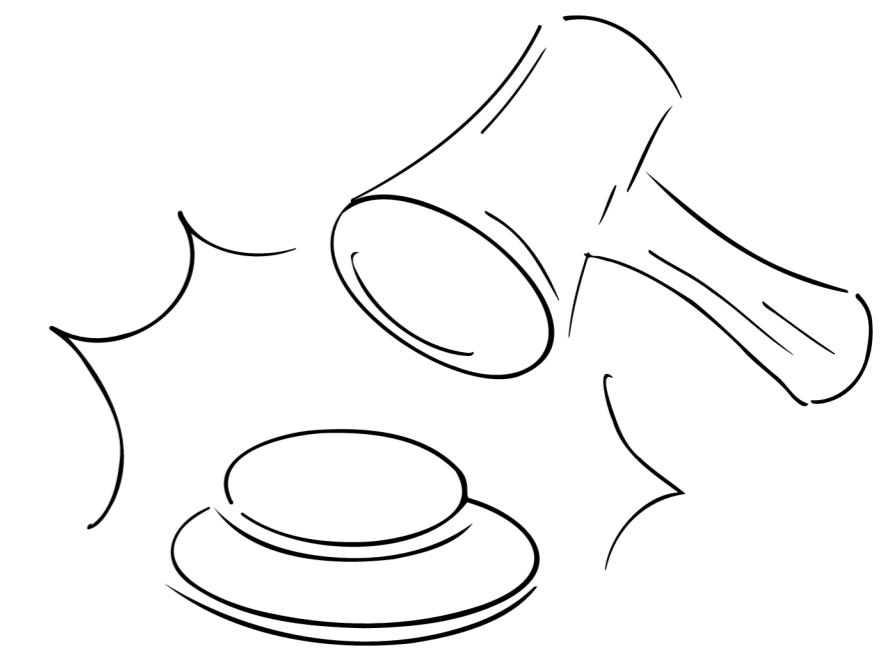
Conclusion

This study successfully developed a regulation-oriented intelligent query system powered by RAG technology, achieving a 73% answer accuracy on occupational safety exam questions. The system is capable of responding to most occupational safety regulation inquiries, demonstrating its feasibility in professional query scenarios.

In the future, we plan to expand the coverage of regulatory sources and further enhance model performance. We also aim to explore the use of knowledge graphs to represent regulatory relationships, allowing articles, definitions, responsibilities, and penalties to be structured more clearly. This will strengthen the interconnectedness of legal provisions and improve the interpretability and visualization of regulatory knowledge.

AI-powered intelligent legal retrieval system

資訊三 劉宸均 黃柏淵 李承祐 徐鍊睿



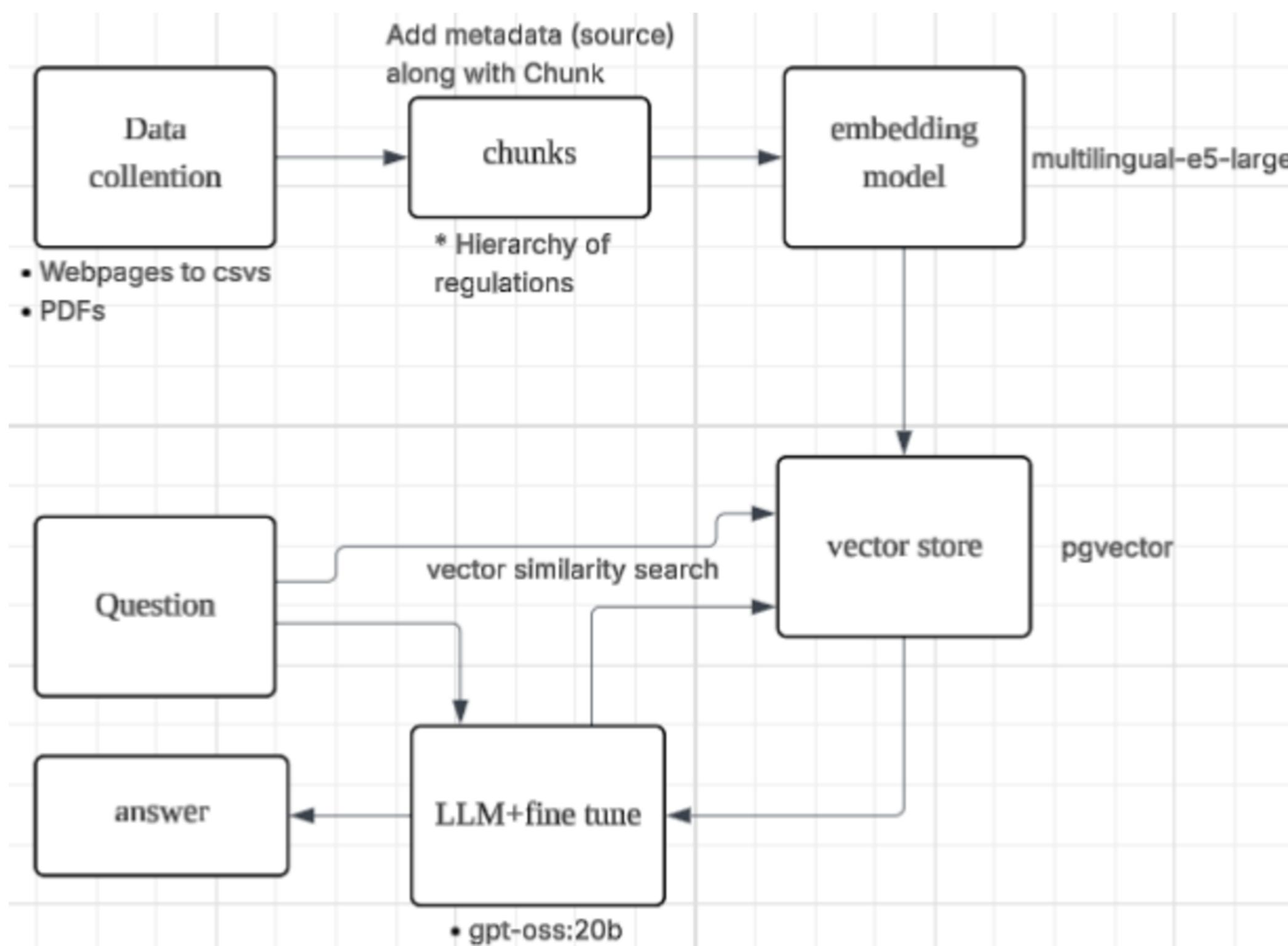
Research motivation

職安法規內容專業且複雜，對於基層勞工往往難以理解，也無法負擔高昂的專業諮詢費用，使得相關法規知識取得門檻過高。因此我們希望建置一套能自動解析法規並提供附有法源依據的智慧查詢系統，促進基層勞工資訊平權。

Research purpose

本研究希望打造一套以 RAG 技術為核心的法規智慧查詢系統，能將政府法規文件轉換成可搜尋的知識庫，並透過向量搜尋與 LLM 生成附帶來源依據的回答。

Workflow



The explanation of workflow steps

Data: raw data來自台南市職安健康處的法規專區

Chunks: 500 字元/塊，200 字元重疊

Embedding Model: multilingual-e5-large (1024 維)
Query/Passage 前綴區分

Vector Store: PostgreSQL + pgvector、HNSW 索引
L2 距離 Top-K 搜尋

Similarity Search: 查詢嵌入化：Query → 1024 維向量
向量搜尋：pgvector L2 距離，Top-K 檢索
結果格式化：來源標註、編號引用

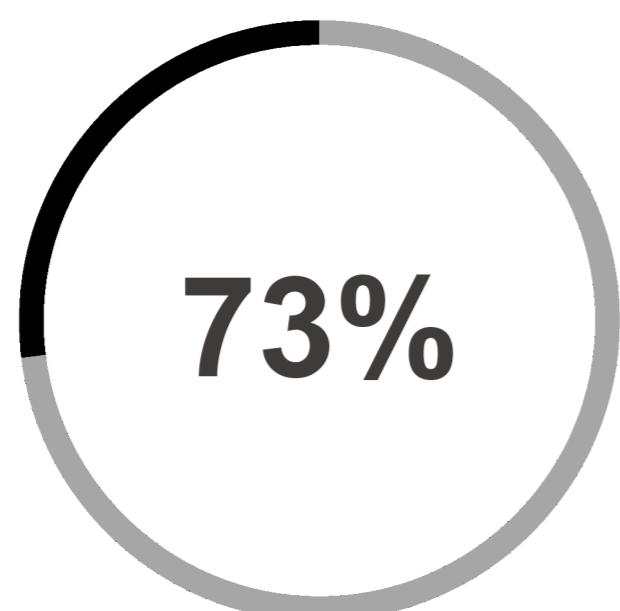
Experiments

Method: 讓模型做考題檢視正確率

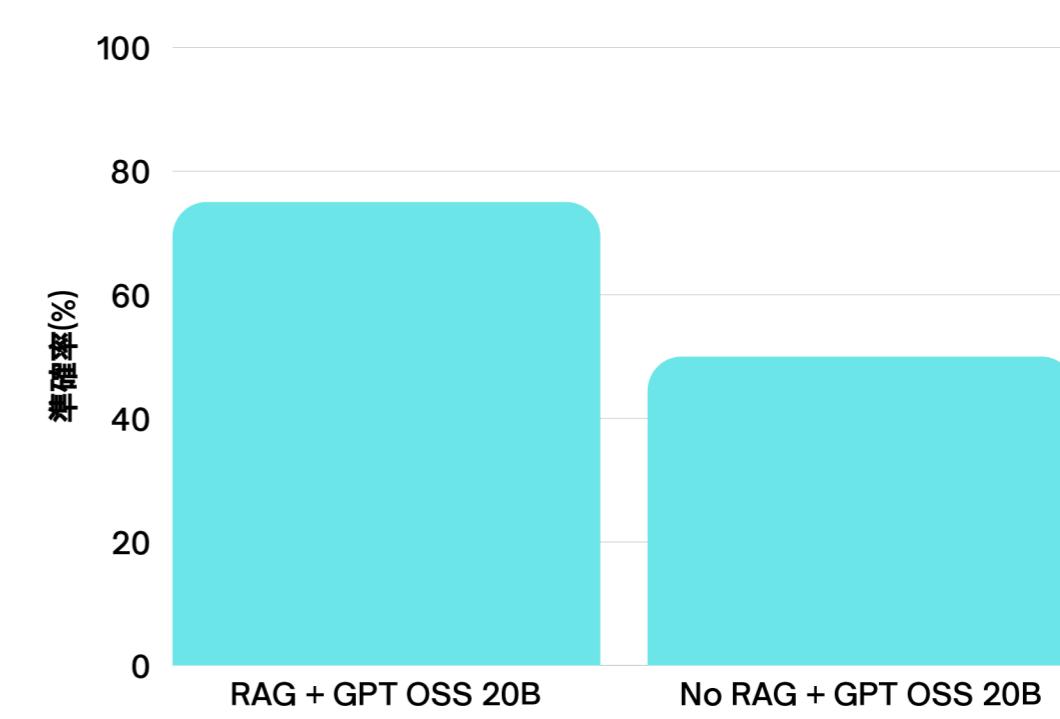
題目來源：22200_-職業安全衛生管理學科(乙員)
做題結果如下：

file	number	model_answer	correct_answer
0	22200_-職業安全衛生管理學科.csv	③	3
1	22200_-職業安全衛生管理學科.csv	3	3
2	22200_-職業安全衛生管理學科.csv	1	1
3	22200_-職業安全衛生管理學科.csv	2	2
4	22200_-職業安全衛生管理學科.csv	2	2

Results



模型回答準確率



有無RAG的比較

Conclusion

本研究成功建置一套以 RAG 技術為核心的法規智慧查詢系統，並在職安相關考題中達到 73% 的回答準確率。系統能大致回答職安法規問題，展現其在專業查詢場景的可行性。

未來，將擴充更多法規來源並提升模型效能，並進一步探索以知識圖譜方式呈現法規關係，使法條、定義、責任、罰則等內容能以更結構化的方式呈現，強化法規之間的連結性與可視化理解。