HỆ THỐNG HỎI ĐÁP PHÁP LUẬT VIỆT NAM

Trần Đăng Khoa - 230101009

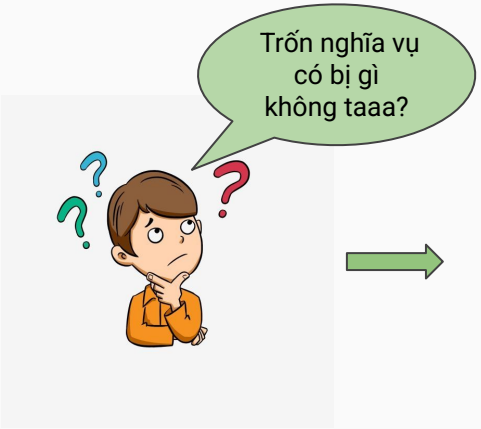
Tóm tắt

- Lớp: CS2205.CH183
- Link Github của nhóm: https://github.com/khoatdds/CS2205.CH183
- Link YouTube video: https://youtu.be/IFfXpxQkwzQ

Trần Đăng Khoa

Giới thiệu

Hệ thống Hỏi đáp Pháp luật Việt Nam ứng dụng **RAG (Retrieval Augmented Generation)** và **LLM (Large Language Model)** để tra cứu và phân tích luật nhanh chóng, chính xác. Hệ thống kết hợp truy xuất thông tin và mô hình tạo sinh văn bản, cung cấp câu trả lời có dẫn nguồn, giúp người dùng tiếp cận pháp luật dễ dàng, hiệu quả.



Trả lời: "Trốn nghĩa vụ quân sự có thể bị phạt hành chính hoặc truy cứu trách nhiệm hình sự tùy mức độ vi phạm, có thể bị phạt tiền, cấm xuất cảnh, hoặc phạt tù."

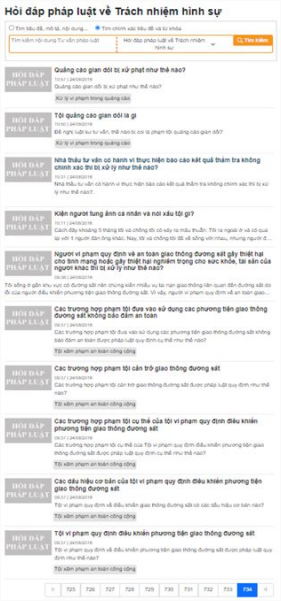
Dẫn chứng: Khoản 7 Điều 1 Nghị định 37/2022/NĐ-CP

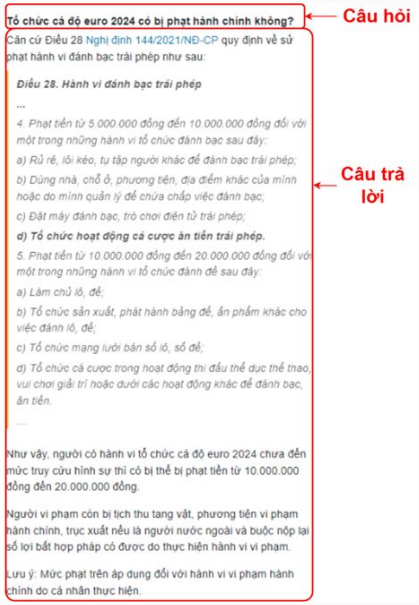
Mục tiêu

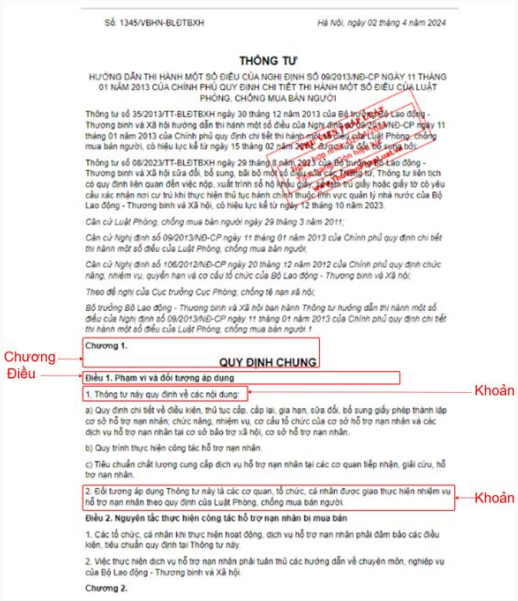
- Xây dựng bộ dữ liệu pháp luật Việt Nam đầy đủ và hoàn chỉnh.
- Phát triển mô hình truy xuất và tạo sinh thông minh bằng phương pháp **RAG** kết hợp **LLM**.
- Cung cấp một hệ thống hỗ trợ tra cứu pháp luật nhanh chóng, chính xác.

Nội dung và Phương pháp

- Xây dựng bộ dữ liệu pháp luật Việt Nam từ https://thuvienphapluat.vn/ gồm các cặp câu hỏi đáp và các văn bản pháp luật.



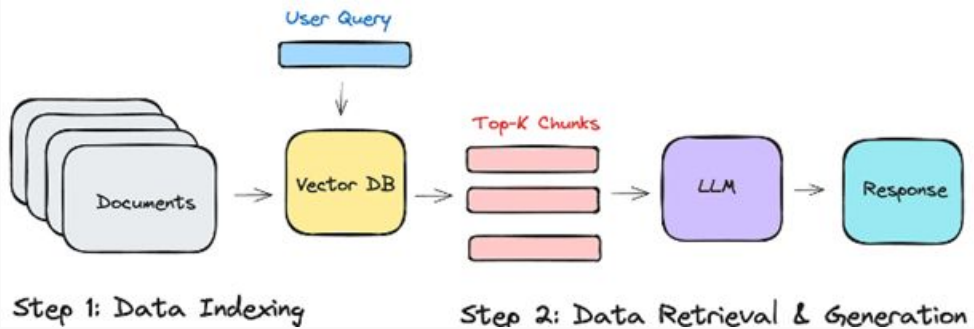




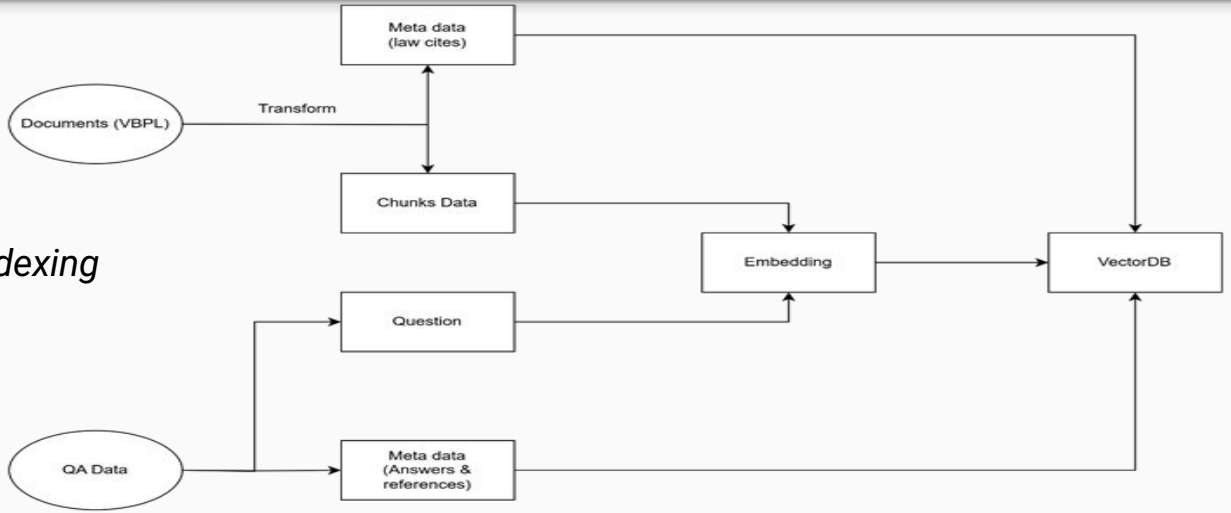
Nội dung và Phương pháp

RAG (Retrieval-Augmented Generation)

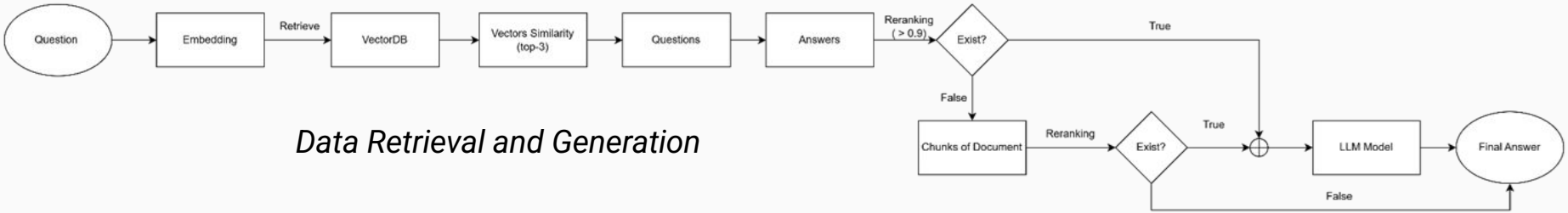
- Là kỹ thuật nâng cao độ chính xác cho câu trả lời của mô hình gen-AI với dữ liệu được truy xuất từ các nguồn dữ liệu bên ngoài.
- Giải quyết nhiều vấn đề mà các LLM thường hay gặp phải (ảo giác - hallucination; độ tin cậy; giảm chi phí, khối lượng công việc,..).



Nội dung và Phương pháp



Data Indexing



Data Retrieval and Generation

Nội dung và Phương pháp

- Model embedding: **bge-m3** (*https://huggingface.co/BAAI/bge-m3*)
- Model Reranking: **bge-reranker-v2-m3** (*https://huggingface.co/BAAI/bge-reranker-v2-m3*)
- Base model : **Vistral-7B** (*https://huggingface.co/Viet-Mistral/Vistral-7B-Chat*)

Kết quả dự kiến

- Xây dựng được một bộ dữ liệu pháp luật Việt Nam hoàn chỉnh và đầy đủ nhất.
- Xây dựng được một hệ thống hỏi đáp pháp luật Việt Nam, câu trả lời sẽ được dẫn chứng cụ thể, rõ ràng.

Tài liệu tham khảo

[1]. P. S. H. Lewis et al., “Retrieval-Augmented Generation for Knowledge-Intensive NLP tasks,” Neural Information Processing Systems, vol. 33, pp. 9459–9474, May 2020, [Online]. Available: https://proceedings.neurips.cc/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf

[2]. Jianlyu Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. 2024. M3-Embedding: Multi-Linguality, Multi-Functionality, Multi-Granularity Text Embeddings Through Self-Knowledge Distillation. In Findings of the Association for Computational Linguistics: ACL 2024, pages 2318–2335, Bangkok, Thailand. Association for Computational Linguistics.

[3]. Chaofan Li, Zheng Liu, Shitao Xiao, and Yingxia Shao. 2023. Making large language models a better foundation for dense retrieval. arXiv preprint arXiv:2312.15503.

[4]. Van Nguyen, C., Nguyen, T., Nguyen, Q., Nguyen, H., Plüster, B., Pham, N., ... & Nguyen, T. (2023). Vistral-7b-chat-towards a state-of-the-art large language model for vietnamese.