

TRƯỜNG ĐẠI HỌC BÌNH DƯƠNG
VIỆN TRÍ TUỆ NHÂN TẠO VÀ
CHUYỂN ĐỔI SỐ
KHOA CNTT, ROBOT VÀ TTNT

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

ĐỀ CƯƠNG ĐỒ ÁN TỐT NGHIỆP

TÊN ĐỀ TÀI: XÂY DỰNG CHATBOT HỖ TRỢ TƯ VẤN TUYỂN SINH TẠI TRƯỜNG ĐẠI HỌC BÌNH DƯƠNG SỬ DỤNG RAG

Cán bộ hướng dẫn (CBHD): ThS. Lê Duy Hùng

Thời gian thực hiện: Từ ngày 22.09 -14.12.2025 (12 tuần)

Sinh viên thực hiện:

Nguyễn Bá Trường – 18050082

Nội dung đề tài:

Lý do chọn đề tài:

- Tuyển sinh là hoạt động quan trọng của mọi trường đại học, đặc biệt trong bối cảnh cạnh tranh gay gắt hiện nay. Nhu cầu tư vấn 24/7 từ thí sinh và phụ huynh ngày càng tăng cao khiến việc tư vấn thủ công tốn kém và không hiệu quả, công tác tuyển sinh của nhà trường cần tăng cường tương tác nhanh, đa kênh và hỗ trợ thông tin phong phú (văn bản, hình ảnh, tài liệu PDF, video giới thiệu). Chatbot ứng dụng kỹ thuật Retrieval-Augmented Generation (RAG) đa phương thức (multimodal RAG) sẽ giúp trả lời chính xác các câu hỏi tuyển sinh, xử lý tài liệu nhiều dạng, giảm tải cho cán bộ tuyển sinh và nâng cao trải nghiệm thí sinh.

Mục tiêu:

- Thiết kế và triển khai chatbot tuyển sinh sử dụng Multimodal RAG có khả năng truy vấn và sinh ngôn ngữ dựa trên nguồn dữ liệu tuyển sinh
- Hỗ trợ tra cứu thông tin về ngành, chỉ tiêu, điều kiện xét tuyển, học phí, lịch tuyển sinh, quy trình nộp hồ sơ.
- Đánh giá hiệu năng độ chính xác truy vấn, thời gian phản hồi, độ hài lòng người dùng.

Phạm vi nghiên cứu:

- Dữ liệu: Thông tin tuyển sinh Trường ĐH Bình Dương (quy chế, ngành học, học phí, điểm chuẩn, cơ sở vật chất)
- Công nghệ: Multimodal RAG, Large Language Models (LLM), Vector Database

- Định dạng: Văn bản, hình ảnh, PDF
- Người dùng: Thí sinh, phụ huynh, cán bộ tuyển sinh

Ý nghĩa của đề tài:

- Ý nghĩa khoa học
 - o Nghiên cứu ứng dụng thực tế của Multimodal RAG trong giáo dục
 - o Đóng góp vào việc số hóa quy trình tuyển sinh tại Việt Nam
- Ý nghĩa thực tiễn:
 - o Giảm tải công việc cho cán bộ tuyển sinh
 - o Tăng cường khả năng truy xuất thông tin phức tạp từ tài liệu đa phương thức.
 - o Nâng cao chất lượng dịch vụ tư vấn, phản hồi nhanh chóng
 - o Tiết kiệm chi phí vận hành, tăng khả năng tiếp cận thông tin
 - o Tạo lợi thế cạnh tranh cho trường trong hoạt động tuyển sinh
 - o Là nền tảng nghiên cứu và triển khai các hệ thống hỗ trợ tuyển sinh thông minh cho nhà trường.

Đối tượng nghiên cứu:

- Các phương pháp xây dựng hệ thống hỏi đáp tự động dựa trên Retrieval-Augmented Generation (RAG).
- Ứng dụng Multimodal RAG để xử lý dữ liệu tuyển sinh đa dạng (văn bản, hình ảnh, tài liệu, video).
- Kỹ thuật triển khai chatbot hỗ trợ tư vấn tuyển sinh trong môi trường thực tế tại trường đại học.

Phương pháp thực hiện:

- Thu thập và tiền xử lý dữ liệu tuyển sinh.
- Xây dựng pipeline embedding cho dữ liệu đa phương thức.
- Triển khai RAG: vector store, retrieval module, và generation module.
- Thiết kế giao diện demo để tương tác và thu thập feedback.
- Đánh giá bằng bộ câu hỏi thử nghiệm, đo lường các chỉ số: chính xác thông tin, độ phù hợp câu trả lời, thời gian phản hồi, và khảo sát người dùng.

Kết quả mong đợi:

- Một prototype chatbot hoạt động với dữ liệu tuyển sinh mẫu, có khả năng trả lời 70–85% các câu hỏi phổ biến chính xác và trả lời tài liệu tham chiếu được.
- Chatbot có khả năng tiếp nhận câu hỏi bằng tiếng Việt, xử lý dữ liệu từ nhiều nguồn như FAQ, văn bản, PDF, hình ảnh.

- Hỗ trợ người dùng tra cứu thông tin tuyển sinh cơ bản: ngành học, chỉ tiêu, học phí, điều kiện xét tuyển, thời gian nộp hồ sơ.
- Có giao diện demo trực quan, dễ sử dụng. Giảm tải cho cán bộ tuyển sinh, có thể mở rộng để ứng dụng thực tế tại trường
- Báo cáo đồ án, bộ dữ liệu, mã nguồn + hướng dẫn cài đặt, slide thuyết trình và video demo.

Kế hoạch thực hiện:

Giai đoạn 1: Nghiên cứu lý thuyết và Thu thập dữ liệu (Tuần 1-4)

- Tuần 1 (22/09 - 28/09):
 - o Nghiên cứu lý thuyết về Multimodal RAG
 - o Phân tích yêu cầu hệ thống
 - o Khảo sát các nguồn dữ liệu tuyển sinh hiện có
- Tuần 2-3 (29/09 - 12/10):
 - Thu thập dữ liệu từ website trường:
 - o Thu thập dữ liệu thông tin tuyển sinh
 - o Thu thập thông tin về ngành học, điểm chuẩn, học phí
 - o Lấy dữ liệu về quy chế, thời gian tuyển sinh
 - Thu thập tài liệu PDF:
 - o Download và phân loại các tài liệu PDF tuyển sinh
 - o Tài liệu hướng dẫn, quy chế, thông báo chính thức
- Tuần 4 (13/10 - 19/10):
 - Tiền xử lý và chuẩn hóa dữ liệu:
 - o Xử lý dữ liệu text từ website (làm sạch HTML, format)
 - o Trích xuất text từ file PDF
 - o Xử lý và mô tả nội dung hình ảnh (campus, cơ sở vật chất)
 - o Chuẩn hóa định dạng và cấu trúc dữ liệu
 - Thiết kế kiến trúc hệ thống cơ bản

Giai đoạn 2: Thiết kế chi tiết và Phát triển Core System (Tuần 5-8)

- Tuần 5 (20/10 - 26/10):
 - o Hoàn thiện thiết kế kiến trúc hệ thống
 - o Lựa chọn công nghệ và công cụ phù hợp
 - o Setup môi trường phát triển
- Tuần 6-7 (27/10 - 09/11):
 - o Xây dựng pipeline xử lý dữ liệu đa phương tiện
 - o Thiết kế và triển khai vector database

- Phát triển module embedding cho text, image, PDF
- Tuần 8 (10/11 - 16/11):
 - Phát triển module Multimodal RAG
 - Tích hợp Large Language Model
 - Test cơ bản các module riêng lẻ

Giai đoạn 3: Phát triển Giao diện và Tích hợp (Tuần 9-10)

- Tuần 9 (17/11 - 23/11):
 - Phát triển giao diện người dùng (chatbot interface)
 - Tích hợp frontend với backend
 - Xây dựng API endpoints
- Tuần 10 (24/11 - 30/11):
 - Tích hợp toàn hệ thống
 - Xây dựng hệ thống quản trị cơ bản
 - Test tích hợp và sửa lỗi

Giai đoạn 4: Testing, Đánh giá và Hoàn thiện (Tuần 11-12)

- Tuần 11 (01/12 - 07/12):
 - Kiểm thử chức năng và hiệu suất toàn hệ thống
 - Đánh giá độ chính xác với bộ câu hỏi test
 - Thu thập feedback và tối ưu hóa
 - Chuẩn bị dữ liệu đánh giá

Tuần 12 (08/12 - 14/12):

- Hoàn thiện tài liệu báo cáo đồ án
- Chuẩn bị slide thuyết trình
- Tạo video demo
- Hoàn thiện mã nguồn và hướng dẫn cài đặt
- Chuẩn bị bảo vệ đồ án

Thành phố Hồ Chí Minh, ngày 9 tháng năm 2025

VIỆN TRƯỞNG
(Ký tên và ghi rõ họ tên)

GV HUỐNG DÂN
(Ký tên và ghi rõ họ tên)



Lê Duy Hùng