

BÀI TẬP THỰC TẬP
CÔNG TY TIỀN PHÒNG

LANGCHAIN & HỆ SINH THÁI LLM

MÃ SINH VIÊN: 224890

HỌ VÀ TÊN: TÔ KHÁNH LINH

NGÀNH: KHOA HỌC MÁY TÍNH

TRƯỜNG: ĐẠI HỌC NAM CẦN THƠ

GIẢNG VIÊN HƯỚNG DẪN: TS.NGÔ HỒ ANH KHÔI

BÀI TẬP THỰC TẬP CÔNG TY TIỀN PHONG

I. TỔNG QUAN LANGCHAIN

1. LANGCHAIN LÀ GÌ?

LANGCHAIN LÀ MỘT FRAMEWORK MÃ NGUỒN MỞ ĐƯỢC THIẾT KẾ ĐỂ XÂY DỰNG CÁC ỨNG DỤNG SỬ DỤNG LARGE LANGUAGE MODELS (LLMS) MỘT CÁCH CÓ CẤU TRÚC.

LANGCHAIN KHÔNG PHẢI LÀ MỘT MÔ HÌNH AI, MÀ LÀ MỘT LỚP TRUNG GIAN (ORCHESTRATION LAYER) GIÚP:

- TỔ CHỨC PROMPT
- KẾT NỐI LLM VỚI DỮ LIỆU NGOÀI
- QUẢN LÝ LƯỒNG XỬ LÝ
- XÂY DỰNG HỆ THỐNG AI PHỨC TẠP

BÀI TẬP THỰC TẬP CÔNG TY TIỀN PHONG

II. Vai trò của LangChain trong hệ thống AI

TRONG HỆ THỐNG AI HIỆN ĐẠI, NẾU CHỈ GỌI API LLM TRỰC TIẾP:

- KHÔNG QUẢN LÝ ĐƯỢC CONTEXT DÀI
- KHÔNG CÓ BỘ NHỚ HỘI THOẠI
- KHÔNG KẾT NỐI ĐƯỢC DỮ LIỆU NỘI BỘ
- KHÔNG TÍCH HỢP CÔNG CỤ BÊN NGOÀI

BÀI TẬP THỰC TẬP CÔNG TY TIỀN PHONG

II. Vai trò của LangChain trong hệ thống AI

LANGCHAIN GIẢI QUYẾT CÁC VẤN ĐỀ TRÊN BẰNG CÁCH:

- TẠO CẤU TRÚC XỬ LÝ THEO CHUỖI (CHAIN)**
- TÍCH HỢP BỘ NHỚ (MEMORY)**
- KẾT NỐI VECTOR DATABASE**
- ĐIỀU PHỐI AGENT VÀ TOOL**

LANGCHAIN ĐÓNG VAI TRÒ “BỘ ĐIỀU PHỐI” TRONG KIẾN TRÚC AI.

BÀI TẬP THỰC TẬP CÔNG TY TIỀN PHONG

III Kiến trúc tổng thể LLM + LangChain

KIẾN TRÚC CƠ BẢN GỒM:

- 1. USER INPUT**
- 2. PROMPT TEMPLATE**
- 3. CHAIN / AGENT**
- 4. LLM**
- 5. MEMORY**
- 6. VECTOR DATABASE (NẾU CÓ RAG)**
- 7. OUTPUT RESPONSE**

LANGCHAIN HOẠT ĐỘNG NHƯ MỘT PIPELINE XỬ LÝ DỮ LIỆU.

BÀI TẬP THỰC TẬP CÔNG TY TIỀN PHÒNG

IV. Prompt Template

PROMPT TEMPLATE GIÚP CHUẨN HÓA CẤU TRÚC CÂU HỎI GỬI ĐẾN LLM.

VÍ DỤ:

BẠN LÀ TRỢ LÝ AI CỦA CÔNG TY.

DỰA TRÊN THÔNG TIN SAU: {CONTEXT}

HÃY TRẢ LỜI CÂU HỎI: {QUESTION}

LỢI ÍCH:

- TÁI SỬ DỤNG PROMPT
- TRÁNH PROMPT INJECTION
- CHUẨN HÓA LOGIC HỆ THỐNG

BÀI TẬP THỰC TẬP CÔNG TY TIỀN PHONG

V. Chain trong LangChain

1. LLMCHAIN

LUỒNG XỬ LÝ ĐƠN GIẢN:

PROMPT → LLM → OUTPUT

DÙNG CHO TÁC VỤ ĐƠN GIẢN NHƯ:

- TÓM TẮT VĂN BẢN**
- SINH NỘI DUNG**

BÀI TẬP THỰC TẬP CÔNG TY TIỀN PHONG

V. Chain trong LangChain

2. SEQUENTIAL CHAIN

KẾT NỐI NHIỀU CHAIN LIÊN TIẾP.

VÍ DỤ:

- 1. TÓM TẮT TÀI LIỆU**
- 2. PHÂN TÍCH NỘI DUNG**
- 3. SINH BÁO CÁO**

PHÙ HỢP HỆ THỐNG PHỨC TẠP.

BÀI TẬP THỰC TẬP CÔNG TY TIỀN PHONG

VI. Memory trong LangChain

MEMORY CHO PHÉP HỆ THỐNG:

- LƯU LỊCH SỬ HỘI THOẠI
- GIỮ NGỮ CẢNH GIỮA CÁC CÂU HỎI

VÍ DỤ:

USER: “CÔNG TY CÓ BAO NHIÊU PHÒNG BAN?”

USER: “PHÒNG IT CÓ BAO NHIÊU NHÂN VIÊN?”

MEMORY GIÚP HIỂU “PHÒNG IT” LIÊN QUAN ĐẾN CÂU HỎI TRƯỚC.

CÁC LOẠI MEMORY:

- CONVERSATIONBUFFERMEMORY
- CONVERSATIONSUMMARYMEMORY

BÀI TẬP THỰC TẬP CÔNG TY TIỀN PHONG

VI. Tool và Agent

TOOL

LÀ CÔNG CỤ BÊN NGOÀI NHƯ:

- API THỜI TIẾT**
- DATABASE**
- WEB SEARCH**
- PYTHON EXECUTION**

BÀI TẬP THỰC TẬP CÔNG TY TIỀN PHONG

VI. Tool và Agent

AGENT

AGENT QUYẾT ĐỊNH:

- KHI NÀO DỪNG LLM
- KHI NÀO GỌI TOOL
- TOOL NÀO CẦN SỬ DỤNG

AGENT GIÚP HỆ THỐNG AI CÓ KHẢ NĂNG RA QUYẾT ĐỊNH.

VII. Vector Database là gì?

VECTOR DATABASE LÀ HỆ THỐNG LƯU TRỮ EMBEDDING VECTOR.

EMBEDDING LÀ BIỂU DIỄN SỐ HỌC CỦA VĂN BẢN TRONG KHÔNG GIAN NHIỀU CHIỀU.

VÍ DỤ:

“NGHỈ PHÉP” VÀ “XIN NGHỈ” SẼ CÓ VECTOR GẦN NHAU.

VECTOR DB CHO PHÉP:

- SEMANTIC SEARCH**
- SIMILARITY SEARCH**
- RETRIEVAL BASED AI**

BÀI TẬP THỰC TẬP CÔNG TY TIỀN PHONG

VIII. Vai trò Vector DB trong RAG

RAG = RETRIEVAL AUGMENTED GENERATION

QUY TRÌNH:

- 1. CHUYỂN TÀI LIỆU THÀNH EMBEDDING**
- 2. LƯU EMBEDDING VÀO VECTOR DB**
- 3. KHI CÓ CÂU HỎI:**
- 4. CHUYỂN CÂU HỎI THÀNH EMBEDDING**
- 5. TÌM CÁC VECTOR GẦN NHẤT**
- 6. TRUY XUẤT NỘI DUNG LIÊN QUAN**
- 7. GỬI NỘI DUNG + CÂU HỎI VÀO LLM**
- 8. LLM SINH CÂU TRẢ LỜI**

RAG GIÚP LLM

- TRẢ LỜI DỰA TRÊN DỮ LIỆU NỘI BỘ**
- HẠN CHẾ HALLUCINATION**
- TĂNG ĐỘ CHÍNH XÁC**

BÀI TẬP THỰC TẬP

CÔNG TY TIỀN PHÒNG

IX. So sánh Vector Database

FAISS

- MÃ NGUỒN MỞ
- CHẠY LOCAL
- TỐC ĐỘ CAO
- KHÔNG CÓ API CLOUD

CHROMADB

- DỄ TÍCH HỢP LANGCHAIN
- PHÙ HỢP PROJECT NHỎ – TRUNG BÌNH
- CHẠY LOCAL HOẶC SERVER NHỎ

PINECONE

- CLOUD-BASED
- MỞ RỘNG TỐT
- PHÙ HỢP DOANH NGHIỆP LỚN

WEAVIATE

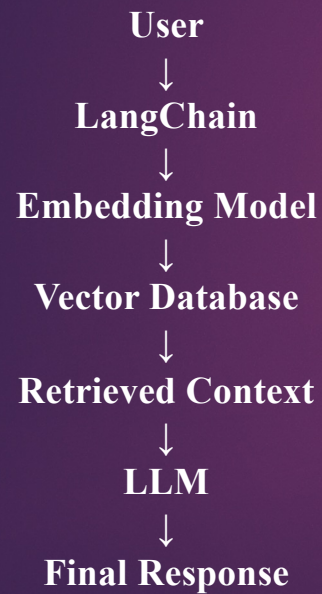
- HỖ TRỢ HYBRID SEARCH
- TÍCH HỢP AI MODULE SẴN

BÀI TẬP THỰC TẬP CÔNG TY TIỀN PHONG

X. Quy trình triển khai hệ thống hỏi đáp nội bộ

- 1. THU THẬP TÀI LIỆU CÔNG TY (PDF, DOCX)**
- 2. TÁCH NỘI DUNG THÀNH CHUNKS**
- 3. TẠO EMBEDDING**
- 4. LƯU VÀO VECTOR DATABASE**
- 5. USER ĐẶT CÂU HỎI**
- 6. TRUY VẤN SEMANTIC SEARCH**
- 7. GỬI KẾT QUẢ VÀO LLM**
- 8. TRẢ LỜI CHÍNH XÁC DỰA TRÊN DỮ LIỆU NỘI BỘ**

XI. Mô hình tổng thể



Giải thích kiến trúc

- User gửi câu hỏi
- LangChain tạo embedding cho câu hỏi
- Truy vấn Vector Database bằng similarity search
- Lấy context phù hợp
- Gửi context + câu hỏi vào LLM
- LLM sinh câu trả lời

XII. Kết luận chuyên sâu

LANGCHAIN GIÚP XÂY DỰNG HỆ THỐNG AI CÓ KIẾN TRÚC RÕ RÀNG VÀ MỞ RỘNG ĐƯỢC.

VECTOR DATABASE LÀ THÀNH PHẦN CỐT LÕI TRONG HỆ THỐNG RAG.

MÔ HÌNH LLM + LANGCHAIN + VECTOR DB GIÚP XÂY DỰNG HỆ THỐNG HỎI ĐÁP DỮ LIỆU NỘI BỘ HIỆU QUẢ, CHÍNH XÁC VÀ CÓ KHẢ NĂNG MỞ RỘNG.