

BÁO CÁO CUỐI KỲ
NHẬP MÔN NGÔN NGỮ HỌC THỐNG KÊ & ỨNG DỤNG

**CHATBOT RAG VỀ CHƯƠNG
TRÌNH ĐÀO TẠO FITHCMUS
CHƯƠNG TRÌNH CHUẨN**

**GVHD: Lê Thanh Tùng
Nhóm 5**

THÀNH VIÊN NHÓM

MSSV	Họ và Tên
20120145	Đường Yến Ngọc
21120198	Nguyễn Thị Lan Anh
21120417	Nguyễn Thị Ngọc Châm
21120426	Huỳnh Phát Đạt

NỘI DUNG

- I. Giới thiệu đồ án**
- II. Giới thiệu tập dữ liệu**
- III. Xây dựng ứng dụng**
- IV. Đánh giá kết quả**
- V. Demo sản phẩm**

I. GIỚI THIỆU ĐỒ ÁN

NỘI DUNG ĐỒ ÁN

- **Mục tiêu:** Xây dựng một hệ thống chatbot RAG có thể truy xuất thông tin chương trình đào tạo và trả lời các câu hỏi bằng ngôn ngữ tự nhiên.

NỘI DUNG ĐỒ ÁN

- **Output:** “Bạn cần tích lũy tối thiểu 16 tín chỉ cho các học phần bắt buộc chuyên ngành”
- **Input:** “Tôi cần tích lũy bao nhiêu tín chỉ bắt buộc chuyên ngành?”

PHƯƠNG PHÁP TIẾP CẬN

- Retrieval-based (Dựa trên truy xuất thông tin)
 - Generation-based (Dựa trên mô hình sinh câu trả lời)
- Retrieval-Augmented Generation (RAG)

III. GIỚI THIỆU TẬP DỮ LIỆU

NGUỒN DỮ LIỆU

- **Dữ liệu được thu thập và tổng hợp từ các nguồn:**
 - **Đề cương môn học:** Trang web khoa FITHCMUS > Đào tạo > Chương trình đào tạo > Chương trình Chuẩn (<https://www.fit.hcmus.edu.vn/vn/Default.aspx?tabid=36>)
 - **Chương trình đào tạo:** Trang web khoa FITHCMUS > Hệ thống sinh viên > CQuy > Chương trình đào tạo > CQuy Khóa tuyển 2023
 - (<https://www.fit.hcmus.edu.vn/vn/Default.aspx?tabid=289>)
 - **Câu hỏi và trả lời mẫu:** Trang Q&A FIT (<https://courses.fit.hcmus.edu.vn/q2a/>)

THU THẬP DỮ LIỆU

7. NỘI DUNG CHƯƠNG TRÌNH ĐÀO TẠO

7.1. KIẾN THỨC GIÁO DỤC ĐẠI CƯƠNG

Tích lũy tổng cộng 56 tín chỉ (không kể Ngoại ngữ, Giáo dục thể chất và Giáo dục quốc phòng – an ninh):

7.1.1. Lý luận chính trị – Pháp luật

STT	MÃ HỌC PHẦN	TÊN HỌC PHẦN	SỐ TC	SỐ TIẾT			Loại học phần	Ghi chú
				Lý thuyết	Thực hành	Bài tập		
1	BAA00101	Triết học Mác - Lênin	3	45	0	0	BB	
2	BAA00102	Kinh tế chính trị Mác - Lênin	2	30	0	0	BB	
3	BAA00103	Chủ nghĩa xã hội khoa học	2	30	0	0	BB	
4	BAA00104	Lịch sử Đảng Cộng sản Việt Nam	2	30	0	0	BB	
5	BAA00003	Tư tưởng Hồ Chí Minh	2	30	0	0	BB	
6	BAA00004	Pháp luật đại cương	3	45	0	0	BB	
TỔNG CỘNG			14					

7.1.2. Khoa học xã hội – Kinh tế – Kỹ năng

STT	MÃ HỌC PHẦN	TÊN HỌC PHẦN	SỐ TC	SỐ TIẾT			Loại học phần	Ghi chú
				Lý thuyết	Thực hành	Bài tập		
1 Chọn 01 học phần (02 tín chỉ) trong các học phần sau:								
	BAA00005	Kinh tế đại cương	2	30	0	0	TC	
	BAA00006	Tâm lý đại cương	2	30	0	0	TC	
	BAA00007	Phương pháp luận sáng tạo	2	30	0	0	TC	

The screenshot shows the FIT@HCMUS website interface. At the top right, there are links for 'ĐĂNG NHẬP' (Login), 'Webmail | Trang môn học | Liên hệ' (Webmail | Course page | Contact), 'Trang chủ | Góp ý | English' (Home | Suggestion | English), and a dropdown menu. The main header features the logo 'TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN ĐHQG-HCM' and the text 'fit@hcmus'. Below the header is a navigation bar with icons for 'TRANG CHỦ' (Home), 'TIN TỨC' (News), 'GIỚI THIỆU' (Introduction), 'NGHIÊN CỨU' (Research), and 'ĐÀO TẠO' (Education). The central content area displays two boxes: 'Lĩnh vực đào tạo' (Fields of study) listing 'Khoa học máy tính' (Computer Science) and its sub-fields: Kỹ thuật phần mềm, Hệ thống thông tin, Công nghệ thông tin, and Trí tuệ nhân tạo; and 'Kiến thức giáo dục đại cương (bắt buộc)' (Compulsory general education knowledge) listing 19 courses from BAA00003 to MTH00050.

Hình 1. Dữ liệu về chương trình đào tạo được tổng hợp từ trang web chính thức FITHCMUS

THU THẬP DỮ LIỆU

Các qui định		
Tiêu đề	Category	Modified Date
QĐ 2380 sửa đổi bổ sung QĐ 1625 K2022		17/09/2024 Tài Xuống
Quy định về việc gia hạn thu học phí QĐ 2165/QĐ-KHTN, 26/10/2023		10/11/2023 Tài Xuống
Tiêu chuẩn đăng ký KLTN áp dụng cho Khóa 2011-2014		11/10/2022 Tài Xuống
Hướng dẫn cách thức nộp KLTN- Cử nhân Tài năng K2015	CNTN	29/09/2021 Tài Xuống
Quy định chuẩn đầu ra ngoại ngữ CT Cử nhân tài năng Khóa 2023 về sau	CNTN	11/03/2024 Tài Xuống
DHCQ-Quy chế đào tạo 2021	QCHV	11/10/2022 Tài Xuống
QĐ 1175-DHCQ-Quy chế đào tạo 2021	QCHV	11/10/2022 Tài Xuống
Quy chế đào tạo HCTC 2016	QCHV	11/10/2022 Tài Xuống
QĐ 534 - sửa đổi chuẩn TA và trình độ TA đối với ĐHCQ_ngày 15062020	QĐ TA	11/10/2022 Tài Xuống
QĐ về xét chuẩn trình độ ngoại ngữ đối với chứng chỉ VNU-EPT - 04/2022	QĐ TA	11/10/2022 Tài Xuống

Hình 2. Các qui định, hướng dẫn chung

THU THẬP DỮ LIỆU

The image displays two screenshots of the Q&A FIT website. The left screenshot shows a list of recent questions under the heading "Những câu hỏi gần đây". The right screenshot shows a detailed view of a specific question.

Left Screenshot (Recent Questions):

- Kỳ 2 có thường mở lớp Phát triển ứng dụng web/mobile/game nâng cao không?**
0 k thích | 0 trả lời
đã hỏi ngày 28 tháng 8 năm 2023 trong KTPM bởi Trung Nguyên Hồ (120 điểm)
- hỏi về các môn tốt nghiệp**
0 k thích | 1 trả lời
đã hỏi ngày 21 tháng 8 năm 2023 trong Giáo vụ đại học bởi Minh Chánh Cái (120 điểm)
- Dạ cho em hỏi muốn du học đối với ngành cnnt thì cần những điều kiện gì ạ**
0 k thích | 1 trả lời
đã hỏi ngày 19 tháng 8 năm 2023 trong Học phí, học bổng bởi Thủ Ngữ 1 (120 điểm)
- Xét chuyên ngành**
0 k thích | 1 trả lời
đã hỏi ngày 29 tháng 7 năm 2023 trong Giáo vụ đại học bởi Toan NGuyen (120 điểm)
- Về cách tính điểm xét chuyên ngành**
0 k thích | 1 trả lời
đã hỏi ngày 24 tháng 7 năm 2023 trong Chương trình chuẩn bởi Kitka Tula (140 điểm)
- Có cần trả hết các môn còn nợ hay không?**
0 k thích | 1 trả lời
đã hỏi ngày 22 tháng 7 năm 2023 trong Chương trình chuẩn bởi Thanh Lộc Ngô (120 điểm)

Right Screenshot (Detailed View):

Dạ cho em hỏi là điều kiện để xét chuyên ngành của nhóm ngành máy tính và công nghệ thông tin cuối năm 2 là gì vậy ạ!

0 k thích | 326 đã xem

1 Câu trả lời

0 k thích | 359 đã xem

trả lời | bình luận

Chào em, điều kiện đối với sinh viên để được xét chuyên ngành cuối học kỳ 4 của khóa học là sinh viên **đang học** có tổng số tín chỉ tích lũy đạt **ít nhất 50 tín chỉ** (sau 4 học kỳ chính).

Hình 3. Thu thập câu hỏi - câu trả lời từ web Q&A FIT

THU THẬP DỮ LIỆU

◀ Data 10 items

Name	Last modified	File size
BSMS.txt	Dec 4, 2024	5 KB
CNTT.txt	Dec 14, 2024	47 KB
DKTN.txt	Dec 14, 2024	9 KB
HTTT.txt	Dec 14, 2024	32 KB
KHMT.txt	Dec 14, 2024	67 KB
KTPM.txt	Dec 14, 2024	32 KB
MonHoc.txt	Dec 4, 2024	474 KB
NN.txt	Dec 14, 2024	13 KB
QADATA.xlsx	Dec 14, 2024	9 KB
TTNT.txt	Dec 14, 2024	33 KB

Chương trình: Đại học Chính quy

Ngành: Hệ thống thông tin

Khóa tuyển: 2023

(Ban hành kèm theo Quyết định số 1712/QĐ - KHTN ngày 07/9/2023
của Hiệu trưởng Trường Đại học Khoa học Tự nhiên, ĐHQG-HCM)

1. Thông tin chung về chương trình đào tạo:

1.1. Tên ngành:

- Tiếng việt: Hệ thống thông tin

- Tiếng Anh: Information Systems

1.2. Mã ngành: 7480104

1.3. Trình độ đào tạo: Đại học

1.4. Tên chương trình: Cử nhân Hệ thống thông tin

1.5. Loại hình đào tạo: Chính quy

1.6. Thời gian đào tạo: 4 năm

1.7. Tên văn bằng sau khi tốt nghiệp:

- Tiếng Việt: Cử nhân Hệ thống thông tin

- Tiếng Anh: Bachelor of Science in Information Systems

1.8. Ngôn ngữ giảng dạy: tiếng Việt

1.9. Nơi đào tạo:

- Cơ sở 1: 227 Nguyễn Văn Cừ, P4, Q5, Tp. HCM

- Cơ sở 2: Phường Linh Trung, Thành phố Thủ Đức, Tp. HCM

Hình 4. Dữ liệu được chuyển sang dạng text

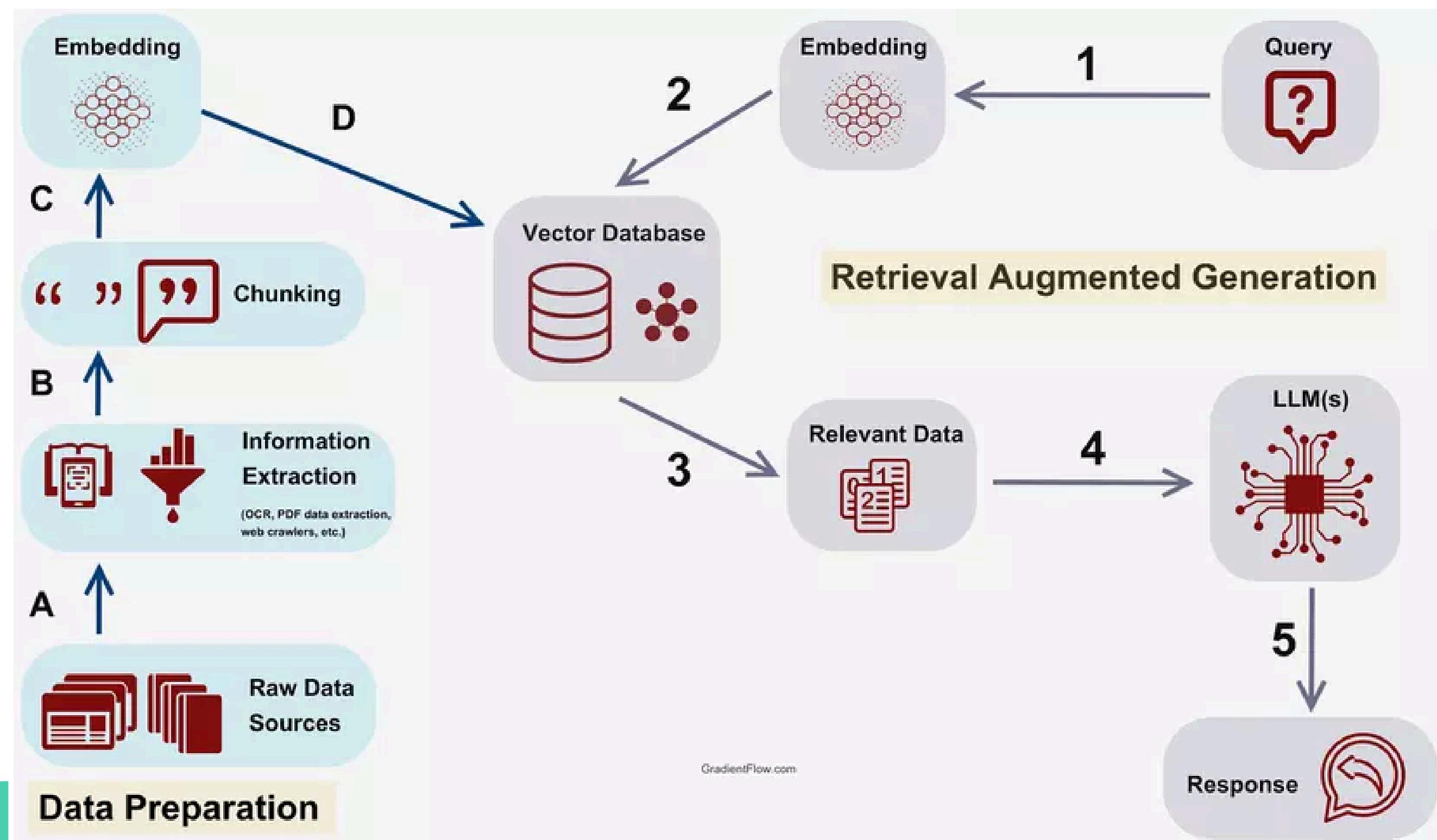
III. XÂY DỰNG ỨNG DỤNG

1. NỀN TẢNG HỖ TRỢ

NỀN TẢNG HỖ TRỢ

- **Ngôn ngữ lập trình:**
 - Back-end: Python, thư viện chính Langchain.
 - Front-end: JavaScript, thư viện Vite React.
- **Mô hình generate:** Vi-Qwen2-3B-RAG, gemini-1.5-flash
- **Mô hình embedding:** LaBSE, multilingual-e5-base
- **Môi trường:** Google Colab, VS Code

RAG PIPELINE



2. XỬ LÝ VÀ NẠP DỮ LIỆU

TẬP DỮ LIỆU DẠNG TEXT

Số lượng tài liệu: 10, Kiểu dữ liệu của documents: <class 'list'>

Nội dung của tài liệu đầu tiên: Ngành Hệ thống thông tin

Chương trình: Đại học Chính quy

Khóa tuyển: 2023

(Ban hành kèm theo Quyết định số 1712/QĐ - KHTN ngày 07/9/2023

của Hiệu trưởng Trường Đại học Khoa học Tự nhiên, ĐHQG-HCM)

1. Thông tin chung về chương trình đào tạo Ngành Hệ thống thông tin:

1.1. Tên ngành:

- Tiếng việt: Hệ thống thông tin
- Tiếng Anh: Information Systems

1.2. Mã ngành: 7480104

1.3. Trình độ đào tạo: Đại học

1.4. Tên chương trình: Cử nhân Hệ thống thông tin

1.5. Loại hình đào tạo: Chính quy

1.6. Thời gian đào tạo: 4 năm

1.7. Tên văn bằng sau khi tốt nghiệp:

- Tiếng Việt: Cử nhân Hệ thống thông tin
- Tiếng Anh: Bachelor of Science in Information Systems

1.8. Ngôn ngữ giảng dạy: tiếng Việt

1.9. Nơi đào tạo:

- Cơ sở 1: 227 Nguyễn Văn Cừ, P4, Q5, Tp. HCM
- Cơ sở 2: Phường Linh Trung, Thành phố Thủ Đức, Tp. HCM

CHUNKING

Chunk 33:

7.1.3. Toán - Khoa học tự nhiên - Công nghệ - Môi trường

- Tên học phần: Vi tích phân 2B
- Mã học phần: MTH00004
- Số tín chỉ: 3
- Loại học phần: Bắt buộc

- Tên học phần: Thực hành Vi tích phân 2B

- Mã học phần: MTH00082
- Số tín chỉ: 1
- Loại học phần: Bắt buộc

- Tên học phần: Đại số tuyến tính

- Mã học phần: MTH00007
- Số tín chỉ: 3
- Loại học phần: Bắt buộc

- Tên học phần: Toán rời rạc

- Mã học phần: MTH00041
 - Số tín chỉ: 3
 - Loại học phần: Bắt buộc
- Metadata: {'source': '/content/Data/KTPM.txt'}
-

In 5 chunk đầu tiên

Chunk 34:

7.1.3. Toán - Khoa học tự nhiên - Công nghệ - Môi trường

- Mã học phần: MTH00083
- Số tín chỉ: 1
- Loại học phần: Bắt buộc

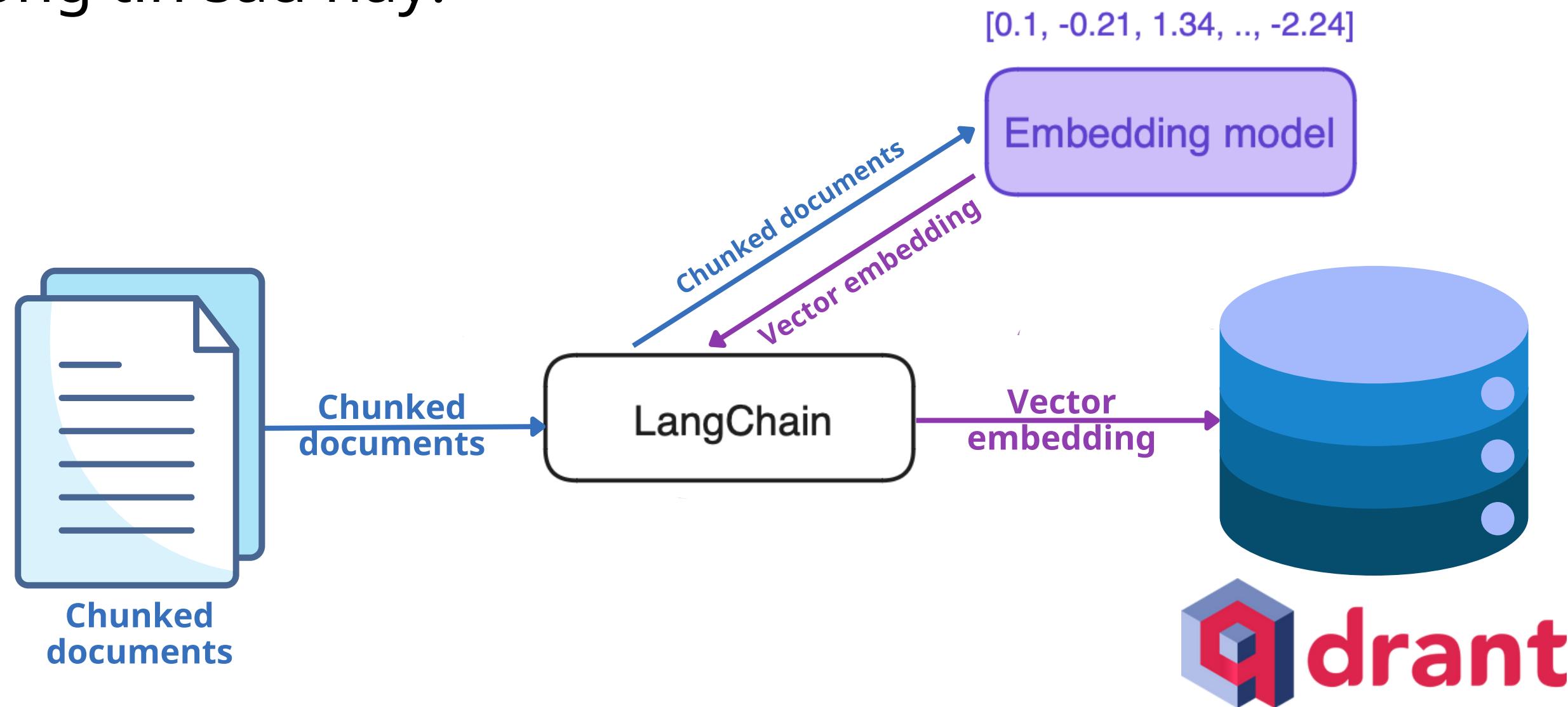
- Tên học phần: Xác suất thống kê

- Mã học phần: MTH00008
- Số tín chỉ: 3
- Loại học phần: Bắt buộc

- Tên học phần: Thực hành Xác suất thống kê

VECTOR DATABASE

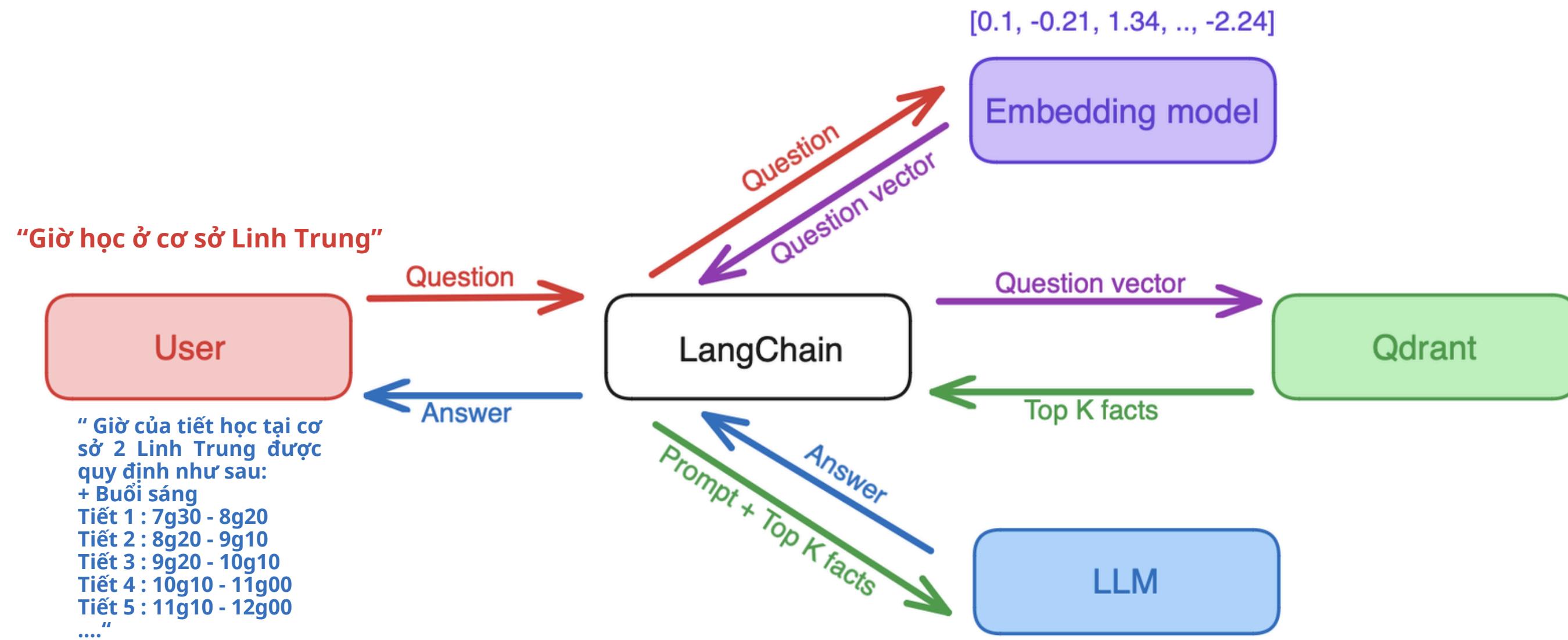
- **Vector Database:** Các vector nhúng được lưu trữ trong cơ sở dữ liệu chỉ mục (index) để tối ưu hóa việc truy xuất thông tin sau này.



3. TRUY XUẤT THÔNG TIN

TRUY XUẤT THÔNG TIN

- Kết hợp mô hình ngôn ngữ lớn để truy xuất và sinh câu trả lời tự nhiên trả về cho người dùng.



TRUY XUẤT THÔNG TIN

- Promt template:

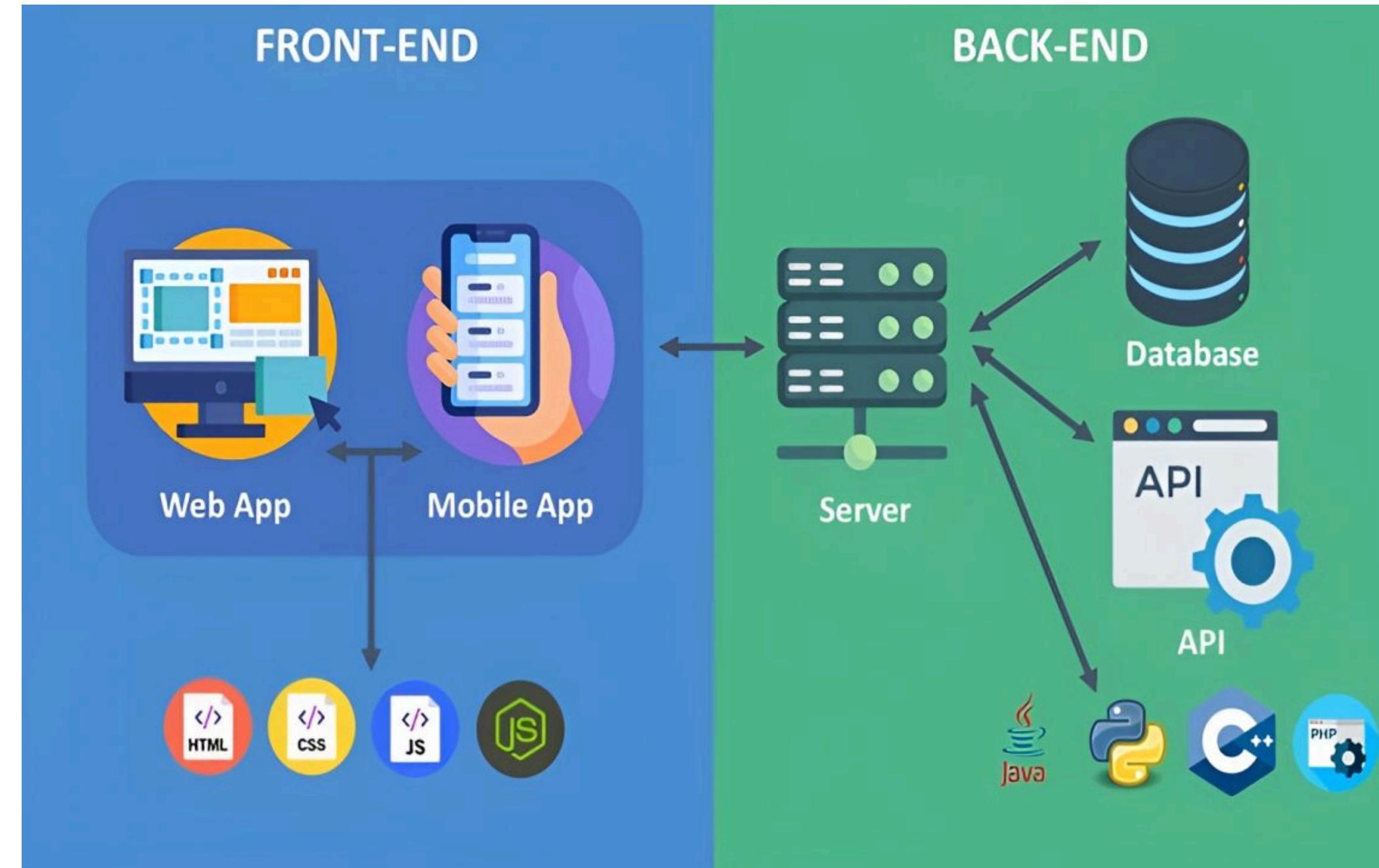
```
<|system|>
'''Thực hiện trả lời câu hỏi từ thông tin có trong ngữ cảnh được cho. Chú ý các yêu cầu sau:
- Câu trả lời phải chính xác, đầy đủ và có liên quan đến câu hỏi.
- Chỉ sử dụng các thông tin có trong ngữ cảnh được cung cấp.
- Nếu có nhiều ngữ cảnh liên quan với câu hỏi thì kết hợp các ngữ cảnh để tổng hợp thông tin.
- Nếu ngữ cảnh không chứa thông tin về câu trả lời thì từ chối trả lời và không suy luận gì th
"Vui lòng liên lạc Khoa Công Nghệ Thông Tin, trường Đại học Khoa Học Tự Nhiên - Đại học Quốc G
Địa chỉ: Phòng I.54, tòa nhà I, 227 Nguyễn Văn Cừ, Q.5, TP.HCM
Điện thoại: (028) 62884499
Email: info@fit.hcmus.edu.vn"
</s>
<|user|>
Hãy trả lời câu hỏi sau dựa trên ngữ cảnh sau:

{context}
---
Câu hỏi: {question}
</s>
<|assistant|>
'''
```

4. KẾT NỐI FRONT-END

KẾT NỐI FRONT-END

- Thực hiện kết nối với Front-end sử dụng thư viện Vite React của ngôn ngữ Javascript



KẾT NỐI FRONT-END

- Trang chủ của ứng dụng giới thiệu ngắn gọn về mục đích hoạt động của Chatbot

FITHCMUS Chatbot

TRANG CHỦ Trò chuyện FAQs Báo lỗi/ Góp ý

Xin chào! Mình là
FITHCMUS Chatbot

Giúp bạn giải đáp thắc mắc, tra cứu thông tin một cách nhanh chóng và chính xác nhất!

BẮT ĐẦU NGAY

KẾT NỐI FRONT-END

- Kết quả sau khi kết nối mô hình và Front-end để hoàn thiện ứng dụng.

FITHCMUS Chatbot

Trang chủ **TRÒ CHUYỆN** FAQs Báo lỗi/ Góp ý

Lịch sử trò chuyện

- Môn học "Nhập môn t..."
- Môn học CSC13116 là ...

Xin chào! Đây là FITHCMUS Chatbot, trợ lý đắc lực dành cho bạn! Bạn muốn tìm kiếm thông tin về những gì? Đừng quên chọn nguồn tham khảo phù hợp để mình có thể giúp bạn tìm kiếm thông tin chính xác nhất nha. 😊

Môn học CSC13116 là môn học gì?
Tham khảo: FITHCMUS

Môn học có mã học phần CSC13116 là môn Đồ án Công nghệ phần mềm

Môn học "Nhập môn tư duy tính toán" (CSC17105) yêu cầu môn học tiên quyết nào?
Tham khảo: FITHCMUS

Môn học "Nhập môn tư duy tính toán" (CSC17105) không yêu cầu môn học tiên quyết.

Nhập câu hỏi tại đây... 

Lưu ý: Mô hình có thể đưa ra câu trả lời không chính xác ở một số trường hợp, vì vậy hãy luôn kiểm chứng thông tin

Nguồn tham khảo

Wikipedia
 FITHCMUS

Những câu hỏi phổ biến

- Điều kiện nhận học bổng?
- Bao nhiêu điểm thì học lực Xuất sắc?
- Bao nhiêu điểm thì học lực Giỏi?
- Bao nhiêu điểm thì học lực Khá?

KẾT NỐI FRONT-END

- Vào trang FAQS để xem các thông tin cần thiết của ứng dụng

The screenshot shows a web page titled "Những câu hỏi thường gặp (FAQs)" (FAQs). The page has a header with navigation links: FITHCMUS Chatbot, Trang chủ, Trò chuyện, FAQS (which is highlighted with a blue border), and Báo lỗi/ Góp ý. Below the title, there are four expandable sections:

- Chatbot hoạt động như thế nào?** (minus sign icon)
Chatbot hoạt động bằng cách từ câu hỏi của người dùng, sử dụng kỹ thuật tìm văn bản liên quan đến câu hỏi trong bộ dữ liệu đã được vector hóa (text similarity) và lưu trữ thông qua vector database. Giúp lấy ra những đoạn văn bản có liên quan sau đó dùng mô hình ngôn ngữ lớn (LLM) Vietcuna để sinh câu trả lời.
- Cách sử dụng chatbot để tra cứu thông tin** (plus sign icon)
- Thông tin từ chatbot có đáng tin cậy không?** (minus sign icon)
Vì là một mô hình xác xuất nên thông tin chatbot đưa ra có thể không chính xác ở một số trường hợp, bạn nên kiểm chứng thông tin hoặc liên hệ hỗ trợ nếu cần thiết nhé
- Tôi có thể liên hệ hỗ trợ như thế nào?** (plus sign icon)

KẾT NỐI FRONT-END

- Trang Góp ý / Báo lỗi để nhóm có thể chỉnh sửa ứng dụng cho tốt hơn

The screenshot shows a web page with a light purple header containing the text 'Báo lỗi hoặc góp ý'. Below the header, there is a message: 'Sự đóng góp ý kiến từ các bạn sẽ là sự hỗ trợ đắc lực giúp chúng tôi ngày càng hoàn thiện sản phẩm hơn.' A text input field contains the placeholder 'Nêu cải thiện để trải nghiệm người dùng tốt hơn'. Below the input field is an email address: 'huynhphatdat@gmail.com'. At the bottom is a large blue button labeled 'GỬI Ý KIẾN'.

FITHCMUS Chatbot

Trang chủ Trò chuyện FAQs BÁO LỖI/ GÓP Ý

Báo lỗi hoặc góp ý

Sự đóng góp ý kiến từ các bạn sẽ là sự hỗ trợ đắc lực giúp chúng tôi ngày càng hoàn thiện sản phẩm hơn.

Nêu cải thiện để trải nghiệm người dùng tốt hơn

huynhphatdat@gmail.com

GỬI Ý KIẾN

KẾT NỐI FRONT-END

- Kết quả nhận được từ nội dung Góp ý / Báo lỗi

New message from huynhphatdat@gmail.com ⏺ Hộp thư đến x

Đ

Dat <huynhphatdat92@gmail.com>

đến tôi ▾

Hello ,

You got a new message from huynhphatdat@gmail.com:

Nên cải thiện để trải nghiệm người dùng tốt hơn

Best wishes,

huynhphatdat@gmail.com

Email sent via [EmailJS.com](#)

← Trả lời

→ Chuyển tiếp



KẾT NỐI FRONT-END

- Thực hiện Functional Testing bằng phương pháp Manual Testing:

1	Function/Feature ID	Case ID	Test case name	Test step	Expected Result (ER)	Actual Result	Status
2 Function 01: Header							
3	UC01	TC01	Điều hướng qua các trang khác nhau trên thanh navigation	Ở trang chủ ứng dụng, chọn nút "Trò chuyện" để đến trang Chat. Chọn nút "FAQS" để đến trang FAQ. Chọn nút "Báo lỗi/Góp ý" để đến trang Feedback.	Điều hướng người dùng đến trang Chat, FAQ, Feedback.	Giống expected result	Pass
4	UC01	TC02	Trở về trang chủ	Ở trang Chat, trang FAQ hoặc trang Feedback. Có thể trở về trang chủ thay vì chọn nút "Trang chủ" ở trên navigation, người dùng có thể chọn logo ở phía bên trái Header.	Điều hướng người dùng về trang chủ	Giống expected result	Pass
5 Function 02: Trang chủ							
6	UC02	TC01	Từ trang chủ điều hướng đến hỏi đáp với chatbot	Ở trang chủ, chọn nút "BẮT ĐẦU NGAY" ở giữa màn hình để chuyển đến trang hỏi đáp với chatbot.	Điều hướng người dùng đến trang Chat.	Giống expected result	Pass
7 Function 03: Chat page							
8	UC03	TC01	Hỏi đáp với Chatbot bằng cách nhập bàn phím	Ở trang Chat, người dùng nhập câu hỏi sau đó chọn nút bên phải màu xanh hoặc "Enter" để thực hiện hỏi đáp.	Chatbot trả lời câu hỏi	Giống expected result	Pass
	UC03	TC02	Hỏi đáp với Chatbot bằng câu hỏi có sẵn	Ở trang Chat, người dùng chọn câu hỏi có sẵn ở mục "Những câu hỏi phổ biến" ở bên phải sau đó chọn nút bên phải màu xanh hoặc	Chatbot trả lời câu hỏi	Giống expected result	Pass

IV. ĐÁNH GIÁ KẾT QUẢ

1. ĐÁNH GIÁ MÔ HÌNH

QA TEST DATA

question	expected_output	ground_truth
Tổng số tín chí bắt buộc của ngành Hệ thống thông tin là bao nhiêu?	Tổng số tín chí bắt buộc của ngành Hệ thi HTTT.txt	
Tên văn bằng ngành Khoa học máy tính sau khi tốt nghiệp	Tên văn bằng ngành Khoa học máy tính sá KHMT.txt	
Sinh viên phải làm gì để đăng ký học phần đầu môi học kỳ	Đầu môi học kỳ, sinh viên phải theo dõi tl QuyChe.txt	
Sinh viên phải học môn nào bắt buộc trong chuyên ngành	Sinh viên chuyên ngành Khoa học máy tín KHMT.txt	
Sinh viên phải hoàn thành bao nhiêu tín chí thuộc Kiến thức	Sinh viên phải hoàn thành 82 tín chí thuộc CNTT.txt	
Sinh viên phải đăng ký tối thiểu bao nhiêu tín chí trong học kỳ	Đối với chương trình Tiên tiến, Liên kết, C QuyChe.txt	
Sinh viên có thể kéo dài thời gian học tập tối đa bao lâu?	Sinh viên được phép kéo dài thêm không QuyChe.txt	
Sinh viên chuyên ngành Khoa học máy tính cần tích lũy bao nhiêu tín chí	Sinh viên chuyên ngành Khoa học máy tín KHMT.txt	
Sinh viên cần tích lũy ít nhất bao nhiêu tín chí từ học phần	Sinh viên cần tích lũy ít nhất 16 tín chí từ KHMT.txt	
Sinh viên cần tích lũy bao nhiêu tín chí cho phần tự chọn	Sinh viên cần tích lũy tối thiểu 8 tín chí ch KTPM.txt	
Sinh viên cần làm gì để được hoãn thi trong trường hợp bất thường	Trong trường hợp đột xuất và có lý do ch QuyChe.txt	
Sinh viên ngành Kỹ thuật phần mềm cần hoàn thành bao nhiêu tín chí	Sinh viên ngành Kỹ thuật phần mềm cần t KTPM.txt	
Sinh viên cần đạt điều kiện gì để đăng ký thực tập dự án tại nước ngoài	Sinh viên cần đạt 4 học phần Anh văn tro DKTN.txt	
Sinh viên cần đáp ứng những điều kiện nào để được công nhận tốt nghiệp	Sinh viên được xét và công nhận tốt nghi QuyChe.txt	
Những học phần nào được xem là bắt buộc trong chương trình	Học phần bắt buộc là học phần chứa đực QuyChe.txt	
Ngành Khoa học máy tính có tổng cộng bao nhiêu tín chí	Ngành Khoa học máy tính có tổng cộng 1 KHMT.txt	
Ngành Công nghệ thông tin được giáng dạy bằng ngôn ngữ	Ngành Công nghệ thông tin được giáng d: CNTT.txt	
Nếu không đạt học phần bắt buộc, sinh viên phải thực hiện	Sinh viên không đạt một học phần bắt bu QuyChe.txt	
Mục tiêu học phần của khóa học Thị giác robot bao gồm những gì	Mục tiêu học phần của khóa học Thị giác MonHoc.txt	
Mục tiêu học phần CSC17104 là gì?	Mục tiêu học phần CSC17104 bao gồm: S MonHoc.txt	
Mục tiêu của khóa học 'Thực hành Hóa đại cương 2' là gì?	Sau khi hoàn thành thành công học phần MonHoc.txt	
Một tín chí trong chương trình đào tạo đại học tương ứng	Một tín chí tương đương với 15 tiết học I QuyChe.txt	
Một tiết học được tính bằng bao nhiêu phút giảng dạy trực tuyến	Một tiết học được tính bằng 50 phút gián QuyChe.txt	
Một năm học đại học thường có bao nhiêu học kỳ và thời gian	Một năm học của Trường được tổ chức r QuyChe.txt	
Môn học CSC16112 yêu cầu bao nhiêu tín chỉ và có những nội dung	Môn học CSC16112 - Chuyên đề Xử lý ản MonHoc.txt	
Môn học CSC14008 Phương pháp nghiên cứu khoa học cụ thể	Môn học CSC14008 Phương pháp nghiên MonHoc.txt	
Môn học CSC13116 là môn học gì?	Môn học CSC13116 là Đồ án Công nghệ f MonHoc.txt	
Môn học CSC11103 - Thiết kế mang yêu cầu bao nhiêu tín	Môn học CSC11103 - Thiết kế mang yêu c MonHoc.txt	

ĐÁNH GIÁ RETRIEVAL

- Các collection chứa vector embedding được lưu tên theo dạng:
ITUS_Tên mô hình_Chunksize

Collection	Phương pháp chunking	Precision@5	Recall @5	MRR @5	MAP @5	nDCG @5
multilingual-e5-base						
ITUS e5 R 600	Recursive	0.7750	0.9838	0.9450	0.7911	0.8361
ITUS e5 R 800	Recursive	0.7708	0.9792	0.9277	0.8183	0.8562
ITUS e5 R 1000	Recursive	0.7264	0.9676	0.8983	0.7170	0.7726
ITUS e5 600	Bổ sung đề mục	0.8417	0.9630	0.9653	0.8524	0.8730
ITUS e5 800	Bổ sung đề mục	0.8236	0.9595	0.9528	0.8345	0.8548
ITUS e5 1000	Bổ sung đề mục	0.7806	0.9676	0.8980	0.7691	0.8137
LaBSE						
ITUS LaBSE R 600	Recursive	0.5750	0.9329	0.8220	0.6521	0.7168
ITUS LaBSE R 800	Recursive	0.5694	0.9062	0.7755	0.6538	0.7103
ITUS LaBSE R 1000	Recursive	0.5611	0.8981	0.7684	0.6322	0.6904
ITUS LaBSE 600	Bổ sung đề mục	0.6333	0.8935	0.8124	0.6667	0.7175
ITUS LaBSE 800	Bổ sung đề mục	0.6264	0.9132	0.8017	0.6785	0.7300
ITUS LaBSE 1000	Bổ sung đề mục	0.5847	0.9074	0.7799	0.6470	0.7019

ĐÁNH GIÁ RETRIEVAL

- Với cùng một mô hình embedding và phương pháp chunking giống nhau, các độ đo có xu hướng giảm khi chunk size tăng, điều này cho thấy chunk size càng lớn thì khả năng bị mất ngữ cảnh càng cao.
- Mô hình multilingual-e5-base luôn cho ra kết quả cao hơn mô hình LaBSE.

ĐÁNH GIÁ RETRIEVAL

- Dựa trên kết quả đánh giá, ITUS_e5_600 - Bổ sung đề mục có kết quả tổng thể cao nhất, nên nhóm sẽ chọn collection này để sử dụng cho retrieval:
 - **Mô hình embedding:** multilingual-e5-base
 - **Chunk size - overlap:** 600 - 200
 - **Phương pháp chunking:** Bổ sung đề mục

ĐÁNH GIÁ GENERATION

- Nhóm đã cài đặt RAG_pipeline trên 2 mô hình Vi-Qwen2-3B-RAG và gemini-1.5-flash, sau đó sử dụng các độ đo đánh giá để so sánh hiệu quả của 2 mô hình này.

Mô hình	BLEU	ROUGE-1	ROUGE-2	ROUGE-L	METEOR	SBERT Similarity
gemini-1.5-flash	0.6022	0.8253	0.7593	0.7705	0.7655	0.8808
Vi-Qwen2-3B-RAG	0.0614	0.2397	0.1992	0.2156	0.3339	0.8032

ĐÁNH GIÁ GENERATION

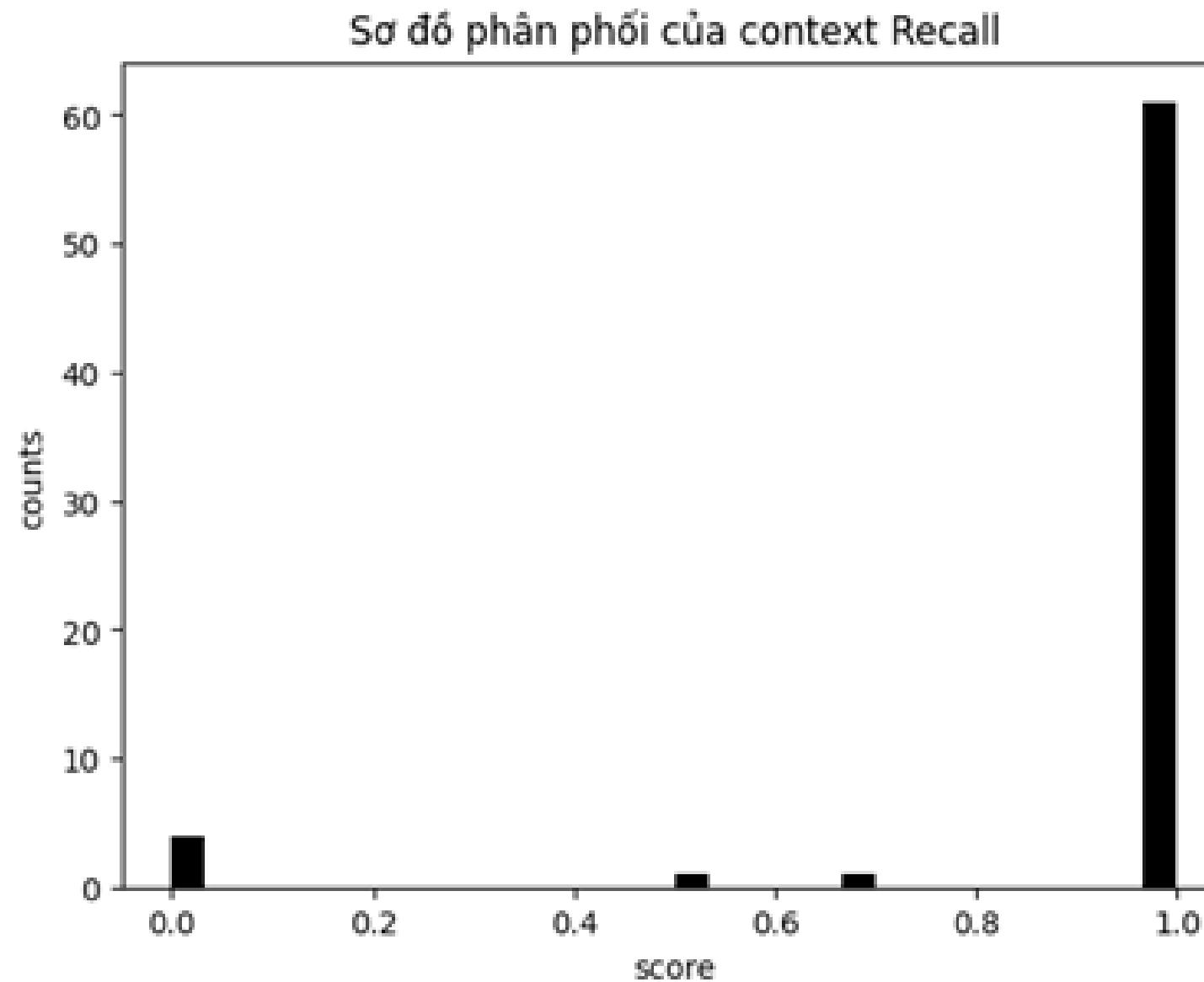
- gemini-1.5-flash thể hiện sự vượt trội rõ rệt ở hầu hết các chỉ số, từ BLEU, ROUGE, METEOR cho đến SBERT Similarity. Điều này cho thấy mô hình này tạo ra câu trả lời có sự trùng khớp cao về từ vựng, cấu trúc câu và ngữ nghĩa với câu trả lời chuẩn.
- Vi-Qwen2-3B-RAG có điểm số thấp, đặc biệt là ở các chỉ số ROUGE và BLEU. Mặc dù vẫn có sự tương đồng ngữ nghĩa (SBERT Similarity), mô hình này không thể tái tạo câu trả lời chính xác hoặc giữ vững cấu trúc tốt như gemini-1.5-flash.

MÔ HÌNH DÙNG CHO CHATBOT

- Dựa vào các đánh giá trước đó, nhóm lựa chọn mô hình '**gemini-flask-1.5**' kết hợp với collection '**ITUS_e5_600**' để xây dựng chatbot.

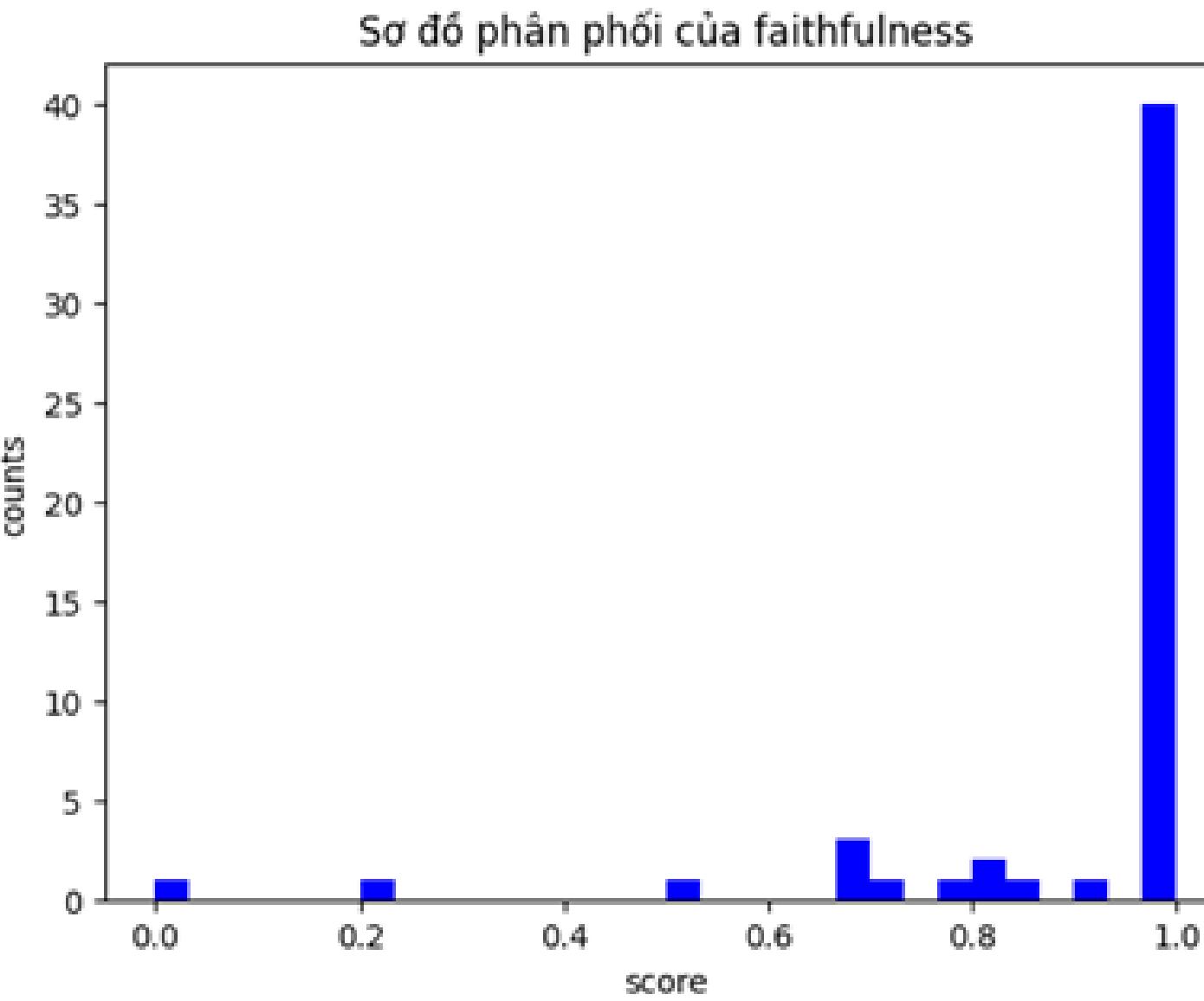
ĐÁNH GIÁ RAGAS

- **Context Recall:** Với chỉ số 0.9279, mô hình sử dụng ngữ cảnh truy xuất rất hiệu quả, gần như 93% câu trả lời sinh ra đều bám sát vào tài liệu truy xuất. Điều này cho thấy mô hình đã thành công trong việc tận dụng thông tin từ tài liệu nguồn.



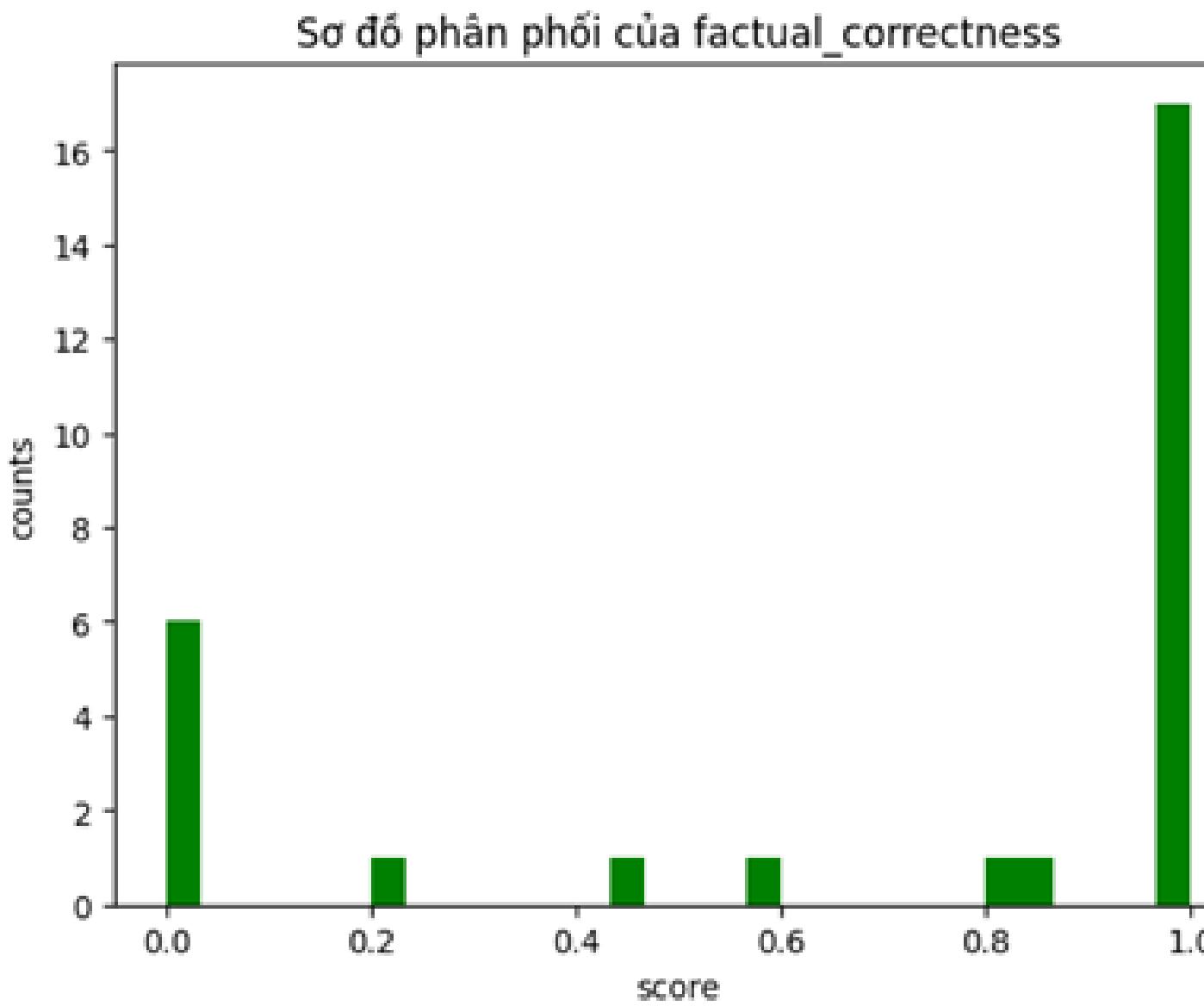
ĐÁNH GIÁ RAGAS

- **Faithfulness:** Mô hình khá trung thực với các tài liệu truy xuất, với 0.9144, cho thấy rằng mô hình hầu như không bị hiện tượng "hallucination" (bịa đặt) thông tin. Mô hình tuân thủ đúng ngữ cảnh và không đưa vào các chi tiết không chính xác.



ĐÁNH GIÁ RAGAS

- **Factual Correctness:** Đây là chỉ số khá quan trọng, phản ánh độ chính xác thực tế của câu trả lời. Mặc dù 0.7107 cho thấy rằng hơn 71% câu trả lời mô hình đưa ra là chính xác, vẫn còn khoảng 28% câu trả lời chưa đúng.



2. CÁC HƯỚNG CẢI TIẾN

CẢI THIỆN

- Tăng cường chất lượng dữ liệu đầu vào
 - Mở rộng tập dữ liệu đào tạo
 - Loại bỏ nhiễu
- Tối ưu hóa mô hình RAG
 - Cải thiện cách truy vấn như: reranking, cải thiện cấu trúc chia chunks, bổ sung metadata,...
 - Cải thiện prompt nếu có thể.
- Kiểm tra lại bộ câu hỏi kiểm tra, bổ sung các độ đo đánh giá khác để tăng tính tin cậy.
- Tối ưu hóa thời gian phản hồi và tăng khả năng chịu tải

V. DEMO SẢN PHẨM

KẾT QUẢ HỎI ĐÁP VỚI CHATBOT

FITHCMUS Chatbot Trang chủ TRÒ CHUYỆN FAQs Báo lỗi/ Góp ý

Lịch sử trò chuyện

Sinh viên cần chuẩn ...

Xin chào! Đây là FITHCMUS Chatbot, trợ lý đắc lực dành cho bạn! Bạn muốn tìm kiếm thông tin về những gì? Đừng quên chọn nguồn tham khảo phù hợp để mình có thể giúp bạn tìm kiếm thông tin chính xác nhất nha. 😊

Sinh viên cần chuẩn trình độ ngoại ngữ là bao nhiêu để tốt nghiệp?

Tham khảo: FITHCMUS

Sinh viên cần đạt tối thiểu bậc 3/6 theo khung năng lực ngoại ngữ 6 bậc dùng cho Việt Nam để tốt nghiệp đại học hệ chính quy chương trình đại trà. Điều này được quy định tại Thông tư số 01/2014/TT-BGDDT ngày 24 tháng 01 năm 2014 của Bộ Giáo dục và Đào tạo và Quyết định số 170/QĐ-DHQG ngày 27 tháng 02 năm 2018 của Giám đốc ĐHQG-HCM.

Tham khảo: Chương trình đào tạo khóa 2023 Chương trình đào tạo khóa 2023 Chương trình đào tạo khóa 2023
Chương trình đào tạo khóa 2023 Chương trình đào tạo khóa 2023 Chương trình đào tạo khóa 2023
Chương trình đào tạo khóa 2023 Chương trình đào tạo khóa 2023 Chương trình đào tạo khóa 2023
Chương trình đào tạo khóa 2023

Nguồn tham khảo

Wikipedia

FITHCMUS

của ngành/Chuyên ngành là bao nhiêu?

Liệt kê một số môn học bắt buộc của chuyên ngành Công nghệ tri thức?

Điều kiện thực hiện để tài tốt nghiệp là gì?

Sinh viên cần chuẩn trình độ ngoại ngữ là bao nhiêu để tốt nghiệp?

Chương trình đào tạo bao gồm bao nhiêu tín

Nhập câu hỏi tại đây...

Lưu ý: Mô hình có thể đưa ra câu trả lời không chính xác ở một số trường hợp, vì vậy hãy luôn kiểm chứng thông tin bạn nhé!

**CẢM ƠN THẦY VÀ CÁC BẠN
ĐÃ THEO DÕI**