**BÁO CÁO KẾT THÚC DỰ ÁN**

|  |  |
| --- | --- |
| **Thành viên:**  Phan Trung Thuận  Nguyễn Quang Tùng  Trần Trăng Sáng  Ung Khánh Như |  |
| **Table of Contents**  [I. Bài toán: 1](#_Toc177332400)  [II. Tổng quan hệ thống 1](#_Toc177332401)  [III. Phương pháp đề xuất: 2](#_Toc177332402)  [1. Chuẩn bị dữ liệu 2](#_Toc177332403)  [2. Xây dựng Retriever 4](#_Toc177332404)  [3. Xây dựng và huấn luyện Generator 4](#_Toc177332405)  [IV. Kết quả thực nghiệm: 5](#_Toc177332406)  [V. Kết luận: 9](#_Toc177332407) |  |

[Hình 1: Kiến trúc tổng quát của hệ thống 1](#_Toc177332349)

[Hình 2: Hình ảnh ví dụ về trang web thuvienphapluat.vn 3](#_Toc177332350)

[Hình 3: Mẫu lưu trữ cấu trúc dữ liệu 3](#_Toc177332351)

[Hình 4: Phân phối độ dài sequence của tập dữ liệu 4](#_Toc177332352)

[Hình 5: Danh sách các bộ dữ liệu đã được dùng để huấn luyện Vietnamese Llama2 7B 4](#_Toc177332353)

[Hình 6: Kiến trúc mô hình Llama2 5](#_Toc177332354)

[Hình 7: Biểu đồ hàm loss của Generator 6](#_Toc177332355)

[Bảng 1: Mô tả chi tiết mục tiêu và vai trò của các thành phần 2](#_Toc177332356)

[Bảng 2: Thông tin chi tiết về mô hình và tham số huấn luyện 5](#_Toc177332357)

[Bảng 3: Kết quả đánh giá Retriever 6](#_Toc177332358)

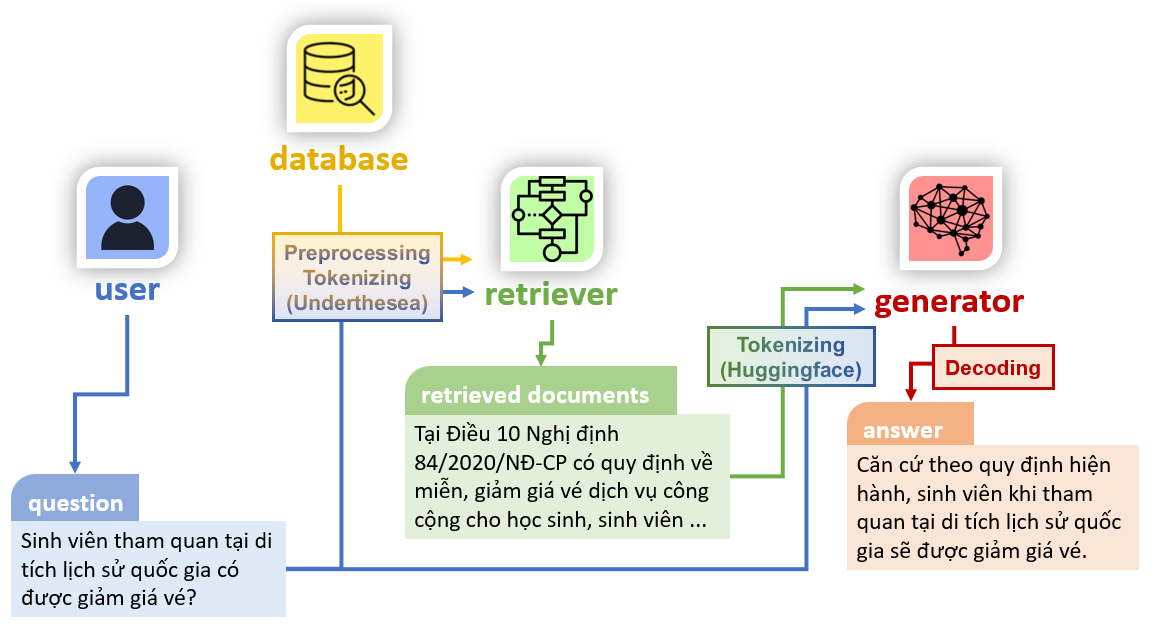
[Bảng 4: Một số câu trả lời được tạo ra khi tài liệu trích dẫn được cung cấp chính xác 7](#_Toc177332359)

[Bảng 5: Một số câu trả lời được sinh ra dựa vào các tài liệu trích dẫn từ Retriever 9](#_Toc177332360)

# I. Bài toán:

Xây dựng hệ thống hỏi đáp (question-answering system) về luật giáo dục đại học Việt Nam sử dụng kiến trúc retrieval augmentation generation (RAG).

# II. Tổng quan hệ thống



Hình 1: Kiến trúc tổng quát của hệ thống

Mô tả hệ thống: khi người dùng nhập vào câu hỏi, hệ thống sẽ xử lý câu hỏi đó và tìm tài liệu phù hợp, dựa vào tài liệu đó, hệ thống sẽ sinh ra câu trả lời phù hợp với câu hỏi.

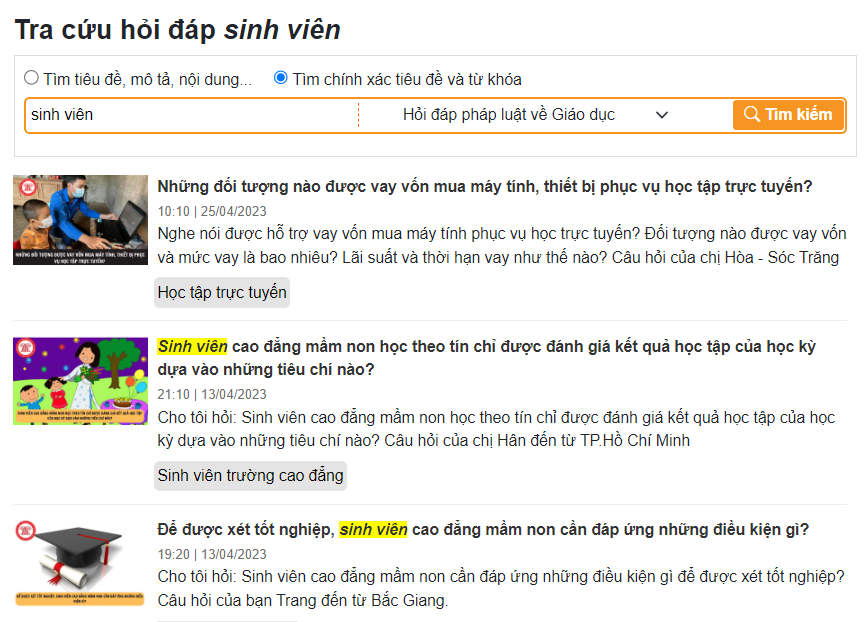
|  |  |  |
| --- | --- | --- |
|  | Vai trò | Mục tiêu/Yêu cầu |
| User | Nhập vào câu hỏi tiếng Việt liên quan đến các vấn đề giáo dục đại học | Nhắm đến nhóm người dùng là sinh viên, giảng viên đại học và cao đẳng trên toàn quốc có nhu cầu tìm hiểu về luật giáo dục đại học |
| Question | Là lời truy vấn để hệ thống có thể tìm kiếm tài liệu và sinh câu trả lời phù hợp | Câu hỏi rõ ràng, liên quan đến luật giáo dục đại học, không được lang mang, khó hiểu |
| Preprocessing | Xử lý các từ ngữ dư thừa, dấu câu để hỗ trợ Retriever tìm kiếm tài liệu chính xác hơn |  |
| Tokenizing | Tách các từ ra thành các token để giảm độ dài sequence, tăng hiệu suất tính toán của mô hình | Độ dài sequence tối đa là 4000 tokens |
| Retriever | Hệ thống thuật toán tìm kiếm tài liệu phù hợp từ bộ dữ liệu dựa trên câu hỏi được đặt ra | Cài đặt TFIDF với độ chính xác Acc@3 đạt 50% |
| Retrieved documents | Tài liệu đã được Retriever tìm thấy và phù hợp với câu hỏi | Chọn tối đa 3 tài liệu từ bộ dữ liệu để hỗ trợ trả lời câu hỏi |
| Generator | Mô hình ngôn ngữ cỡ vừa | Scale down pretrain model để phù hợp với lượng tài nguyên giới hạn |
| Decoding | Chuyển các token được tạo ra từ Generator thành câu trả lời |  |
| Answer | Câu trả lời hoàn chỉnh cho câu hỏi | BLEU đạt 0.5 và Perplexity đạt 0.5 |

Bảng 1: Mô tả chi tiết mục tiêu và vai trò của các thành phần

# III. Phương pháp đề xuất:

## Chuẩn bị dữ liệu

Dữ liệu được cào từ trang web thuvienphapluat.vn với từ khóa tìm kiếm là “sinh viên” và “giáo dục đại học”. Mỗi trang có thể có một hoặc nhiều cặp câu hỏi và câu trả lời thường sẽ được liệt kê ra ở mục “Nội dung chính”. Một trong những công việc khó nhất là select được thẻ HTML của các câu hỏi, những thẻ nằm giữa các câu hỏi hoặc câu hỏi đến cuối trang sẽ là tài liệu và câu trả lời cho câu hỏi trước nó. Các câu trả lời được nhận ra bằng các từ khóa như “theo đó”, “như vậy”, “trên đây là”, “căn cứ theo quy định nêu trên”, “căn cứ vào”, “đã trình bày như trên”, … Vậy phần còn lại giữa câu hỏi và câu trả lời là phần tài liệu trích dẫn. Tài liệu trích dẫn được tách ra thành câu trích dẫn và nội dung bằng cách xác định các từ khóa như “tại điều”, “theo mục”, “căn cứ khoản”, “căn cứ theo mục”, … Mỗi câu hỏi có thể trích dẫn nhiều tài liệu để phục vụ cho câu trả lời. Một số câu hỏi chỉ cần trích dẫn mà không cần câu trả lời.



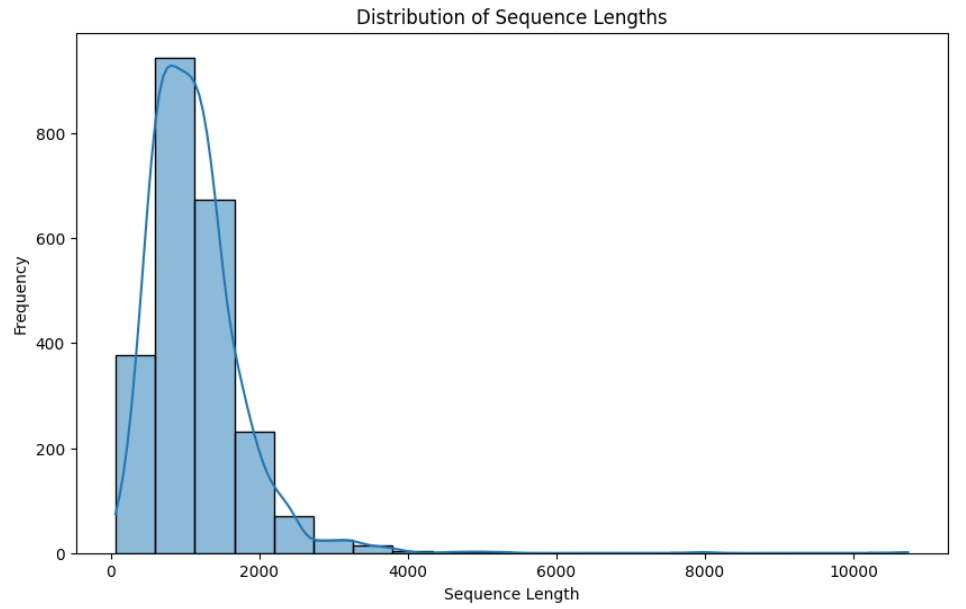
Hình 2: Hình ảnh ví dụ về trang web thuvienphapluat.vn

Sau khi có được dữ liệu thô, chúng em cũng cần phải xử lý những ngoại lệ như tài liệu trích dẫn có thể dạng bảng (khi cào về sẽ có rất nhiều thẻ và mỗi thẻ có rất ít nội dung) hoặc dạng ảnh (khi cào về thì phần nội dung trích dẫn sẽ không có gì cả vì không cào được ảnh) cần được loại bỏ. Sau khi xử lý đầy đủ, chúng em thu được bộ dữ liệu với 2345 bộ câu hỏi – câu trả lời với 2036 tài liệu để retriever truy xuất. Các tài liệu này gần như không thể chia nhỏ vì chúng thống nhất trong 1 bộ/điều/khoản/mục trong luật pháp nên bộ dữ liệu có thể lớn là điều có thể hiểu. Các thẻ rời rạc sẽ được gộp lại bằng ký tự xuống dòng và lưu bằng format như ảnh dưới.



Hình 3: Mẫu lưu trữ cấu trúc dữ liệu

Cấu trúc câu từ được tạo ra cho 1 sample như sau: “CÂU HỎI: ” + sample[‘question’] + “\nTRẢ LỜI: \n” + to\_string(sample[‘documents’]) + “\n” + sample[‘answer’]. Chúng em đã thống kê độ dài của sequence hoàn thiện này như hình bên dưới. Có thể thấy rằng sequence có độ dài tối đa 4000 là những gì chúng em cần cho bài toán này.



Hình 4: Phân phối độ dài sequence của tập dữ liệu

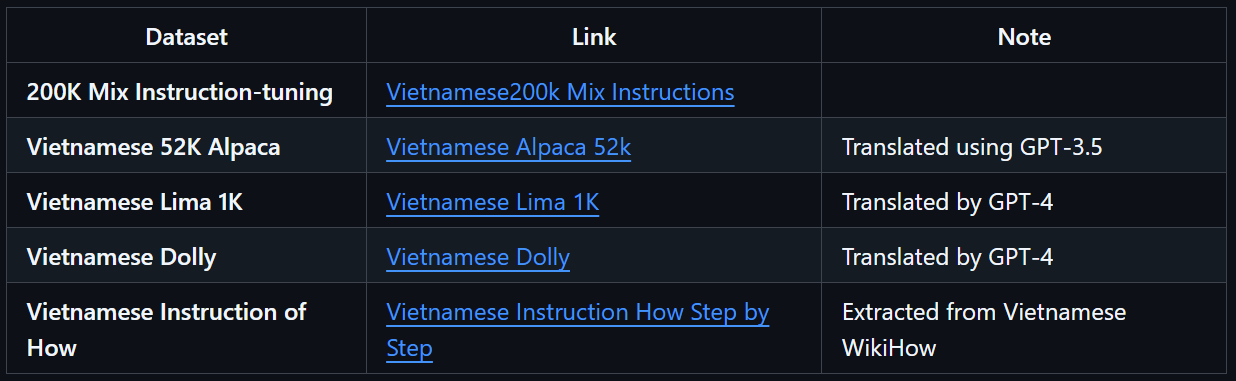
Chúng em chia bộ dữ liệu cho huấn luyện và đánh giá theo tỉ lệ 6:4 vì sự giới hạn về mặc tài nguyên cũng như chúng em muốn tìm tòi, học hỏi để thiết kế một mô hình nhỏ huấn luyện trên bộ dữ liệu vừa nhưng đạt được hiệu quả ấn tượng.

## Xây dựng Retriever

Chúng em sử dụng TF-IDF Retriever để tìm các tài liệu phù hợp với câu hỏi để hỗ trợ Generator sinh ra câu trả lời có chất lượng cao. Để xây dựng TFIDFIVectorizer từ thư viện Sklearn, chúng em cần tiền xử lý nội dung của tài liệu qua 3 bước chính: chuyển thành chữ viết thường, tokenizing (bằng thư viện Underthesea) và xóa stopword ([nguồn](https://github.com/stopwords/vietnamese-stopwords)). Chúng em sẽ áp dụng phương pháp tiền xử lý này với câu hỏi được người dùng nhập vào để vector hóa và tính cosine similarity giữa câu hỏi và các tài liệu để tìm 3 tài liệu phù hợp nhất.

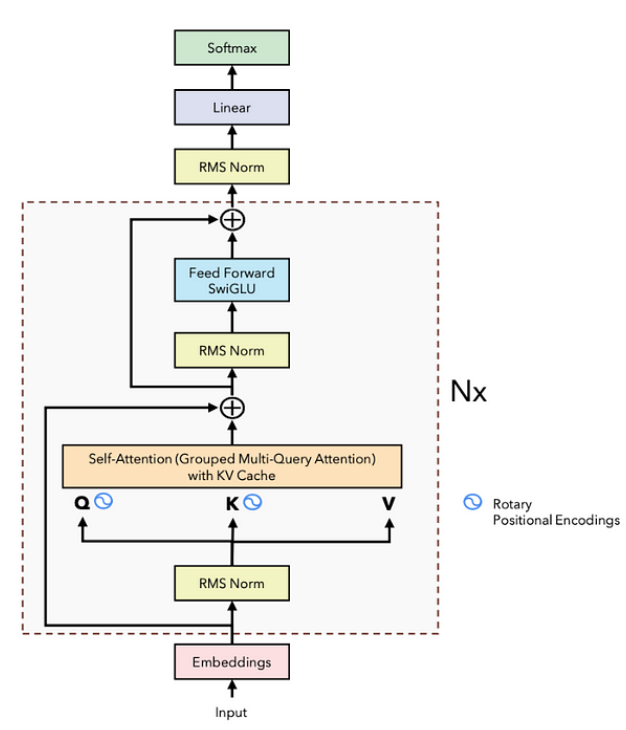
## Xây dựng và huấn luyện Generator

Để sinh ra câu trả lời chúng em sử dụng mô hình Llama2 7B 8K được huấn luyện trước dành cho tiếng Việt ([nguồn](https://huggingface.co/VietnamAIHub/Vietnamese_llama2_7B_8K_SFT_General_domain)) đã được huấn luyện trên các bộ dữ liệu rất lớn và tổng quát nhưng chúng em buộc phải giảm số lượng lớp decoder từ 32 xuống còn 16 vì lý do phần cứng. Chúng em sẽ fine tuning mô hình này với bộ dữ liệu đã được chuẩn bị trước với kỹ thuật qLoRA.



Hình 5: Danh sách các bộ dữ liệu đã được dùng để huấn luyện Vietnamese Llama2 7B

[Llama2](https://arxiv.org/pdf/2307.09288) là một kiến trúc Decoder-Only Transformer mạnh mẽ, áp dụng những ký thuật tiên tiến nhất hiện có bao gồm Rotary Positional Embedding (từ [Roformer](https://arxiv.org/pdf/2104.09864)) để biểu diễn thông tin về vị trí tốt hơn đặc biệt khi độ dài sequence lớn, Grouped Multi-Query Attention ([GQA](https://arxiv.org/pdf/2305.13245)) được đánh giá là đạt hiệu quả ngang ngữa với Multi-Head Attention nhưng tốc độ nhanh như Multi-Query Attention, kết hợp với KV Cache để tăng tốc độ inference (nguồn) và thay thế lớp Feedforward trong Tranformer bằng lớp SwiGLU đã được [chứng minh](https://arxiv.org/pdf/2002.05202v1) đạt hiệu quả cao.



Hình 6: Kiến trúc mô hình Llama2

Bên dưới là thông số mô hình và một số siêu tham số huấn luyện cho mô hình của chúng em:

|  |  |
| --- | --- |
| **Mô tả** | **Giá trị** |
| Tổng số lượng tham số | 1 895 305 216 |
| Số lượng tham số có thể huấn luyện | 14 024 704 (chiếm 0.74%) |
| Khích thước bộ vocab | 32000 |
| Độ dài sequence tối đa | 4000 |
| Độ lớn của d\_model | 4096 |
| Số lớp decoder | 16 |
| Rank của LoRA | 16 |
| Alpha của LoRA | 32 |
| Tỉ lệ Dropout của LoRA | 0.4 |
| Tỉ lệ học | 0.001 |
| Thuật toán tối ưu | AdamW |
| Kích thước batch | 3 |
| Số bước cộng dồn đạo hàm | 4 |
| Số lượng epoch | 2 |

Bảng 2: Thông tin chi tiết về mô hình và tham số huấn luyện

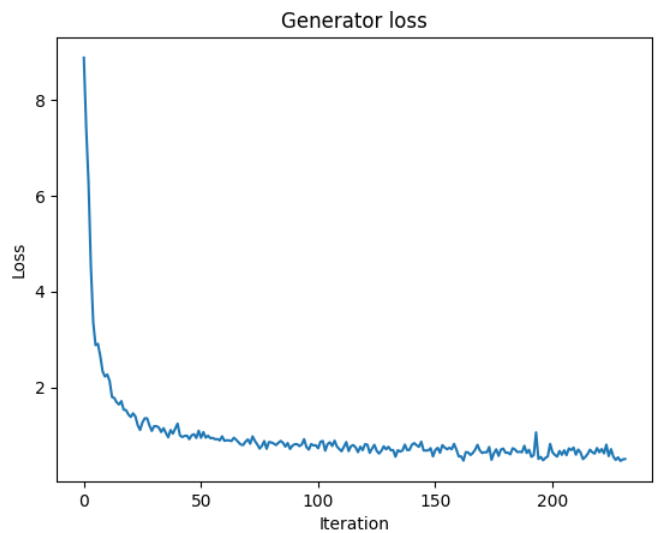
# IV. Kết quả thực nghiệm:

Do bộ dữ liệu chưa được xử lý tốt, chi tiết và Retriever TFIDF cũng gặp khó khăn khi tìm mối liên hệ giữa câu hỏi và tài liệu đặc biệt là các tài liệu về luật lại càng khó khăn hơn nên độ chính xác khi trích dẫn tài liệu không đạt như kỳ vọng..

|  |  |  |  |
| --- | --- | --- | --- |
| Retriever type | Acc@1 | Acc@3 | Acc@5 |
| TFIDF | 0.1842 | 0.3186 | 0.3928 |
| **BM25** | **0.2222** | **0.3667** | **0.4239** |

Bảng 3: Kết quả đánh giá Retriever

Đối với Generator thì do kiến trúc được thiết kế tỉ mỉ, mô hình không quá nhỏ và câu trả lời thường không quá dài và phức tạp nên nếu dẫn trích tài liệu (luật) đúng với câu hỏi thì câu trả lời được sinh ra tương đối tốt. Chúng em tính được Generator đạt **0.6355** điểm BLEU khi tài liệu được trích dẫn đúng, vượt qua kỳ vọng của chúng em.



Hình 7: Biểu đồ hàm loss của Generator

|  |  |  |
| --- | --- | --- |
| Câu hỏi | Tài liệu trích dẫn (từ test dataset, top3 tài liệu) | Câu trả lời |
| Học sinh có được trực tiếp đứng tên vay vốn mua máy tính, thiết bị phục vụ học tập trực tuyến không? | Căn cứ tại **Điều 4 Quyết định 09/2022/QĐ-TTg** có quy định về mức vốn cho vay như sau:  Phương thức cho vay  1. Việc cho vay đối với học sinh, sinh viên được thực hiện theo phương thức cho vay thông qua hộ gia đình. Đại diện hộ gia đình của học sinh, sinh viên là người đứng tên vay và giao dịch với Ngân hàng Chính sách xã hội.  2. Trường hợp đối tượng tại điểm b khoản 1 Điều 3 Quyết định này đã đủ 18 tuổi thì được trực tiếp đứng tên vay vốn tại Ngân hàng Chính sách xã hội nơi cư trú hoặc nơi nhà trường đóng trụ sở nếu hộ gia đình không còn thành viên nào đủ 18 tuổi trở lên hoặc thành viên còn lại không còn sức lao động, không có đủ năng lực hành vi dân sự đầy đủ theo quy định của pháp luật.  3. Việc cho vay của Ngân hàng Chính sách xã hội được thực hiện theo phương thức ủy thác cho các tổ chức chính trị - xã hội theo hợp đồng ủy thác hoặc trực tiếp cho vay đến người vay. | Như vậy, đối với học sinh, sinh viên thuộc trường hợp đối tượng tại điểm b khoản 1 Điều 3 Quyết định này đã đủ 18 tuổi thì được trực tiếp đứng tên vay vốn tại ngân hàng Chính sách xã hội nơi cư trú hoặc nơi nhà trường đóng trụ sở nếu hộ gia đình không còn thành viên nào đủ 18 tuổi trở lên hoặc thành viên còn lại không còn sức lực hành vi dân sự đầy đủ theo quy định của pháp luật.  Trân trọng! |
| Xếp hạng tốt nghiệp của sinh viên cao đẳng mầm non được xác định như thế nào? | **Khoản 3 Điều 14 Quy chế đào tạo trình độ cao đẳng ngành Giáo dục Mầm non** ban hành kèm theo **Thông tư 07/2023/TT-BGDĐT** quy định về việc xếp hạng tốt nghiệp của sinh viên cao đẳng mầm non như sau:  Công nhận tốt nghiệp và cấp bằng tốt nghiệp  ...  3. Xếp hạng tốt nghiệp được xác định căn cứ vào điểm trung bình tích lũy toàn khóa được quy định tại khoản 5 Điều 10 của Quy chế này, trong đó, hạng tốt nghiệp của sinh viên có điểm trung bình tích lũy loại xuất sắc và giỏi sẽ bị giảm đi một mức nếu thuộc một trong các trường hợp sau:  a) Khối lượng của các học phần phải học lại vượt quá 5% so với tổng số tín chỉ quy định cho toàn chương trình;  b) Sinh viên đã bị kỷ luật từ mức cảnh cáo trở lên trong thời gian học.  ... | Như vậy, đối với sinh viên cao đẳng mầm non thì bạn có thể xếp hạng tốt nghiệp được xác định căn cứ vào điểm trung bình tích lũy toàn khóa được quy định tại khoản 5 Điều 10 của Quy chế này.  Trân trọng! |
| Điều kiện lưu trữ hồ sơ bệnh án điện tử thay cho hồ sơ bệnh án giấy? | Theo **Điều 6 Thông tư 46/2018/TT-BYT** quy định lưu trữ hồ sơ bệnh án điện tử như sau:  Cơ sở khám bệnh, chữa bệnh được phép lưu trữ hồ sơ bệnh án điện tử thay cho hồ sơ bệnh án giấy khi đáp ứng các yêu cầu sau đây:  - Phần mềm hồ sơ bệnh án điện tử đạt mức nâng cao theo quy định tại Thông tư số 54/2017/TT-BYT ngày 29/12/2017 của Bộ trưởng Bộ Y tế ban hành bộ tiêu chí ứng dụng công nghệ thông tin tại các cơ sở khám bệnh, chữa bệnh.  - Thiết bị lưu trữ phải có đủ dung lượng để lưu trữ hồ sơ bệnh án điện tử đáp ứng thời gian lưu trữ hồ sơ bệnh án theo quy định tại khoản 3 Điều 59 Luật Khám bệnh, chữa bệnh.  - Hồ sơ bệnh án điện tử phải được lưu trữ dự phòng tại một cơ quan, tổ chức cung cấp dịch vụ lưu trữ dữ liệu (data center) đáp ứng tiêu chuẩn theo quy định của Bộ trưởng Bộ Thông tin và Truyền thông. Đối với các cơ sở khám bệnh, … <Còn tiếp> | Như vậy, điều kiện lưu trữ hồ sơ bệnh án điện tử thay cho hồ sơ bệnh án giấy được quy định trong quy định của Bộ Quốc phòng. |

Bảng 4: Một số câu trả lời được tạo ra khi tài liệu trích dẫn được cung cấp chính xác

|  |  |  |
| --- | --- | --- |
| Câu hỏi | Tài liệu trích dẫn (từ BM25 Retriever, top3 tài liệu) | Câu trả lời và nhận xét |
| Sinh viên ngành sư phạn sẽ được hỗ trợ chi phí sinh hoạt bao nhiêu? | Tại **Điều 8 Nghị định 116/2020/NĐ-CP** quy định về chi phí bồi hoàn và cách tính chi phí bồi hoàn như sau:  Chi phí bồi hoàn và cách tính chi phí bồi hoàn  1. Chi phí bồi hoàn bao gồm kinh phí hỗ trợ tiền đóng học phí và chi phí sinh hoạt đã được ngân sách nhà nước hỗ trợ cho người học.  2. Sinh viên sư phạm … <Còn tiếp>  Tại **Điều 8 Nghị định 116/2020/NĐ-CP** quy định về Chi phí bồi hoàn và cách tính chi phí bồi hoàn như sau:  1. Chi phí bồi hoàn bao gồm kinh phí hỗ trợ tiền đóng học phí và chi phí sinh hoạt đã được ngân sách nhà nước hỗ trợ cho người học.  2. Sinh viên sư phạm … <Còn tiếp>  Căn cứ **Điều 8 Nghị định 116/2020/NĐ-CP** quy định về chi phí bồi hoàn và cách tính chi phí bồi hoàn như sau:  1. Chi phí bồi hoàn bao gồm kinh phí hỗ trợ tiền đóng học phí và chi phí sinh hoạt đã được ngân sách nhà nước hỗ trợ cho người học.  2. Sinh viên sư phạm … <Còn tiếp> | Như vậy, đối với sinh viên ngành sư phạm thuộc đối tượng quy định tại điểm a và điểm c khoản 1 Điều 6 của Nghị định này phải bồi hoàn toàn bộ kinh phí đã được ngân sách nhà nước hỗ trợ.  Trân trọng!  Nhận xét: Một số tài liệu trích dẫn có thể bị trùng và không liên quan dẫn đến câu trả lời thiếu chất lượng |
| Sinh viên chỉ bài cho bạn có bị thông báo về gia đình hay không? | Căn cứ **Điều 9 Nghị định trên quy định về thu hồi chi phí bồi hoàn** như sau:  1. Hằng năm, căn cứ vào kết quả rèn luyện, học tập của sinh viên sư phạm, cơ sở đào tạo giáo viên thông báo danh sách sinh viên thuộc đối tượng quy định tại điểm c khoản 1 Điều 6 Nghị định này cho Ủy ban nhân dân cấp tỉnh để thông báo thu hồi kinh phí đã hỗ trợ cho sinh viên sư phạm và gia đình theo quy định tại khoản 2 Điều 8 Nghị định này…<Còn tiếp>  Căn cứ **Khoản 3 Điều 9 Nghị định 116/2020/NĐ-CP** quy định thu hồi chi phí bồi hoàn như sau:  3. Trong thời hạn 30 ngày, kể từ ngày nhận được quyết định của cơ quan nhà nước có thẩm quyền, sinh viên hoặc gia đình phải có trách nhiệm liên hệ với cơ quan thu hồi kinh phí bồi hoàn để làm thủ tục bồi hoàn… <Còn tiếp>  Tại Điều 3 Quyết định 157/2007/QĐ-TTg quy định về tín dụng đối với học sinh, sinh viên do Thủ tướng Chính phủ ban hành, có quy định:  Điều 3. Phương thức cho vay:  - Việc cho vay đối với học sinh, sinh viên được thực hiện theo phương thức cho vay thông qua hộ gia đình. Đại diện hộ gia đình là người trực tiếp vay vốn và có trách nhiệm trả nợ Ngân hàng Chính sách xã hội. Trường hợp học sinh, sinh viên mồ côi cả cha lẫn mẹ hoặc chỉ mồ côi cha hoặc mẹ nhưng người còn lại không có khả năng lao động, được trực tiếp vay vốn tại Ngân hàng Chính sách xã hội nơi nhà trường đóng trụ sở…<Còn tiếp> | Như vậy, đối với sinh viên sư phạm thuộc đối tượng phải bồi hoàn kinh phí theo quy định tại khoản 1 Điều 6 Nghị định này, nếu thuộc đối tượng chính sách, khó khăn thì bạn có thể thuộc đối tượng xét tuyển thẳng và được xét tuyển thẳng vào học trước khi kết thúc thời hạn đóng học.  Trân trọng!  Nhận xét: Tài liệu trích dẫn hoàn toàn sai nên câu trả lời không khớp với câu hỏi |
| Đang là sinh viên năm nhất có được chuyển trường đại học? | Tại **Điều 18 Quy chế đào tạo đại học và cao đẳng hệ chính quy** theo hệ thống tín chỉ ban hành kèm theo **Quyết định 43/2007/QĐ-BGDĐT** như sau:  1. Sinh viên được xét chuyển trường nếu có các điều kiện sau đây:  a) Trong thời gian học tập, nếu gia đình chuyển nơi cư trú hoặc sinh viên có hoàn cảnh khó khăn, cần thiết phải chuyển đến trường gần nơi cư trú của gia đình để thuận lợi trong học tập;  b) Xin chuyển đến trường có cùng ngành hoặc thuộc cùng nhóm ngành với ngành đào tạo mà sinh viên đang học;  c) Được sự đồng ý của Hiệu trưởng trường xin chuyển đi và trường xin chuyển đến;  d) Không thuộc một trong các trường hợp không được phép chuyển trường quy định tại khoản 2 Điều này.  2. Sinh viên không được phép chuyển trường trong các trường hợp sau:  a) Sinh viên đã tham dự kỳ thi tuyển sinh theo đề thi chung, nhưng không trúng tuyển vào trường hoặc có kết quả thi thấp hơn điểm trúng tuyển của trường xin chuyển đến;  b) Sinh viên thuộc diện nằm ngoài vùng tuyển quy định của trường xin chuyển đến;  c) Sinh viên năm thứ nhất và năm cuối khóa;  d) Sinh viên đang trong thời gian bị kỷ luật từ cảnh cáo trở lên.  3. Thủ tục chuyển trường:… <Còn tiếp>  Theo như bạn trình bày là hiện tại ban đang học năm nhất do học phí cao quá nên bạn muốn nghỉ học và năm sau muốn chuyển. Trường hợp này bạn vẫn có thể chuyển trường nếu bạn có đủ các điều kiện theo quy định trên.  Khi bạn chuyển trường thì bạn cần sự đồng ý của hiệu trưởng trường xin chuyển đi và trường chuyển đến. Trường hợp này bạn cần học hết năm nhất và kết quả trong kỳ thi tuyển sinh theo đề thi chung phải có kết quả cao hơn của ngành học tại trường dự định chuyển đến.  Trường hợp bạn nghỉ học tại trường đang học hiện tại thì bạn không thể lấy kết quả trong kỳ thi chung năm 2018 để nộp vào trường khác được. Để học trường mới thì bạn cần phải tham dự kỳ thi chung năm 2019 để được xét tuyển.  **Điều 9 Quy chế đào tạo Đại học và Cao đẳng hệ chính quy** ban hành kèm theo **Quyết định số 25/2006/QĐ-BGDĐT** có quy định:  Sinh viên được xét chuyển trường nếu có các điều kiện sau đây:  - Trong thời gian học tập, nếu gia đình chuyển nơi cư trú hoặc sinh viên có hoàn cảnh khó khăn, cần thiết phải chuyển đến trường gần nơi cư trú của gia đình để thuận lợi trong học tập;  - Xin chuyển đến trường có ngành đào tạo trùng hoặc thuộc cùng nhóm ngành với ngành đào tạo mà sinh viên đang học;  - Được sự đồng ý của Hiệu trưởng trường xin chuyển đi và trường xin chuyển đến.  **Khoản 2 Điều 9 Quy chế đào tạo đại học và cao đẳng chính quy** ban hành kèm theo **Quyết định 25/2006/QĐ-BGDĐT** quy định về việc chuyển trường đại học như sau:  Sinh viên không được phép chuyển trường trong các trường hợp sau:  - Sinh viên đã dự thi tuyển sinh nhưng không trúng tuyển hoặc có điểm thi thấp hơn điểm xét tuyển của trường xin chuyển đến trong trường hợp chung đề thi tuyển sinh;  - Sinh viên có hộ khẩu thường trú nằm ngoài vùng tuyển của trường xin chuyển đến;  - Sinh viên năm thứ nhất và năm cuối khóa;  - Sinh viên đang chịu mức kỷ luật từ cảnh cáo trở lên. | Như vậy, đối với sinh viên không được chuyển trường trong các trường hợp sau:  - Sinh viên đã tham dự kỳ thi tuyển sinh theo đề thi chung, nhưng không trúng tuyển vào trường hoặc có kết quả thi thấp hơn điểm trúng tuyển của trường xin chuyển đến;  - Sinh viên thuộc diện nằm ngoài vùng tuyển quy định của trường xin chuyển đến;  - Sinh viên năm thứ nhất và năm cuối khóa;  - Sinh viên đang trong thời gian bị kỷ luật từ cảnh cáo trở lên.  Nhận xét: cả tài liệu trích dẫn và câu trả lời đều tốt mặc dù tài liệu trích dẫn chưa được xử lý kỹ càng (có thông tin dư thừa được tô đỏ) |

Bảng 5: Một số câu trả lời được sinh ra dựa vào các tài liệu trích dẫn từ Retriever

# V. Kết luận:

Đây là một trong những dự án lớn nhất mà chúng em từng làm, nhờ đó chúng em học được rất nhiều thứ, từ cào và xử lý dữ liệu, kiến trúc Llama2, cách sử dụng các model từ Huggingface, các huấn luyện và lưu trữ LoRA, … Tuy nhiên bộ dữ liệu vẫn chưa được xử lý tốt và Retriever khá đơn giản nên độ chính xác về trích xuất tài liệu chưa cao. Bộ dữ liệu cũng chưa đủ lớn, mô hình vẫn chưa được huấn luyện lâu nên vẫn còn nhiều việc để hoàn thành sản phẩm này.