Journal of Industrial Information Technology and Application *ISSN(Online): 2586-0852*
Vol. 9 No. 1

LangChain and RAG-Based Q&A System for University Policies

In-Hye Park [1), Min-Jeong Kim [2), Kyung-Ae Cha [3)*

[1,2,3) Dept. of Artificial Intelligence, Daegu University, Gyeongsan-si, Korea

Abstract. Recent advancements in language models have led to their widespread application across various fields, achieving remarkable success. However, these models often exhibit a phenomenon known as 'hallucination', where they generate responses that present false information as factual. This issue has emerged as a critical challenge, particularly in scenarios where reliability and accuracy are paramount. To address this problem, the RAG (Retrieval-Augmented Generation) technique has gained significant attention. RAG combines generative language models with information retrieval systems, allowing the model to search for relevant information from external databases before generating a response. This approach helps mitigate hallucinations and delivers more trustworthy and accurate information.

This study leverages the LangChain framework to develop a RAG-based Q&A system using university policies and administrative regulations as the dataset, which serves as the foundation for students' academic and institutional operations. By systematically organizing and managing complex data, the proposed system aims to provide accurate and reliable responses, offering users valuable and actionable information efficiently.

Keywords; Large Language Models (LLMs), Retrieval-Augmented Generation, LangChain

Cite this paper as: *In-Hye Park, Min-Jeong Kim, Kyung-Ae Cha* (2025) "LangChain and RAG-Based Q&A System for University Policies", Journal of Industrial Information Technolgoy and Application, Vol. 9. No.1, pp.1039 - 1050

1. Introduction

The recent advancements in Large Language Models (LLMs) have introduced new possibilities in the field of Natural Language Processing (NLP), establishing themselves as innovative tools across various domains.[1] Notably, state-of-the-art models such as OpenAI's GPT-4 leverage vast training datasets to provide advanced capabilities in language understanding and generation. These models have been widely applied in areas such as conversational AI, content creation, translation, and question-answering systems [2, 3].

However, large-scale language models are often limited by the phenomenon of hallucination.[4] Hallucination refers to the generation of nonexistent or inaccurate information by the model, which undermines the reliability of its responses. This issue has become particularly critical in fields where trustworthy information is essential.

To address this challenge, the Retrieval-Augmented Generation (RAG) technique has gained significant attention in recent years.[5] RAG combines generative language models with information retrieval systems, allowing the model to search external databases for relevant information before generating responses. This approach improves the reliability of model outputs while ensuring that responses reflect the most up-to-date information.

In this study, the RAG technique is employed to design and implement a question-answering system based on university regulations and administrative information.

2. Related research content

LangChain is an open-source framework designed to facilitate the use of large language models (LLMs), and various studies have been proposed that combine its information retrieval and generation capabilities. Choi Jongmyung et al. (2024) developed a KTAS(Korean Triage and Acuity Scale) determination system using LangChain and GPT-4[6]. This study demonstrated an effective use case of Retrieval-Augmented Generation (RAG) technology by integrating voice and text data to automatically classify patient severity in emergency medical settings. LangChain was utilized to search external data and integrate with GPT-4, delivering rapid and reliable emergency care guidelines.

So Hoon et al. (2024) conducted a study to improve the interaction with NPCs in the game system by combining LangChain and Unreal Engine.[7] The system was designed to provide real-time, contextual responses to user questions, and the combination of the vector database and LangChain enabled NPCs to have more natural and personalized conversations.

Additionally, Namhyun Kim et al. (2024) developed a system to process and cluster tourism research data using LangChain[8]. This study clustered semantically similar data to analyze research trends and explore new possibilities in the tourism industry. LangChain’s search and summarization capabilities contributed to effectively managing large datasets and improving the quality of the analysis.

These studies have shown that LangChain and RAG technologies can be applied in each specific domain to provide appropriate responses to domain-specific requirements. However, university school regulations and administrative information have different structures and contents for each school, and user needs are also diverse, so research that reflects this is still needed.

Currently, Daegu University does not provide a chatbot service to search for school regulations and administrative information, and this information is provided in the form of several documents through the administrative service on the Daegu University website. Each document contains different regulations and detailed information, making it challenging for users to quickly find the information they need. To solve this problem, this study utilized LangChain-based RAG technology and implemented it as an app-based interface so that users can search for information that is most relevant to school regulations and administration-related questions and quickly check the necessary information. Through this, we would like to propose a system that increases the efficiency of information access and reduces the time and effort required to search for information.

3. System design and configuration

Figure 1 is a diagram illustrating the overall process of the RAG -based Q&A system, which is designed to manage Daegu University's school regulations and administrative information and provide reliable responses. The system includes the following steps from data processing to user response generation.

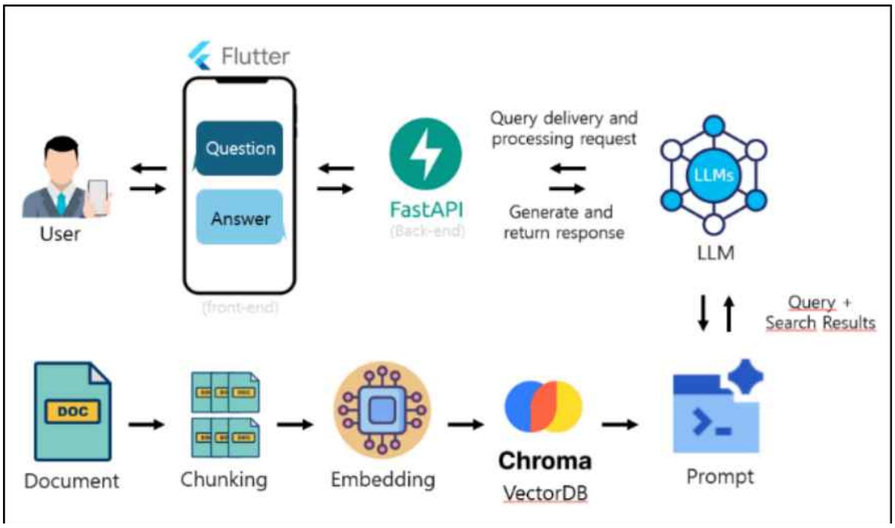
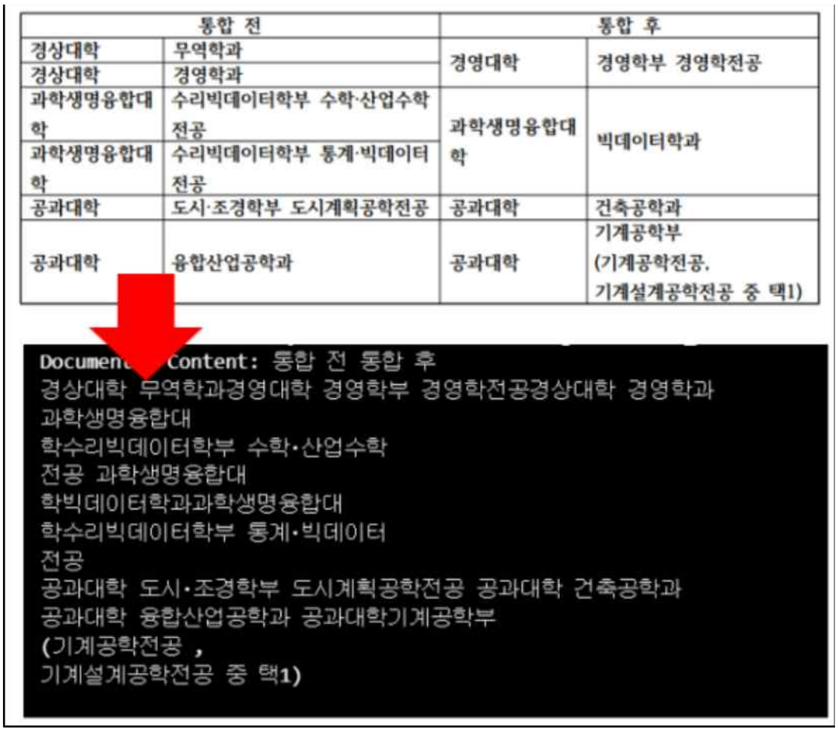


Figure 1 RAG-based Q&A system structure diagram

3.1 Data collection and preprocessing

Daegu University Academic Regulations and Administrative Data is available in the Regulations section of the Daegu University website and contains various documents related to university administration and academic operations, such as “Tuition Regulations,” “Academic Grade Evaluation Regulations,” “Leave of Absence and Reinstatement Procedures,” etc. The materials are primarily provided in Hangul document format (HWP) and consist of a mixture of tabular data and unstructured text. In order for a large-scale language model (LLM) to process it efficiently, preprocessing to reconstruct and clean the data was essential.

For example, in the Daegu University school regulations document, tabular data including "Transitional measures following college integration and department (major) name change" contained clear relationships between columns, such as "Before integration" and "After integration.". However, when such table data was simply converted to text, the relationships between columns and structural context were lost, making it difficult to understand the meaning of the information. Figure 2 shows this original table data. For example, if a relationship indicating that the "Urban Planning Engineering major in the Department of Urban Planning and Landscape" was changed to "Architecture Department" was converted to a text simply listed, the relationship between the columns was not clearly revealed, and it was possible that the language model would not understand it correctly.

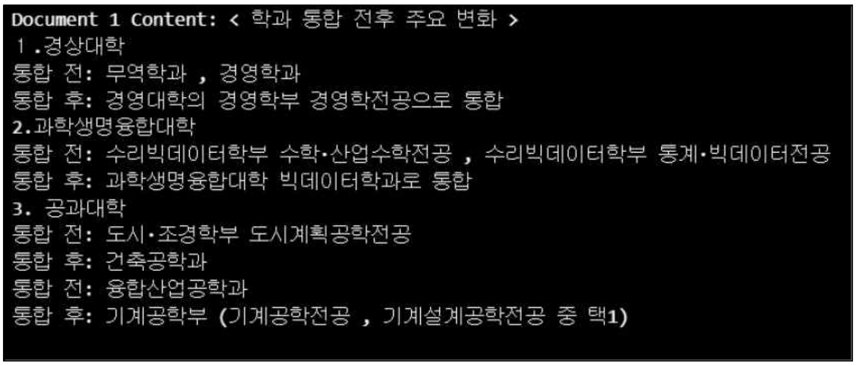


통합 전		통합 후	
경상대학	무역학과	경영대학	경영학부 경영학전공
경상대학	경영학과		
과학생명융합대학	수리빅데이터학부 수학·산업수학전공	과학생명융합대학	빅데이터학과
과학생명융합대학	수리빅데이터학부 통계·빅데이터전공		
공과대학	도시·조경학부 도시계획공학전공	공과대학	건축공학과
공과대학	융합산업공학과	공과대학	기계공학부 (기계공학전공, 기계설계공학전공 중 택1)

```
Documen[illegible] Content: 통합 전 통합 후
경상대학 무역학과경영대학 경영학부 경영학전공경상대학 경영학과
과학생명융합대
학수리빅데이터학부 수학·산업수학
전공 과학생명융합대
학빅데이터학과과학생명융합대
학수리빅데이터학부 통계·빅데이터
전공
공과대학 도시·조경학부 도시계획공학전공 공과대학 건축공학과
공과대학 융합산업공학과 공과대학기계공학부
(기계공학전공 ,
기계설계공학전공 중 택1)
```

Figure 2 Some output in a structured format, such as a table

To solve this problem, the table data was manually reconstructed and converted into a form that maintains the relationship and context between columns. Figure 3 shows the reconstructed text data and is designed to maintain the relationship between "before integration" and "after integration". This work was an important step in preserving the consistency of the data and helping the language model to handle it. In addition, in addition to the table format data, the unstructured text of school regulations and administrative data was classified based on the title, main item, and keyword of the regulation to increase the relevance. Unnecessary information was removed, and the structure and meaning of the document were succinctly organized.



```
Document 1 Content: < 학과 통합 전후 주요 변화 >
1 .경상대학
통합 전: 무역학과 , 경영학과
통합 후: 경영대학의 경영학부 경영학전공으로 통합
2.과학생명융합대학
통합 전: 수리빅데이터학부 수학·산업수학전공 , 수리빅데이터학부 통계·빅데이터전공
통합 후: 과학생명융합대학 빅데이터학과로 통합
3. 공과대학
통합 전: 도시·조경학부 도시계획공학전공
통합 후: 건축공학과
통합 전: 융합산업공학과
통합 후: 기계공학부 (기계공학전공 , 기계설계공학전공 중 택1)
```

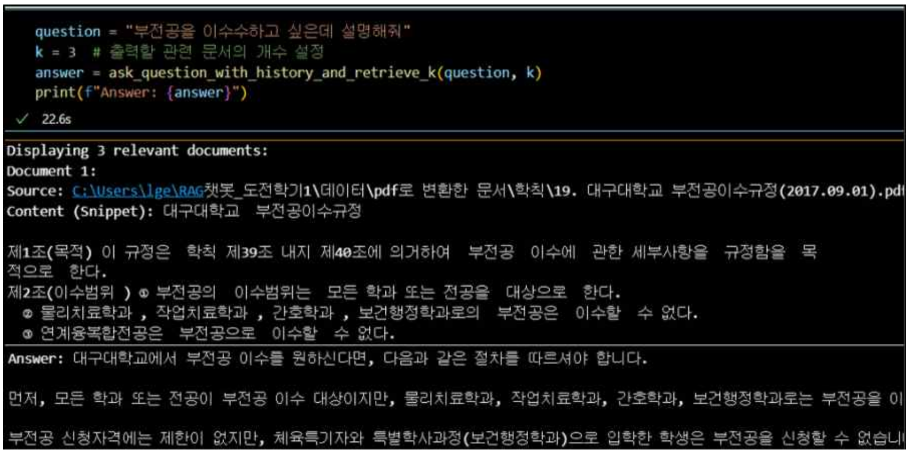
Figure 3 Convert structured data to text format

Then, the organized data was divided into approximately 1000 character units using RecursiveCharacterTextSplitter. Each text chunk maintained the continuity of the context by adding a superposition of 300 characters. This allowed the language model to better understand the back-and-forth relationship of the text and to generate responses suitable for user queries.

3.2 Build embedding and search algorithms

The preprocessed Daegu University academic regulations and administrative data were converted into high-dimensional vectors using OpenAI's GPT-4 embedding model, and the converted data was stored in a vector database called ChromaDB. ChromaDB is a database designed to search for the most appropriate data by calculating the similarity between user queries and data.

During the search process, the MMR (Maximal Marginal Relevance) algorithm was applied to maintain relevance to user queries while minimizing redundant information, thereby producing highly reliable search results. For example, as shown in Figure 4, when the query "tuition fee" is entered, the MMR algorithm searches for the most relevant information within the Daegu University academic policy data, such as "tuition payment standards," "international student tuition regulations," and "installment payment procedures." Returns information as search results. Through this search process, users can clearly understand key information related to the scope of minor completion and quickly check the school policy information most appropriate for your inquiry.



```
question = "부전공을 이수수하고 싶은데 설명해줘"
k = 3  # 출력할 관련 문서의 개수 설정
answer = ask_question_with_history_and_retrieve_k(question, k)
print(f"Answer: {answer}")
```

```
✓ 22.6s
```

```
Displaying 3 relevant documents:
Document 1:
Source: C:\Users\lge\RAG챗봇_도전학기1\데이터\pdf로 변환한 문서\학칙\19. 대구대학교 부전공이수규정(2017.09.01).pdf
Content (Snippet): 대구대학교  부전공이수규정

제1조(목적) 이 규정은  학칙 제39조 내지 제40조에 의거하여  부전공  이수에  관한 세부사항을  규정함을  목
적으로  한다.
제2조(이수범위 ) ① 부전공의  이수범위는  모든 학과 또는 전공을  대상으로  한다.
  ② 물리치료학과 , 작업치료학과 , 간호학과 , 보건행정학과로의  부전공은  이수할  수 없다.
  ③ 연계융복합전공은  부전공으로  이수할  수 없다.
```

```
Answer: 대구대학교에서 부전공 이수를 원하신다면, 다음과 같은 절차를 따르셔야 합니다.

먼저, 모든 학과 또는 전공이 부전공 이수 대상이지만, 물리치료학과, 작업치료학과, 간호학과, 보건행정학과로는 부전공을 이

부전공 신청자격에는 제한이 없지만, 체육특기자와 특별학사과정(보건행정학과)으로 입학한 학생은 부전공을 신청할 수 없습니
```

Figure 4 Example of MMR Similarity Based Search

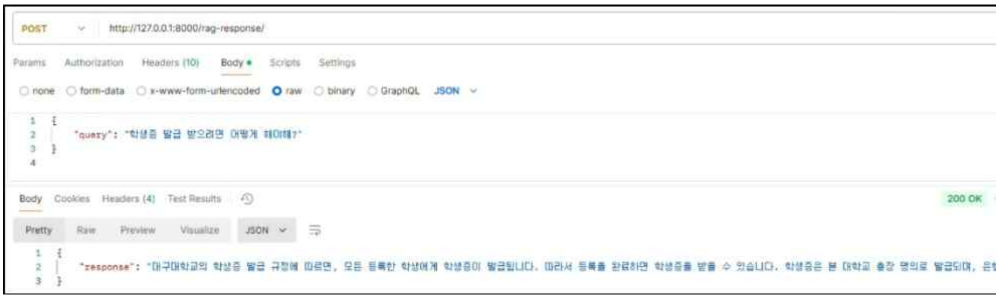
Through this search process, we go beyond simply listing data and selectively provide key information most relevant to user queries, allowing users to search for data that suits their needs.

3.3 Server integration and application implementation

The built system is designed in a structure that interfaces FastAPI and ChromaDB to process user questions, and the GPT-4 model generates the final response. FastAPI is a lightweight web framework that provides high performance and fast request handling through support for asynchronous and concurrent processing [7]. This study leveraged these advantages of FastAPI to design a query-answering system for Daegu University's academic regulations.

In particular, we optimized the system's scale-up performance by using Hypercorn, an Application Server Gateway Interface (ASGI) server.[8] Hypercorn is an asynchronous ASGI server that supports HTTP/1.1, HTTP/2, and WebSockets, enabling fast data transfer and efficient handling of multiple requests simultaneously. Through this, we built a system that processes large-scale school policy data and provides real-time responses.

Figure 5 shows a question-answering test performed using Postman, an API testing tool, to verify the FastAPI question-answering system.[9] When a user enters the question “What do I need to do to get a student ID card?”, the question is sent to the server through an HTTP request. The request data is in JSON format, and after processing it, the server searches ChromaDB for documents related to student ID processing and generates and returns a response through the GPT-4 model. As a result of the test through Postman, it was confirmed that the FastAPI server returned an accurate response to the input student ID issuance method.



```
{
  "query": "학생증 발급 받으려면 어떻게 해야해?"
}
```

```
{
  "response": "대구대학교의 학생증 발급 규정에 따르면, 모든 등록한 학생에게 학생증이 발급됩니다. 따라서 등록을 완료하면 학생증을 받을 수 있습니다. 학생증은 본 대학교 총장 명의로 발급되며, 은
}
```

Figure 5 API testing using Postman

A Flutter-based mobile application was developed for the front-end implementation.[10] The application is designed to provide an intuitive user interface (UI) so that users can effectively utilize the question-and-answer system related to school rules.

In Figure 6, the application consists of a splash screen and a question and answer screen. The splash screen is the first thing that appears when the application is launched, and it conveys the purpose and identity of the service to the user through the Daegu University logo and the chatbot title. The question and answer screen displays the title “Academic Rules Response System” at the top, and a question input field and a send button are placed at the bottom. The input field defaults to the prompt “How can I help you?” to prompt the user to type in a question, which can be sent via the Send button (arrow icon) in the bottom right corner.

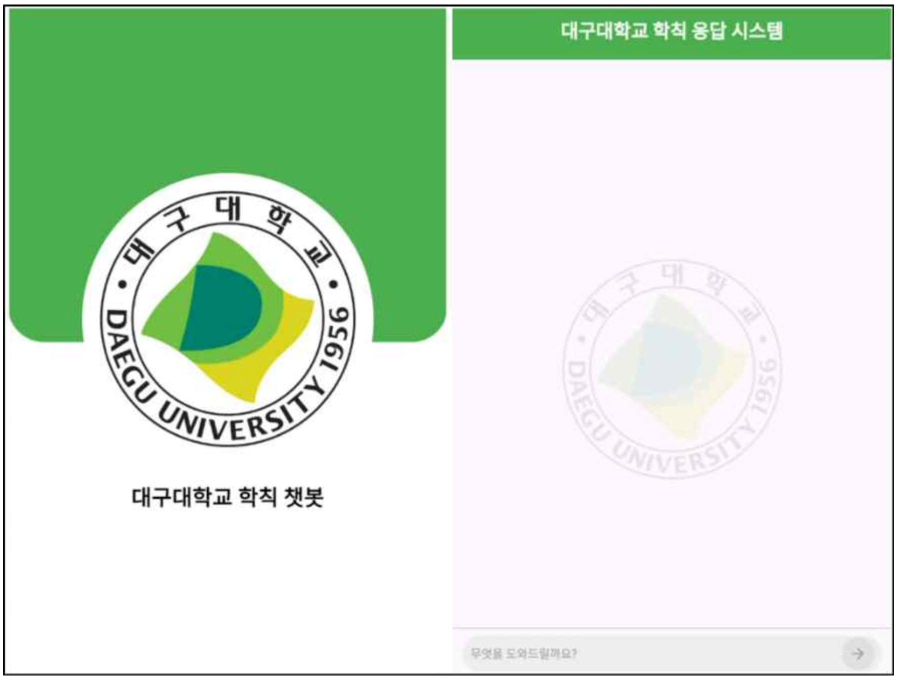


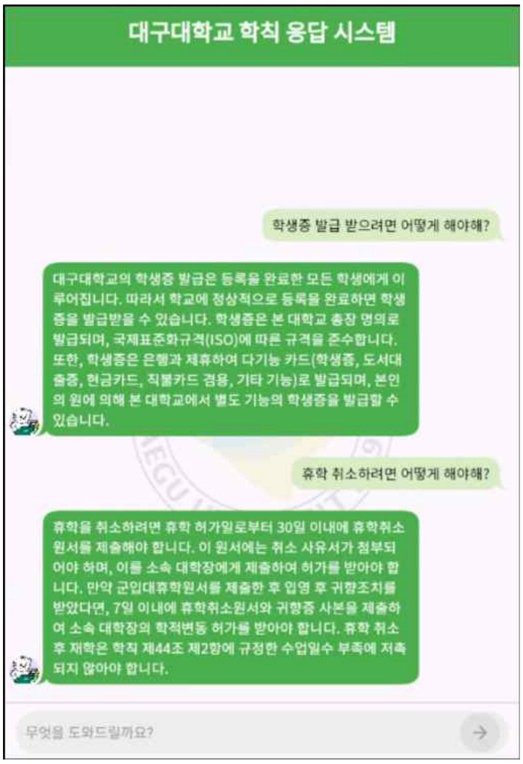
Figure 6 Application implemented in Flutter

4. Experiment results

The figures below show examples of responses provided by the system to various questions entered by the user in the Flutter-based application.

Figure 7 shows how the system searches for relevant information from Daegu University's school regulations data and guides the documents necessary for issuing student cards and application procedures when the user enters the question, "How do I issue a student card?". In addition, in response to the question, "How do I cancel my leave of absence?" the system explained the procedure for applying for cancellation of leave of absence and precautions step by step. This allows users to easily understand and execute the necessary administrative procedures.

In Figure 8, for the question “What are the criteria for issuing an employment recommendation letter?”, the system retrieved the relevant academic policy data to show the grade point average requirement (3.5 or higher) and the application process for obtaining an employment recommendation letter. Meanwhile, Figure 9 illustrates how the system responds to queries that are not included in the dataset, informing users that the requested information could not be found.



대구대학교 학칙 응답 시스템

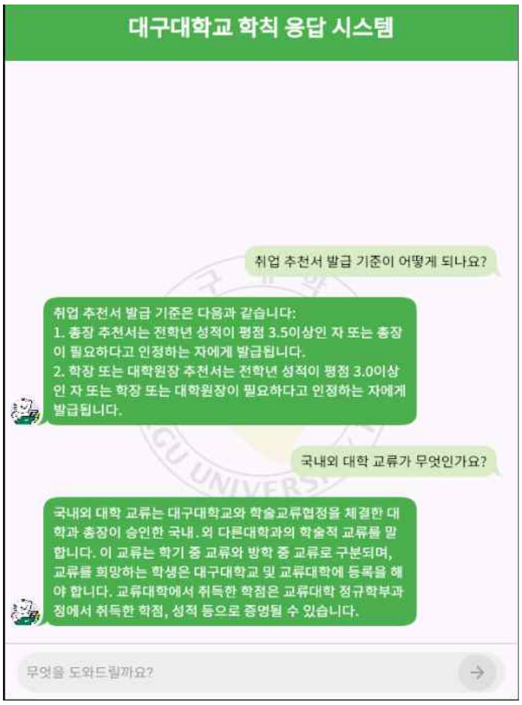
학생증 발급 받으려면 어떻게 해야해?

대구대학교의 학생증 발급은 등록을 완료한 모든 학생에게 이루어집니다. 따라서 학교에 정상적으로 등록을 완료하면 학생증을 발급받을 수 있습니다. 학생증은 본 대학교 총장 명의로 발급되며, 국제표준화규격(ISO)에 따른 규격을 준수합니다. 또한, 학생증은 은행과 제휴하여 다기능 카드(학생증, 도서대출증, 현금카드, 직불카드 겸용, 기타 기능)로 발급되며, 본인의 원에 의해 본 대학교에서 별도 기능의 학생증을 발급할 수 있습니다.

휴학 취소하려면 어떻게 해야해?

휴학을 취소하려면 휴학 허가일로부터 30일 이내에 휴학취소원서를 제출해야 합니다. 이 원서에는 취소 사유서가 첨부되어야 하며, 이를 소속 대학장에게 제출하여 허가를 받아야 합니다. 만약 군입대휴학원서를 제출한 후 입영 후 귀향조치를 받았다면, 7일 이내에 휴학취소원서와 귀향증 사본을 제출하여 소속 대학장의 학적변동 허가를 받아야 합니다. 휴학 취소 후 재학은 학칙 제44조 제2항에 규정한 수업일수 부족에 저촉되지 않아야 합니다.

Figure 7 Q&A with Implemented Applications (1)



대구대학교 학칙 응답 시스템

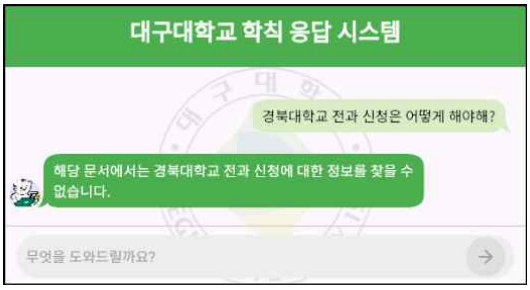
취업 추천서 발급 기준이 어떻게 되나요?

취업 추천서 발급 기준은 다음과 같습니다:
1. 총장 추천서는 전학년 성적이 평점 3.5이상인 자 또는 총장이 필요하다고 인정하는 자에게 발급됩니다.
2. 학장 또는 대학원장 추천서는 전학년 성적이 평점 3.0이상인 자 또는 학장 또는 대학원장이 필요하다고 인정하는 자에게 발급됩니다.

국내외 대학 교류가 무엇인가요?

국내외 대학 교류는 대구대학교와 학술교류협정을 체결한 대학과 총장이 승인한 국내.외 다른대학과의 학술적 교류를 말합니다. 이 교류는 학기 중 교류와 방학 중 교류로 구분되며, 교류를 희망하는 학생은 대구대학교 및 교류대학에 등록을 해야 합니다. 교류대학에서 취득한 학점은 교류대학 정규학부과정에서 취득한 학점, 성적 등으로 증명될 수 있습니다.

Figure 8 Q&A with Implemented Applications (2)



대구대학교 학칙 응답 시스템

경북대학교 전과 신청은 어떻게 해야해?

해당 문서에서는 경북대학교 전과 신청에 대한 정보를 찾을 수 없습니다.

Figure 9 Responding to questions that do not exist in the dataset

5. Conclusion

In this study, we developed a system that provides academic and administrative information of Daegu University in the form of questions and answers by utilizing the RAG(Retrieval-Augmented Generation) technique. Users were able to obtain information more quickly about school regulations-related questions such as tuition regulations and graduation requirements, which greatly improved the efficiency of accessing academic and administrative information.

However, since structured data such as tabular format must be preprocessed manually so that LLM can accurately process it, there is a limitation in that additional time and cost are required. To address these issues, future research will improve the system's processing efficiency by applying technologies that can integrally process images and text using multimodal data.

References

[1] Lanchain, https://python.langchain.com/v0.1/docs/modules/model_io/llms/. Accessed: December 12, 2024.

[2] OpenAI, "GPT-4 Technical Report," Available: https://openai.com/research/gpt-4, arXiv,2303.08774 (2024).

[3] OpenAI, https://openai.com/. Accessed: December 12, 2024.

[4] L. Huang, W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin, and T. Liu, "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions," arXiv, vol. 2311.05232, pp. 1-2, Nov. 2023.

[5] Shangyu Wu, Ying Xiong, Yufei Cui, Haolun Wu, Can Chen, Ye Yuan, Lianming Huang, Xue Liu, Tei-Wei Kuo, Nan Guan, Chun Jason Xue, "Retrieval-Augmented Generation for Natural Language Processing: A Survey," arXiv, 2024.

[6] Choi Jongmyung, Lee Youngho, and Kim Sun Kyung, "Design of a KTAS Decision System using LangChain," (Jour. of KoCon.a) The Journal of the Korea Contents Association, pp. 9-10, 2024.

[7] Hoon So, Seongjun Kang, Chanwoo Jeong, and Hyunjun Jung, "Unreal Engine"s specific NPC chatbot system using Langchains," Proceedings of KIIT Conference, pp. 394-398, 2024.

[8] Namhyun Kim, Sehyoung Kim, and Juyoung Kang, "Analysis of National Research and Development Trends Using LangChain, and Clustering of Research Outcomes: Focused on the NTIS Tourism Industry," The Journal of Society for e-Business Studies, Vol. 29, No. 2, pp. 93-115, 2024, DOI: 10.7838/jsebs.2024.29.2.093

[9] ChromaDB, https://www.trychroma.com/. Accessed: December 12, 2024.

[10] FastAPI, https://fastapi.tiangolo.com/ko/. Accessed: December 12, 2024.

[11] Hypercorn, https://hypercorn.readthedocs.io/en/latest/. Accessed: December 14, 2024.

[12] Postman, https://www.postman.com/. Accessed: December 14, 2024.

[13] Flutter, https://flutter.dev/. Accessed: December 12, 2024.

COPYRIGHT TRANSFER FORM

Title of the Journal: International Society for Information Technology and Application

Title of Work: LangChain and RAG-Based Q&A System for University Policies

Author(s) name(s): In-Hye Park, Min-Jeong Kim, Kyung-Ae Cha

Corresponding Author's name, affiliation, Country and e-mail: Kyung-Ae Cha,Department of Artificial Intelligence, Daegu University, Republic of Korea,E-mail:chaka@daegu.ac.kr

The copyright to the Contribution identified above is transferred to International Society for Information Technology and Application (hereinafter called ISIITA). The copyright transfer covers the sole right to print, publish, distribute and sell throughout the world the said Contribution and parts thereof, including all revisions or versions and future editions, in all forms and media of expression – such as in its electronic form (offline, online) – now known or developed in the future, as well as to translate, print, publish, distribute and sell the Contribution in any foreign languages and throughout the world.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release.

I hereby irrevocably transfer exclusive copyright for this material to ISIITA. I give ISIITA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give ISIITA permission to distribute the materials without restriction to any institutions or individuals. I give ISIITA permission to submit the work for inclusion in article repositories as ISIITA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that ISIITA applies to the published article. ISIITA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. ISIITA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print. ISIITA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that ISIITA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to ISIITA, ISIITA's contractors, and ISIITA's partners to further distribute the work.
The Corresponding Author signs for and accepts responsibility for releasing this material on behalf of any and all Co-Authors.

Signature(s) of Author(s) : 박인혜 Date : January 20

AUTHOR/COMPANY RIGHTS:
If you are employed and you prepared your paper as part of your job, the rights to your work rest initially with your employer. In that case, when you sign the copyright transfer form, we assume you are authorized to do so by your employer and that your employer has consented to all the terms and conditions of this form. If not, it should be signed by someone so authorized.

PLEASE DIRECT ALL QUESTIONS ABOUT ISIITA COPYRIGHT POLICY OR THIS FORM TO:
Conference secretarial / International Symposium on Innovation in Information, E-mail: isiita.paper@gmail.com Office