

RAG 기반 대학 규정 검색 챗봇 개발 및 성능 평가: 경희대학교 사례 연구

장현중¹, 김영훈²

¹경희대학교 인공지능학과 석사과정

²경희대학교 산업경영공학과 교수

hyunjongjang@khu.ac.kr, y.kim@khu.ac.kr

Development and Performance Evaluation of a RAG-based University Regulation Search Chatbot: A Case Study of Kyung Hee University

Hyunjong Jang¹, Younghoon Kim²

¹Dept. of Artificial Intelligence, Kyunghee University

²Dept. of Industrial and Management Engineering, Kyunghee University

요약

본 연구는 Large Language Model (LLM)과 Retrieval Augmented Generation (RAG) 기술을 활용하여 경희대학교 규정 검색 챗봇을 개발한 사례를 제시한다. 방대한 대학 규정집의 효율적 검색과 일관된 해석을 위해 벡터 검색 시스템과 에이전트 Workflow를 구현하였다. PDF/HWP 문서를 JSON으로 구조화하고 FAISS 벡터 데이터베이스를 구축하여 컨텍스트 기반 RAG 시스템을 개발하였다. 성능 평가 결과 Faithfulness가 0.361에서 0.516으로 43% 향상되었으며, 교직원 업무시간 45% 감소, 학생 월간 처리시간 64% 단축(10.6시간→3.8시간)의 실무 개선 효과를 확인하였다. 본 연구는 LLM 기술의 실제 대학 행정 적용 가능성을 입증한 실증 사례이다.

1. 서론

대학 규정은 학사 운영, 조직 인사, 학생 활동, 재정 시설 등 대학 운영 전반을 규율하는 핵심 지침이다. 경희대학교와 같은 대규모 종합대학은 수백 개의 규정 문서와 수천 개의 세부 조항을 보유하고 있어, 행정 직원들이 특정 상황에 맞는 규정을 신속히 찾아 적용하기 어렵다. 특히 직원 간 해석 차이로 인한 일관성 문제와 규정 개정 이력 관리 부재로 최신 버전 확인이 어려워 행정 서비스의 신뢰성이 저하되고 있다. 기존 PDF 규정집이나 키워드 검색 시스템은 문맥 이해 능력이 없어 동일 용어의 다른 맥락(예: 일반 휴학/군 휴학/질병 휴학) 구분이 어렵고, 연관 규정들(예: 휴학-등록금 반환-복학)을 통합 제시하지 못한다. FAQ 시스템 역시 사전 정의된 질문 외에는 대응이 불가능하다. 이러한 비효율적 검색 과정은 반복 업무를 증가시켜 행정 생산성을 저하시킨다. 본 연구는 LLM과 RAG 기술을 활용하여 경희대학교 규정 검색 및 질의응답 AI 시스템을 개발하고자 한다. 제안 시스템은 규정, 시행세칙, 학사제도를 통합 처리하여 맥락 기반 의미 검색과 근거 기반 답변을 제공한다. 이를 통해 규정 해석의 일관성 확보와 행정 업무 효율성 개선을 목표로 한다.

2. 관련 연구

2.1 Foundation Model과 LLM 활용

Foundation Model은 대규모 데이터로 사전 학습되어 다양한 과제를 수행할 수 있는 기초 인공지능 모델이다 [1].

GPT-4, Claude, Gemini 등 최신 LLM 들은 자연어 이해와 생성 능력이 크게 향상되어 전문 도메인 질의응답 시스템 구축에 활용되고 있다. 특히 문제 해결 방법론으로 Prompt Engineering, RAG, Fine-tuning이 주로 활용된다 [2].

2.2 RAG 기술

Lewis et al. [3]이 제안한 RAG는 외부 지식베이스에서 관련 정보를 검색하여 언어모델의 생성 과정에 결합하는 방법이다. 이는 LLM의 환각(hallucination) 문제를 완화하고 최신 정보를 반영할 수 있어 기업 또는 기관의 지식관리시스템 구축에 널리 채택되고 있다. Gao et al. [4]은 RAG 시스템의 성능이 검색 품질과 context 통합 방식에 크게 의존함을 보였다.

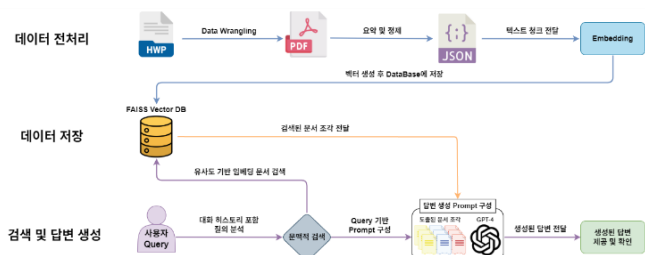
2.3 대학 챗봇 시스템 현황

국내에서는 고려대학교의 '쿠챗(KUChat)' 등 여러 대학이 AI 챗봇을 도입하였으나, 대부분 단순 FAQ 수준에 머물러 있으며, 특히 한국 대학 특유의 복잡한 규정 체계와 HWP 문서 형식을 처리하는 특화 시스템은 아직 부족한 실정이다.

3. 시스템 설계 및 구현

3.1 시스템 아키텍처

본 시스템은 그림 1과 같이 데이터 전처리, 벡터 저장, 검색 및 답변 생성의 3단계 파이프라인으로 구성된다. 경희대학교 규정 문서(HWP, PDF, JSON, etc.)를 수집하여



(그림 1) 3단계 규정 검색 챗봇 파이프라인

unstructured.io 라이브러리로 구조 기반 분할을 수행하였다. 문서는 JSON 형식으로 변환되어 메타데이터(문서 제목, 조항 번호, 페이지, 개정 일자)와 함께 저장된다.

3.2 하이브리드 검색 시스템

문서 검색을 위해 OpenAI의 text-embedding-3-large 모델을 사용하여 임베딩을 생성하고, FAISS 벡터 데이터베이스에 저장하였다. 이후 RecursiveCharacterTextSplitter를 활용하여 2048자 단위로 분할하고 256자의 오버랩을 적용하여 문맥 연속성을 확보하였다. FAISS의 기본 Flat 인덱스를 사용하여 전체 문서에 대한 정확한 유사도 검색을 수행한다.

3.3 History-Aware RAG 구현

검색된 문서와 사용자 질의를 LLM에 입력하여 답변을 생성한다. 시스템은 검색된 context를 기반으로 근거 있는 답변을 제공하며, 답변 시 참조한 문서의 출처를 함께 표시하여 투명성을 강화하였다.

4. 실험 및 결과

4.1 실험 설정

10개의 대표 질의(연구비 관리, 학사 여비, 비전임 교원, BK21 장학금 등)를 선정하여 PDF 직접 임베딩과 JSON 전처리 방식을 비교하였다. 평가는 RAGAS [5] 프레임워크의 Faithfulness와 Answer Relevancy 지표를 사용하였다.

4.2 정량적 평가 결과

표1은 성능 평가 결과를 보여준다. Faithfulness가 0.361에서 0.516으로 43% 향상되었으며, Answer Relevancy는 0.837에서 0.783으로 6.4% 감소하였다. 행정 업무에서는 정확성이 우선되므로 긍정적 변화로 해석된다.

4.3 사용자 만족도 조사

경희대 교직원 4명과 학생 10명 대상 2주 파일럿 운영을 실시하였다. 교직원의 일평균 규정 처리 시간이 45% 감소(2.2시간 절감)하였으며, 학생의 월간 규정 확인 시간은

10.6시간에서 3.8시간으로 64% 단축되었다. 또한 교직원의 50%가 "챗봇이 반복 업무를 효과적으로 대체한다"에 매우 동의하였다.

5. 결론

본 연구는 LLM과 RAG 기술로 경희대학교 규정 검색 챗봇을 구현하였다. JSON 전처리로 Faithfulness가 0.361에서 0.516으로 43% 향상되었으며, 교직원 업무시간 45% 절감, 학생 처리시간 64% 단축을 달성하였다. 또한 HWP 문서 처리를 최적화하여 한국 대학 환경에 특화된 시스템을 구축하였다. 본 시스템은 복잡한 규정 체계를 가진 대규모 조직의 디지털 전환 모델을 제시한다. 현재 한계는 드문 쿼리 처리, 실시간 규정 개정 반영, 복잡한 질의에서의 관련성 저하이다. 향후 규정 온톨로지와 지식 그래프를 구축하여 규정 간 우선순위와 예외 관계를 모델링하고, SPARQL 기반 정답 계산과 RAG 문서 근거를 융합한 KG-RAG 하이브리드 시스템으로 발전시킬 예정이다. 개발된 시스템은 <https://kyunghye-chatbot-aimslab.streamlit.app>에 배포되어 있다. 본 연구는 대학 행정 AI 적용의 실증 사례로서 한국 대학의 지식관리시스템 혁신 기반이 될 것으로 기대된다.

참고문헌

- [1] Bommasani, R., Hudson, D. A., Adeli, E., et al., "On the opportunities and risks of foundation models", arXiv preprint arXiv:2108.07258, 2021.
- [2] Brown, T., Mann, B., Ryder, N., et al., "Language models are few-shot learners", 『Advances in Neural Information Processing Systems』, Vol. 33, 2020, pp. 1877-1901.
- [3] Lewis, P., Perez, E., Piktus, A., et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks", 『Advances in Neural Information Processing Systems』, Vol. 33, 2020.
- [4] Gao, Y., Xiong, Y., Gao, X., et al., "Retrieval-augmented generation for large language models: A survey", arXiv preprint arXiv:2312.10997, 2023.
- [5] Es, S., James, J., Espinosa-Anke, L., Schockaert, S., "RAGAS: Automated evaluation of retrieval augmented generation", arXiv preprint arXiv:2309.15217, 2023.

Acknowledgement

본 연구는 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행되었음(No.RS-2022-00155911, 인공지능융합혁신인재양성(경희대학교)).

(표 1) 성능 평가 결과

평가 지표	PDF 임베딩	Json 전처리	개선율
Faithfulness	0.361	<u>0.516</u>	+43%
Answer Relevancy	<u>0.837</u>	0.783	-6.4%