

국내학술대회 논문집

제10권 제1호

일시 | 2026.6.26(금)-27(토)

장소 | 한국과학기술회관

주관: (사)국제문화기술진흥원(IPACT), (사)국제인공지능학회(IAAI), 지식의숲

후원: 과학기술정보통신부, 한국연구재단, 한국과학기술단체총연합회

모션 센싱 기술을 활용한 실내 운동기구 제작과 회상 요법 기반의 인터랙티브 케어
콘텐츠 연구 / 57

김준호, 이호준, 박상영, 오현석, 박성준 (성결대학교)

퍼즐형 게임에서의 동적 난이도 조정(DDA) 개입 조건에 따른 사용자 경험 분석 / 64

권순형, 김용범, 박서연, 원보경, 김정이 (성결대학교)

현대 게임 시장에서 ‘보는 게임(Observational Gaming)’의 흥미 유지 메커니즘에
대한 심리학적 분석 / 70

이관용, 허석재, 황 샘, 심보연, 이상원, 허원희 (성결대학교)

의료 마이데이터 연계를 위한 계층적 신체 모델 기반 증상 데이터 정규화 및 AI 기반
사전 문진 프레임워크 연구 / 76

양유빈, 정동희, 박재완, 조승윤, 허원희 (성결대학교)

LLM & RAG 기반 저전력 파이썬 코드 최적화 / 84

김재호, 정은진, 김영철 (홍익대학교)

AI 기반 인터넷 유인성 피싱 기사 자동 판별 시스템 개발 / 90

남선우, 송한춘 (서일대학교)

AI를 활용한 딥페이크 이미지 판별 및 신뢰도 평가 시스템 개발 / 95

정준배, 송한춘 (서일대학교)

LLM의 환각 제어를 위한 기업 문서 RAG 프레임워크 연구 / 100

이원재, 강민준, 류강희, 윤성준, 김장환, 서채연, 김영철 (홍익대학교)

유연근무제가 직원의 직무번영에 미치는 영향: 지식획득과 내재적 동기의 연속적 매개효과
및 인지적 유연성의 조절효과 / 103

리자오치, 최명철, 통지아후이 (가천대학교), 천홍(서경대학교) 이훈석 (중앙대학교)

재난 현장 기반 LLM WIFI CSI정보를 통한 피구조자 감지 / 107

권성민, 서채연, 김영철 (홍익대학교)

소비재기업의 B2B 의사결정 최적화를 위한 한국형 합성 페르소나 시뮬레이션 융합 연구
/ 110

민경호(취르르코리아), 문예찬, 조유리, 이한진 (한동대학교)

AI 챗봇을 통한 아동의 정서 표현과 디지털 돌봄 커뮤니케이션 가능성 탐색 -
지역아동센터 현장 테스트를 중심으로 / 114

허예은, 박거성, 이한진 (한동대학교), 황장준(구글 클라우드)

생성형AI 콘텐츠 구조화 품질의 자동 평가 프레임워크 (SPCQI) / 121

조은선(메타큐레이션), 이지수, 최아인, 이한진 (한동대학교)

LLM의 환각 제어를 위한 기업 문서 RAG 프레임워크 연구

A Study on Enterprise Document-Based RAG Framework for Hallucination Control in Large Language Models

이원재*, 강민준*, 류강희*, 윤성준*, 김장환**, 서채연***, 김영철****

Won Jae Lee*, Min Jun Kang*, Kang Hui Ryu*, Seong Jun Yoon*,

Janghwan Kim**, Chaeyun Seo***, Young Chul Kim****

{sojoongyi*, coja0727*, ryukh1129*, ysjysj1212*}@naver.com,

{lentoconstante**, chaeyun***, bob****}@hongik.ac.kr

요 약

최근 기업 내부에서는 인사 규정, 보안 정책, 업무 매뉴얼, 프로젝트 문서 등 다양한 문서가 PDF와 사내 Wiki 형태로 분산되어 관리되고 있다. 이러한 문서들은 기업 운영에 필수적인 정보를 포함하고 있지만, 직원이 필요한 내용을 찾기 위해 여러 문서를 직접 검색 해야 하므로 정보 탐색 비용이 증가한다. 또한, 동일한 질문이 인사, 총무, 보안 담당자에게 반복적으로 전달되어 업무 효율성이 저하되는 문제가 발생한다. 본 논문에서는 기업 내부 문서를 기반으로 자연어 질문에 답변하고, 답변의 근거가 되는 문서 출처를 함께 제공하는 RAG 기반 Q&A 챗봇 시스템 Intra-Q를 제안한다. 제안 시스템은 PDF 업로드, 텍스트 추출, 청크 분할, 문서 관리, 채팅 기반 질의응답, 출처 카드 표시 기능으로 구성된다. 프론트엔드는 Vite 기반 React와 React Router를 활용하여 구현하였고, 백엔드는 FastAPI와 SQLite를 기반으로 문서 업로드 및 질의응답 API를 제공한다. 본 시스템은 기업 내부 문서 검색 시간을 줄이고, 출처 기반 답변을 통해 AI 응답의 신뢰성을 높이는 것을 목표로 한다.

키워드 : RAG, 기업 문서 검색, Q&A 챗봇, FastAPI, React, 출처 기반 답변

I. 서 론

기업의 업무 환경에서는 인사 규정, 복지 제도, 보안 정책, 업무 매뉴얼 등 다양한 내부 문서가 지속적으로 생성되고 축적된다. 이러한 문서는 PDF, 사내 Wiki, 공유 드라이브 등 여러 형태로 분산되어 있어 사용자가 필요한 정보를 신속하게 찾는 데 어려움이 발생한다. 최근 대규모 언어 모델(LLM)을 활용한 챗봇 서비스가 확산되고 있으나, 일반적인 생성형 AI는 기업 내부 문서에 기반하지 않기 때문에 정확한 답변을 제공하기 어렵고 환각 현상의 위험도 존재한다. 이에 본 논문에서는 기업 내부 PDF 문서를 기반으로 사용자의 질문에 답변하고, 문서명, 페이지 번호, 관련 문장 등 출처 정보를 함께 제공하는 RAG 기반 Q&A 챗봇 시스템 Intra-Q를 제안한다.

제안 시스템은 문서 업로드, 텍스트 추출, 청크 분할, 관련 문서 검색, 답변 생성 및 출처 카드 제공 기능을 통해 사용자가 AI 답변의 근거를 직접 확인할 수 있도록 설계한다. 또한 연도별 규정 변화가 복잡한 대학 교과과정 도메인을 테스트베드로 선정하여 제안 시스템의 실효성을 검증한다. 본 논문의 구조는 다음과 같다, 2장은 관련연구에 대해 언급하고, 3장은 LLM환각제어를 위한 RAG 파이프라인에 대해 설명하고, 4장은 RGA 파이프라인을 설명하고 5장은 결론 향후 연구를 언급한다.

II. 관련 연구

대규모 언어 모델의 생성 능력과 외부 문서 검색 기능을 결합한 RAG(Retrieval-Augmented Generation) 구조를 제안한다 [1][5]. 모델 내부 지식만으로는 최신 정보나 특정 도메인 지식을 충분히 반영하기 어렵다는 한계를 보완하기 위해, 검색된 외부 문서를 답변 생성 과정에 활용한다. 이러한 접근은 기업

*홍익대학교 소프트웨어융합학과 학부생

**홍익대학교 소프트웨어융합학과 외래교수

***홍익대학교 소프트웨어융합학과 초빙교수

****홍익대학교 소프트웨어융합학과 교수(교신저자)

내부 PDF 문서를 기반으로 정확한 답변과 출처를 제공하고자 하는 본 연구의 이론적 기반이 된다.

PDF 문서를 주요 데이터 소스로 활용한 RAG 기반 LLM 시스템의 개발 과정을 제시한다. PDF 문서 수집, 전처리, 검색 인덱싱, 답변 생성으로 이어지는 전체 처리 절차를 다루며, 구조화 및 비구조화 문서를 활용할 경우 응답의 정확성과 문맥 적합성을 향상시킨다. 이는 기업 내부 PDF 문서를 대상으로 텍스트 추출, 청크 분할, 관련 문서 검색, 답변 생성 및 출처 제공 기능을 구현하는 시스템 구조와 관련된다 [2].

III. LLM의 환각 제어를 위한 기업 문서 RAG 프레임워크

1. 시스템 개요

LLM 환각 제어를 위한 프레임워크는 기업 및 대학 내부 문서를 활용한 웹 기반 RAG Q&A 챗봇 시스템이다. 사용자는 PDF 문서를 업로드하고 자연어 질문을 통해 문서에 기반한 답변을 받을 수 있다 [3]. 전체 시스템은 프론트엔드, 백엔드, 관계형 데이터베이스, 벡터 데이터베이스, RAG 파이프라인의 5개 컴포넌트로 구성된다 [4].

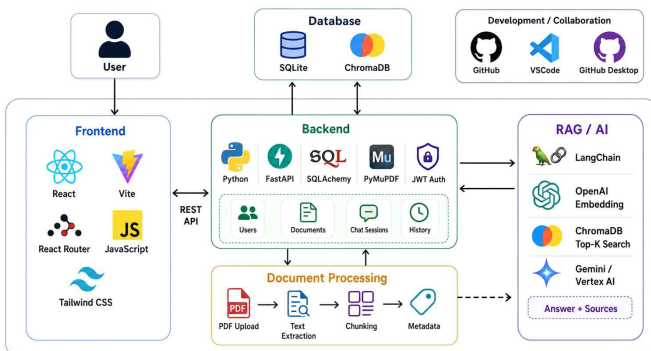


그림 1. Intra-Q 전체 시스템 아키텍처

PDF 업로드 시 PyMuPDF를 활용하여 페이지별 텍스트를 추출하고 청크 단위로 분할 한다. 교과과정 문서의 경우 연도, 대학, 학과 메타데이터가 함께 저장되어 벡터 데이터베이스에 반영된다. 사용자별 문서 접근 권한은 JWT 인증을 통해 관리되며, 다른 사용자의 문서에는 접근할 수 없다.

IV. RAG 파이프라인 설계

RAG 파이프라인은 멀티쿼리 확장, 벡터 검색, RRF융합, 리랭킹, 답변 생성의 5단계로 구성된다.

- 청크 분할 및 임베딩
- 형태소 분석 기반 BM25 하이브리드 검색

c. 멀티쿼리 확장

d. RRF 융합

f. Gemini 2.5 Flash 답변 생성

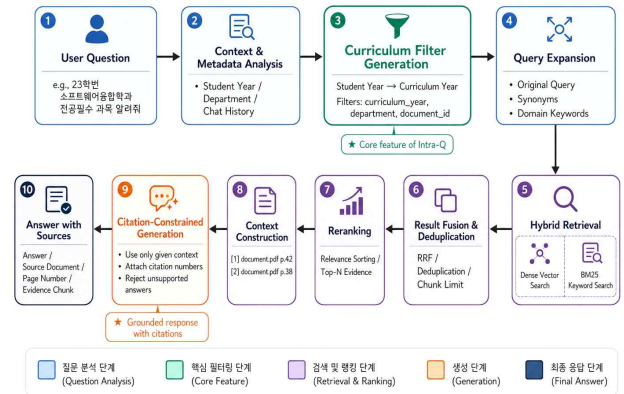


그림 4. RAG 파이프라인 처리 흐름

시스템의 핵심 기능은 출처 기반 답변 UI 설계를 활용하여 AI 답변과 함께 출처를 제공하는 것이다. 답변 하단의 출처 카드에는 문서명, 페이지 번호, 관련 청크 내용, 유사도 점수가 표시된다. 교과과정 문서의 경우 교과과정 연도, 학과명, 과목명, 이수구분, 학점 정보가 추가로 포함되어 사용자가 어떤 연도·학과 기준으로 답변이 생성되었는지 직접 확인할 수 있다.

실험 데이터셋은 홍익대학교 교과과정 PDF 5개년치 (2021~2025)와 비교용 기업 문서인 SK하이닉스 사업보고서로 구성한다. 교과과정 문서는 2021년 452개, 2022년 488개, 2023년 495개, 2024년 503개, 2025년 1,000개의 청크로 분할하였으며, SK하이닉스 사업보고서는 2,786개의 청크로 구성하여 전체 6,224개의 청크를 실험에 활용한다. 특히 SK하이닉스 사업보고서는 교과과정과 무관한 문서로, 연도 필터를 적용하지 않은 환경에서 검색 결과가 다른 도메인 문서에 의해 오염되는지를 검증하기 위해 동일 컬렉션에 포함한다.

IV. 결 론

본 논문에서는 기업 및 대학 내부 문서를 기반으로 자연어 질문에 답변하고, 답변의 출처를 함께 제공하는 RAG 기반 Q&A 챗봇 시스템 Intra-Q를 제안하였다. 제안 시스템은 문서 검색, 멀티쿼리 확장, RRF 융합, 리랭킹, LLM 답변 생성으로 구성된 RAG 파이프라인을 통해 내부 문서 기반 답변의 정확성과 신뢰성을 높이하고자 하였다. 특히 교과과정 연도 기반 필터링 기법을 적용하여 타연도 문서 검색으로 인한 오염률을 감소시키고 출처 정확도를 향상시켰다. 실험 결과, 연도 필터 적용 시 오염률은 62.5%에서 0%로 감소하였으며, 출처 정확도는 46.7%에서 86.7%로 향상되었다. 또한 BM25 기반 검색을

결합하여 검색 속도와 정확도를 개선하였다. 향후에는 스트리밍 응답, 계층적 청킹, 표 구조 인식 기능을 추가하여 교과과정 문서와 같은 복잡한 내부 문서에 대한 검색 품질을 더욱 향상시킬 계획이다.

※ 본 연구는 홍익대학교 소프트웨어융합학과 학부생들의 캡스톤디자인 프로젝트 결과물이다. 또한 본 연구는 정보통신기획평가원 SW전문인재양성사업 홍익대학교 메타버스융합 SW아카데미 제6기 교육생들의 프로젝트 결과물이다. (202301830004, 메타버스 융합SW 아카데미)

참 고 문 헌

- [1] Y. Gao et al., “Retrieval-Augmented Generation for Large Language Models: A Survey,” arXiv preprint arXiv:2312.10997, 2023.
- [2] G. Cormack, C. Clarke, and S. Buettcher, “Reciprocal Rank Fusion Outperforms Condorcet and Individual Rank Learning Methods,” Proc. 32nd Int’l ACM SIGIR Conf., pp. 758-759, 2009.
- [3] Google Cloud, “Vertex AI Ranking API, “
<https://cloud.google.com/generative-ai-app-builder/docs/ranking>, 2024.
- [4] T. Chen et al., “ChromaDB: The AI-native Open-source Embedding Database,”<https://docs.trychroma.com/>, 2023.
- [5] P. Lewis, et al, “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,” Advances in Neural Information Processing Systems, vol. 33, pp. 9459-9474, 2020.