

RAG 기반 대화형 채용공고 추천 시스템

권준혁¹, 신성준², 장경훈³, 정경훈⁴, 최지혁⁵
 인하대학교 인공지능공학과¹, 건국대학교 컴퓨터공학부²,
 세종대학교 데이터 사이언스학과³, 지능기전공학과⁴, 경기대학교 융합데이터공학과⁵
 tree.jkh@gmail.com, shinsj4653@gmail.com, matkimchi_@naver.com, ghoon99@kakao.com, akfktl328@gmail.com

RAG-Based Conversational Job Recommendation System

Junhyuk Kwon, Seongjun Shin, Kyoungsoon Jang, Kyunghoon Jung, JeeHyuk Choi
 Department of Artificial Intelligence Engineering, Inha University
 Department of Computer Science and Engineering, Konkuk University
 Department of Data Science, Department of Intelligent Mechatronics Engineering, Sejong University
 Department of Convergence Data Engineering, Kyonggi University

요 약

본 연구는 구직자의 정보 부족 문제를 해결하기 위해 검색 증강 생성(RAG) 기술을 활용한 대화형 채용공고 추천 시스템을 제안한다. 이 시스템은 웹 크롤링, multi-modal 데이터 처리, ETL 과정을 통해 최신 채용 정보를 수집 및 정제하고, 사용자 의도를 정확히 파악하기 위해 경량화된 LLM (Large Language Model)과 GPT-4를 결합한 다단계 접근 방식을 채택한다. RAG(Retrieval Augmented Generation)를 통해 개인화된 채용 정보를 추천하며, Next.js, Spring, FastAPI를 활용한 통합 아키텍처로 구현되어 사용자에게 실시간으로 맞춤형 취업 정보를 제공한다.

1. 서론

최근 취업 시장에서 구직자의 정보 부족으로 인해 취업 준비가 더욱 어려워지고 있다. 따라서, 구직자들에게 개인화된 채용공고를 실시간으로 추천하는 새로운 형태의 취업 정보 제공 시스템의 필요성이 강조되고 있다.

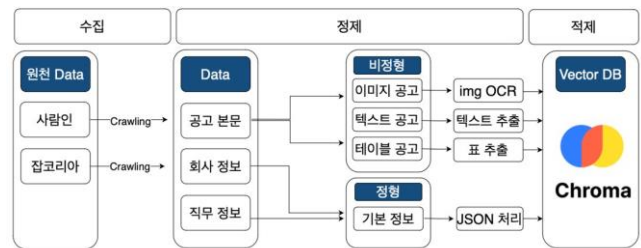
본 연구에서는 이러한 문제를 해결하기 위해 검색 증강 생성(Retrieval Augmented Generation, RAG) 기술을 활용한 실시간 채용공고 추천 챗봇 시스템을 제안한다 [1]. 이 시스템은 최신 채용 정보를 지속적으로 수집하고, 사용자의 개인화된 요구사항을 반영하여 맞춤형 정보를 제공하는 것을 목표로 한다.

2. 본론

본론에서는 데이터 수집 파이프라인, 채팅 의도 분류 및 맞춤형 응답 생성 시스템, 그리고 RAG 기반 대화형 채용정보 추천 시스템의 아키텍처를 설명한다.

2-1. 데이터 수집 파이프라인

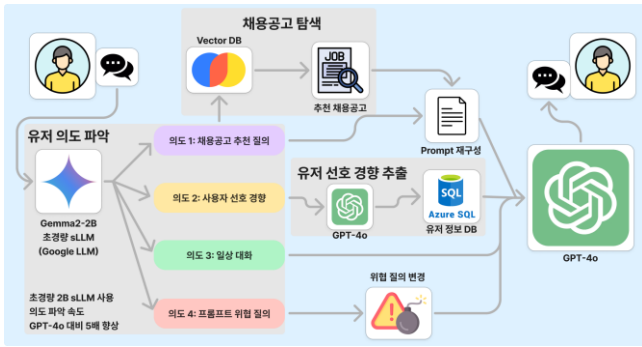
본 시스템의 데이터 수집 파이프라인은 웹 크롤링을 통한 데이터 수집과 수집된 데이터의 정제로 나뉜다. 먼저, 웹 크롤링을 통한 데이터 수집은 취업 포털 사이트인 잡코리아와 사람인에서 채용 정보를 수집하기 위해 BeautifulSoup4(bs4)와 Selenium을 사용한다. JavaScript로 렌더링되는 동적 콘텐츠도 누락 없이 수집한다. 데이터 수집 과정은 Apache Airflow를 통



(그림 1) 데이터 수집 파이프라인

해 자동화된 스케줄링으로 정기적으로 업데이트되며, OCR을 사용하여 이미지 속 텍스트를 추출하고, LLM의 In-context Learning을 통해 OCR 중 발생한 오타자에 대한 교정을 수행해 데이터의 품질을 높인다.

수집된 원시 데이터는 ETL(Extract-Transform-Load) 과정을 통해 정제되며, 특히 Apache Spark를 활용한 Transform 단계에서는 대규모 데이터의 병렬 처리, 텍스트 분석, 키워드 추출, 문서 유사도 계산 등이 수행된다. 이 처리된 데이터는 JSON 형식으로 표준화되어 Azure Cosmos DB와 ChromaDB에 저장되며, 각각 사용자 세션 관리와 벡터 기반 텍스트 검색 및 추천에 활용된다. 더불어, 사용자의 피드백을 실시간으로 반영하여 데이터의 정확성과 개인화된 추천의 품질을 지속적으로 개선하고 있다. 이러한 종합적인 데이터 수집 및 처리 시스템을 통해 본 서비스는 텍스트와 이미지 정보를 정제하고, 최신 채용 공고를 실시간으로 제공하는 기능을 갖추고 있다.



(그림 2) 맞춤형 응답 생성 시스템

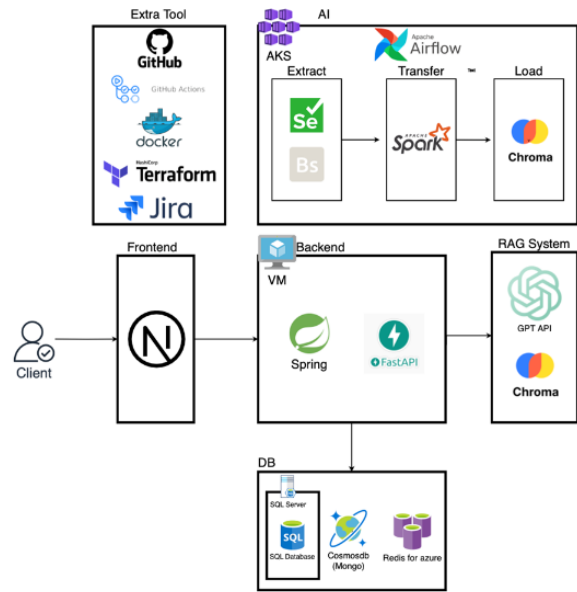
2-2. 채팅 의도 분류 및 맞춤형 응답 생성 시스템

본 시스템은 그림 2 에서와 같이, 사용자 입력의 정확한 의도 파악을 위해 다단계 접근 방식을 채택하고 있다. 초기 단계에서는 Google 사에서 배포한 초경량 2B 크기의 Large Language Model (LLM)인 Gemma-2B¹ 모델을 활용하여 사용자 입력의 의도를 신속하게 분류한다 [2]. 이 모델은 GPT-4² 대비 5 배 빠른 처리 속도를 보여, 실시간 대화 시스템에 적합한 성능을 제공한다. 의도는 '채용공고 추천 요청', '일반적인 취업 관련 질문', '개인 선호도 정보 제공', '잠재적 위협 질의'의 4 가지로 구분된다. 보안 측면에서는 잠재적 위협 질의(프롬프트 탈취 질의 등)를 감지하여 시스템의 안전성을 확보한다.

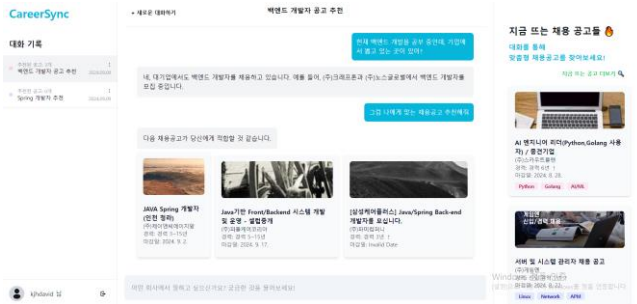
분류된 의도에 따라 시스템은 다양한 동작 모드로 전환되어 최적화된 응답을 제공한다. 채용공고 추천 의도로 분류되면, RAG(Retrieval Augmented Generation) 알고리즘을 통해 ChromaDB 에 저장된 채용 정보 중 사용자의 프로필과 선호도에 가장 적합한 공고를 벡터 유사도 기반으로 추출한다 [1]. 추출된 정보는 GPT-4 모델을 통해 자연스러운 대화 형식으로 재구성되어 사용자에게 전달된다. 개인 선호도 정보 제공의 경우, GPT-4 모델을 사용해 사용자의 발언에서 중요한 선호 사항을 추출하고 분석한 후, 이를 Azure SQL 데이터베이스에 저장하여 향후 추천에 활용한다. 이러한 다양한 동작 모드를 통해 시스템은 각 상황에 최적화된 대응을 할 수 있으며, 사용자에게 항상 관련성 높고 개인화된 정보를 제공할 수 있게 된다.

2-3. 시스템 아키텍처

본 시스템은 그림 3 에서와 같이, 데이터 수집부터 사용자 응답 생성까지 전 과정을 아우르는 통합 아키텍처로 구성되어 있다. 앞서 기술한 데이터 파이프라인과 대화 생성 알고리즘과 더불어 프론트엔드는 Next.js 로 구현되어 반응형 사용자 인터페이스를 제공하고, 백엔드는 Spring 과 FastAPI 를 조합하여 안정적이고 확장 가능한 서비스 로직을 구현한다. 전체 시스템은 Docker 와 Terraform 을 활용해 클라우드 환경에서 안정적으로 운영된다. 이러한 통합 아키텍처를 통해 사용자는 실시간으로 업데이트되는 정확하고 개인화된 채용 정보를 대화형 인터페이스를 통해 쉽게 얻을 수 있다.



(그림 3) 시스템 아키텍처



(그림 4) RAG 기반 채용정보 챗봇 웹 인터페이스

3. 결론

본 연구에서 제안한 RAG 기반 대화형 채용공고 추천 시스템은 구직자의 취업 정보 부족 문제를 해결할 수 있는 솔루션이다. OCR 과 LLM 기반의 데이터 수집 기법, 정교한 의도 분류 알고리즘, 그리고 개인화된 RAG 기반 응답 생성 방식을 통합한 이 시스템은 사용자에게 최신의, 관련성 높은 채용 정보를 실시간으로 제공한다. 또한, 그림 4 와 같이 직관적이고 반응형인 사용자 인터페이스는 구직자들이 복잡한 채용 정보를 쉽게 탐색하고 이해할 수 있게 돕는다.

참고문헌

[1] Gao, Yunfan, et al. "Retrieval-augmented generation for large language models: A survey." arXiv preprint arXiv:2312.10997 (2023).

[2] Team, Gemma, et al. "Gemma 2: Improving open language models at a practical size." arXiv preprint arXiv:2408.00118 (2024).

※ 본 논문은 과학기술정보통신부 대학디지털교육역량강화 사업의 지원을 통해 수행한 ICT 멘토링 프로젝트 결과물입니다.