

생성형 AI 및 RAG 활용한 항만공사 챗봇 서비스 개발

김태강¹, 신성준¹, 이강준¹, 조성빈¹, 박준규², 김인수³

¹인하대학교 정보통신공학과, ²한국항공대학교 소프트웨어학과, ³ECS텔레콤
 xorkd119@gmail.com, sinsung1084@naver.com, ererer1001@naver.com, whtjcdqslqp@naver.com, junju404@naver.com

Development of the Port Authority Chatbot Service Using Generative AI and RAG

Tae-Gang Kim¹, Seong-Jun Shin¹, Kang-Jun Lee¹, Seong-Bin Cho¹, Jun-Gyu Park², In-Soo Kim³

¹Dept. of Information and Communication Engineering, Inha University,

²Dept. of Software Engineering, Korea Aerospace University, ³ECS Telecom

요 약

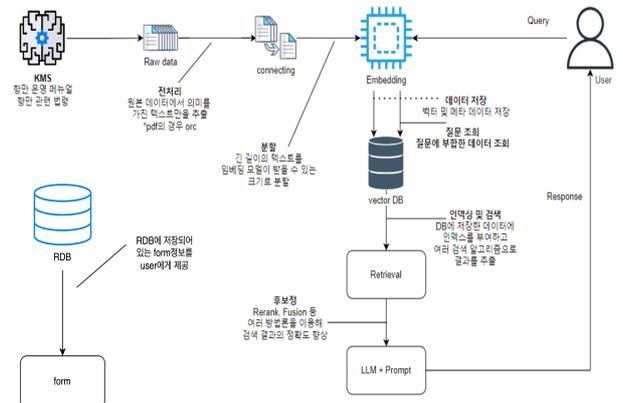
본 연구는 생성형 AI와 검색 증강 생성(Retrieval-Augmented Generation, RAG) 기술을 결합하여 환각(Hallucination)을 방지하고 항만공사 환경에 특화된 챗봇 서비스를 제안한다. 주요 특징은 첫째, KMS 저장소, 벡터 데이터베이스, 검색 및 재조정 처리, LLM을 이용한 프롬프트 처리 등을 통해 사용자 질의에 최적화된 응답을 생성한다. 둘째, ‘챗봇 One-Step-Service’은 사용자에게 업무 처리용 Form을 제공해 업무 처리의 편의성을 향상시킨다. 해당 시스템을 항만공사에 적용하여 한정된 상담 인력 문제를 해결하고 다양한 내부 업무에 적용하여 업무 효율성이 향상될 것으로 기대된다.

1. 서론

디지털 전환이 가속화됨에 따라, 다양한 산업 분야에서 인공지능(AI) 기술의 활용이 필수적 요소로 자리잡고 있다[1]. 수년 전부터 국내외 ICT기업들이 앞다투어 인공지능을 이용한 챗봇 서비스를 출시하며 챗봇 서비스의 효율성은 사실상 입증된 상태이다[2]. ‘Facebook AI Research’의 ‘Retrieval Augmented Generation for Knowledge-Intensive NLP Tasks’ 연구는 생성형 AI의 발전을 촉진하고 한계 극복을 위한 방안으로 RAG를 활용하는 방안을 제시하였다[3]. 본 연구는 생성형 AI와 RAG 기술을 결합하여 항만공사 환경에 특화된 챗봇 서비스를 개발하는 것을 목표로 한다. 고객센터는 한정된 인력으로 운영되며, 고객들은 방문 신청이나 항만 관련 다양한 정보를 신속하게 얻길 원한다. 동시에 내부 직원들은 잦은 인사 이동으로 인해 인수인계가 원활하지 않은 상황에서 업무 처리에 어려움을 겪고 있다. 이러한 문제들을 해결하기 위해 생성형 AI와 RAG모델을 접목한 챗봇이 필수적이다. 이 챗봇 서비스는 실시간으로 사용자 문의에 최적화된 답변과 다양한 정보를 제공할 수 있다. 이를 통해 내부 직원들이 자주 하는 질문이나 업무 절차에 대한 안내를 제공함으로써 업무 효율성을 높이고, 내부 인수인계 과정에서 발생할 수 있는 문제를 줄이는 데 기여할 수 있다. 생성형 AI는 자연스러운 대화를 통해 사용자의 요구를 이해하고, 필요한 정보를 신속하고 정확하게 검색하여 답변을 생성함으로써 기존의 챗봇보다 높은 수준의 서비스를 제공한다[4]. 또한, 사용자의 불편함을 덜기 위해 자주 하는 질문 中 방문 신청과 같은 복잡한 절차도 챗봇 One-step-Service를 통해 해결할 수 있다. 해외 담당자나 사용자가 영어로 문의할 경우에도 즉각적으로 대응할 수 있어, 다국적 대응이 필요한 항만공사에 필수적이다.

2. 본론

2.1 시스템 구성도



(그림 1) 시스템 구성도

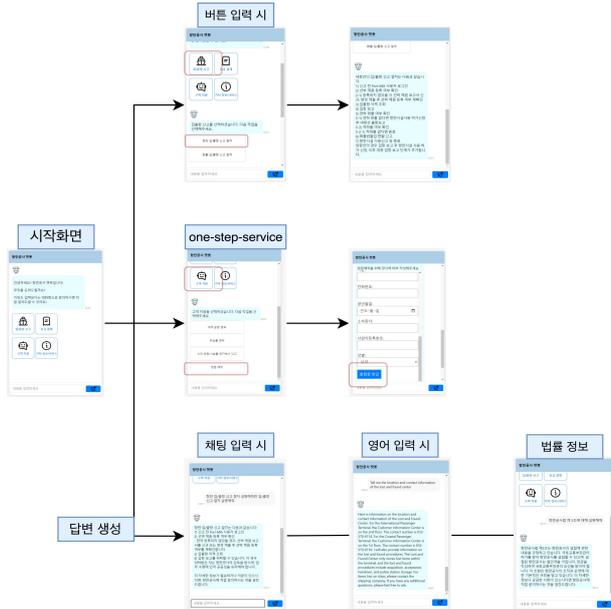
구성요소	구성요소 기능
KMS 저장소	항만공사 내부 지식 데이터는 KMS에 저장된다. 항만공사법 외 6건 (pdf 파일) 항만 출입 신청 절차 외 11종 (pdf 파일)
연결 및 처리	KMS 지식 데이터는 전처리 후 시스템으로 전송된다.
임베딩 처리	텍스트 데이터는 all-MiniLM-L6-v2 모델을 사용해 벡터 형태로 변환한다. 변환된 데이터는 의미 보존 및 검색·처리를 가능하게 한다.
벡터 DB 저장	변환된 벡터 데이터는 벡터 데이터베이스에 저장되며, 이는 검색 및 유사도 비교에 이용된다.
RDB	사용자가 양식에 입력한 정보를 RDB에 저장하고, 관리자는 저장된 정보를 조회하거나 관리한다.
검색 및 재조정 (Rerank) 처리	질의에 기반한 데이터 검색이 수행되고, FAISS와 Chroma의 검색 알고리즘을 활용해 연관성이 높은 데이터를 재조정하는 과정을 거친다.
LLM 이용한 프롬프트 처리	검색·재조정된 데이터는 LLM 모델 GPT 3.5 Turbo 활용하여, 사용자 질의에 대한 최적의 응답을 항만공사 챗봇에 특화된 프롬프트를 거쳐 생성한다.
질의 입력 및 응답 출력	사용자 질의는 시스템에 입력되고, 시스템은 처리된 응답을 실시간으로 사용자에게 제공한다.

<표 1> 시스템 구성요소

(그림 1)은 시스템 구성을 도식화한 것이며, <표 1>은 시스템의 핵심 구성요소 및 구성요소의 기능을 작성했다.

2.2 시스템 기능

2.2.1 모바일 흐름도

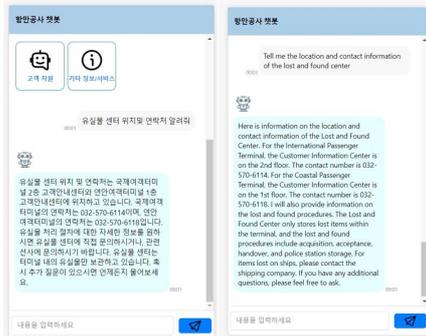


(그림 2) 항만공사 챗봇 서비스 흐름도

(그림 2)는 챗봇 서비스 흐름을 도식화한 것이다. 고객이 자주 하는 질문을 손쉽게 찾을 수 있는 ‘버튼 기능’, 법률 및 항만 관련 질문에 적합한 답변을 생성하는 ‘답변 생성’, 고객의 업무 처리 편의성을 제공하는 ‘챗봇 One-Step-Service’를 제공한다.

2.2.2 답변 생성

(그림 3)은 인공지능의 환각(Hallucination)을 방지하고 알맞은 응답을 생성한 화면이다. Langchain 라이브러리를 통해 구현한 RAG 기술을 통해 항만공사 내부 지식 데이터 기반으로 답변을 제공한다. FAISS를 활용해 거리 기반 검색을 수행하고 Chroma 내부 Rerank 기능 및 Fusion 알고리즘 통해 환각 발생을 최소화한다. 이러한 기술들로 Rule 기반 챗봇에 비해 운용 용이성 및 답변의 정확도를 확보한다.



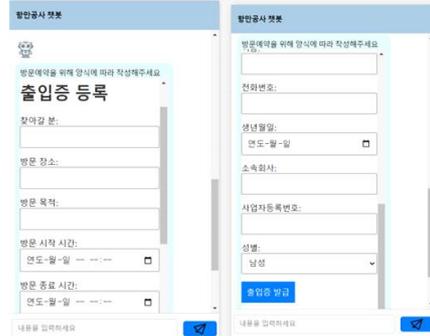
(그림 3) 답변 생성

2.2.3 챗봇 One-Step-Service

(그림 4)는 ‘챗봇 One-Step-Service’ 중 방문 예약 신청 화면이다. 해당 서비스는 항만공사 홈페이지에서 제공하는 ‘공

모전 등록’이나 ‘방문 예약과 같은 복잡한 업무를 챗봇에서 직접 처리할 수 있도록 해 신속한 업무 처리가 가능하다. 위의 두 업무 처리는 울산항만공사의 ‘대국민 공모 절차’와 ‘출입증 신청 절차’를 참고했다.

사용자는 (그림 4)의 양식에 정보를 작성하여 제출하면 항만공사 관리자는 제출된 정보를 검토 및 처리할 수 있다. 관리자는 실시간으로 동기화된 정보를 바탕으로 예약 정보를 확인하고, 출입증 발급 등 필요한 조치를 할 수 있다. 본 기능을 통해 고객은 기존보다 쉽게 방문 예약 등을 처리할 수 있다.



(그림 4) 챗봇 One-Step-Service 中 방문 예약 Form

3. 결론

본 연구에서는 항만공사에 특화된 지능형 챗봇 시스템 개발을 위해 생성형 AI와 RAG 기술을 결합한 접근 방식을 제안했다. ‘실시간 정보 제공’, ‘다국어 지원’, 복잡한 절차를 간소화하는 ‘챗봇 One-Step-Service’ 등의 기능은 기존 챗봇의 한계를 극복하며, 사용자 편의성과 효율성을 크게 개선 시킬 것이다. ‘답변 생성’ 기능을 통해 벡터 데이터베이스와 RAG 프로세스를 활용해 응답의 정확성을 보장했으며, 관리 시스템의 연계를 통해 쉽게 정보 처리가 가능해졌다. 이 시스템을 항만공사에 적용함으로써 한정된 상담 인력 문제를 해결하고 기관 내부의 다양한 업무에 적용하여 업무 효율성이 향상될 것으로 기대된다.

Acknowledgement

본 논문은 해양수산부 실무형 해상물류 일자리 지원사업 (스마트해상물류 x ICT멘토링)을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.

참고문헌

[1] 김민석. "본격 'DX·DT' 시대... 한국, '디지털전환' 격전지 됐다." 서울신문, 2022.
 [2] 한국지능정보사회진흥원. 인공지능 기반의 ‘챗봇(ChatBot)’ 등장과 발전동향. ICT융합 Issues&Trends, 2016, 5p.
 [3] Lewis, P., et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. NeurIPS 2020, Vancouver, Canada, 2020.
 [4] 정찬수. "전통적인 챗봇과 ChatGPT 연계 서비스 방안 연구." 정보기술융용관리학회지, 2023, 12p