동국대학교 맞춤형 정보 제공 챗봇 시스템

동똑이

Team

Renux

팀 소개 및 역할

Renux

조준용
팀장

경찰행정학부

2020111242

백엔드 서버 구현

신원철
팀원

통계학과

2021110445

RAG 모델 구현

육심호
팀원

통계학과

2021110473

웹 프론트 구현

OSSprac의 Team Nux ➡ OSSProject의 Team Re:nux

목차

프로젝트 개요

동똑이란?

"동국대학교 똑똑이"의 줄임말로, 동국대학교의 내규, 학과별 정보, 부서별 정보 등 학교의 전반적인 정보를 **대화형으로 제공**해주는 AI 챗봇 서비스

생성형 AI 기술을 활용하여 기존 챗봇의 한계를 극복하고 학생들의 정보 습득에 편리함을 제공함

내규 정보

학교 정책, 규정, 절차에 대한
정보를 대화형으로 제공

학과별 정보

각 학과별 일정, 과목 정보, 커리큘럼 등
맞춤형 정보 제공

부서별 정보

학생처, 장학처, 입학처 등
부서별 연락처 및 업무 내용 안내

차별점

- 템플릿 기반 응답 구조가 아닌, **자연어 이해 능력**으로적 절한 답변 제공
- 링크 중심이 아닌, **정확한 정보 매칭**으로 부정확한 정보 노출 방지
- **맥락 기반 답변**으로 이어지는 대화에서도 일관된 정보 제공
- **사용자 추가 탐색 없이** 필요한 정보를 즉시 확인할 수 있는 구조

현재 시스템의 문제점

현재 운영중인 동국대학교의 챗봇 서비스의 구조적 한계

1

템플릿 기반 응답 구조

사전에 정해진 질문-답변 형식을 기준으로 작동하기 때문에, 템플릿을 벗어난 질문이나 복합적인 맥락을 요구하는 문의에는 적절한 답변을 제공하지 못합니다.

2

정보 정확도의 리스크

사용자가 의도한 질문과 연결된 링크의 내용이 일치하지 않을 가능성이 존재하여, 부정확한 정보에 노출될 수 있습니다. 이는 챗봇 서비스에 대한 신뢰 저하로 이어질 수 있습니다.

3

링크 중심의 반환 방식

설명이나 요약이 아닌 링크 중심의 반환 방식으로 제한되어 있어, 사용자가 질문에 대한 직접적인 답을 얻기 어렵습니다.

4

추가 탐색 부담

원하는 정보에 접근하기 위해 사용자는 여러 페이지를 추가로 탐색해야 하는 불편을 겪게 됩니다. 이러한 추가 탐색 부담은 챗봇의 핵심 목적과 상충하며, 전체적인 사용자 경험을 저해합니다.

결론: 현재 챗봇 시스템은 보다 유연한 자연어 이해 능력, 맥락 기반 답변 제공, 정확한 정보 매칭, 그리고 사용자가 추가 탐색 없이 필요한 정보를 즉시 확인할 수 있는 구조로의 전환이 필요합니다.

경쟁 서비스 분석

기능 \ 챗봇	ChatGPT	서강대	서울여대	건국대	동국대(기존)	동국대(동똑이)
학과별 맞춤 답변 기능	✗	✗	✗	✗	✗	✓
정보의 최신성	−	−	−	−	✗	1일 4회
LLM 모델 종류	Chat GPT	Chat GPT	정보 없음	정보 없음	✗	Chat GPT
개인별 이전 대화내역 저장	✓	−	−	−	✗	✓
상황별 담당 부서 연락처 안내	✗	✓	✓	✓	✗	✓

핵심 분석 결과

- 모든 대학 챗봇이 **학과별 맞춤 답변** 기능을 제대로 구현하지 못함
- 정보의 최신성 측면에서도 모든 챗봇이 특출난 성능을 보이지 못함

동국대 동똑이의 차별점

- **학과별 맞춤 답변** 기능을 성공적으로 구현
- 매일 4회 자동 업데이트되는 **최신 정보** 제공

기술 동향 분석

1 대형 언어 모델 (LLM)

- **OpenAI GPT-4o / GPT-4o mini**: 한국어에 최적화된 최신 멀티모달 LLM, 다국어 질의응답과 지식 보강
- **Google Gemini 1.5 Pro**: 컨텍스트 윈도우가 길어 대용량 문서 기반 RAG 답변 정확도를 높임
- **Meta Llama-3 70B**: 오픈소스 LLM 중 상위 성능으로, 사내 데이터와 결합해 RAG 구축에 활용

2 임베딩 검색 인프라

- **Sentence Transformers** (KU RE-v1, KoSim CSE 등): 한국어 문장 임베딩 품질이 높아 공지 규정 같은 비정형 문서 검색에 적합
- **FAISS / Chroma DB**: 수십만 청크를 빠르게 탐색하는 벡터 DB로, 하이브리드 검색 및 재빌드 자동화를 지원
- **Elasticsearch + ELSER**: 전통적 키워드 · BM25와 Dense Retrieval을 결합한 하이브리드 검색으로 RAG 정밀도를 끌어 올림

3 파이프라인 / 워크플로우

- **LangChain / LlamaIndex**: 프롬프트 템플릿, 체인, 메모리 등을 모듈화해 RAG 워크플로를 빠르게 조립
- **Airflow / Prefect**: 크롤링 전처리 인덱스 재생성 작업을 스케줄링해 최신 데이터를 유지 기반 채팅 및 세션 모니터링
- **FastAPI / Flask**: 경량 REST API로 외부 서비스와 연동하며, 세션 모니터링 엔드포인트를 제공

참고 연구

- "KURE : Korea University Retrieval Embedding model" 고려대학교 NLP & AI 연구실 + HIAI.
- "IRAG Frameworks: LangChain vs LangGraph vs LlamaIndex vs Haystack vs DSPy" (Cem Dilmegani, Ekrem Sarı, 2025)

프로젝트 필요성

기존 시스템의 한계

템플릿 기반 응답 구조
미리 정해진 질문-답변 형식에 묶여 있어 새로운 질문에는 적절한 답변을 제공하지 못함

맥락 기반 답변 부족
이어지는 대화에서 일관된 답변을 제공하지 못해 사용자 혼란 초래

링크 중심 응답 방식
사용자가 질문에 대한 직접적인 답을 얻기 어렵고, 추가 탐색 부담 발생

정보 정확도 문제
연결된 링크 내용과 실제 질문에 대한 정보가 일치하지 않을 가능성이 존재

동똑이가 해결할 수 있는 문제

유연한 자연어 이해
템플릿에 구애받지 않고 다양한 질문 형태를 이해하여 보다 자연스러운 대화 가능

맥락 기반 답변 제공
이어지는 대화에서도 일관되고 관련성 높은 답변을 제공

정확한 정보 매칭
원하는 정보에 대한 직접적인 답을 제공하여 추가 탐색 필요 없음

출처 인용 시스템
답변에 출처를 포함하여 정보의 정확도와 신뢰성 보장

학생 편의성 향상

시간 절약: 추가 탐색 없이 필요한 정보 즉시 확보

이용 편의: 친숙한 대화형 인터페이스로 정보 접근

정보 신뢰: 정확하고 최신화된 정보로 신뢰성 확보

개발 목표

챗봇 시스템 개발 총 목표

동국대학교 재학생들이 학교의 다양한 정보와 학과별 맞춤 정보를 보다 편리하고 빠르게 접근할 수 있도록, **생성형 AI 기술을 활용한 '동똑이'를 오픈 소스를 이용하여 개발**하는 것입니다.

자연어 이해 능력 향상

사용자의 자연어 질문에 적절한 답변을 생성할 수 있는 AI 모델을 개발합니다.

현재 시스템 목표 시스템

하이브리드 검색 시스템 구축

하이브리드 검색 구조를 통해, 정확하고 빠른 정보를 제공 가능하도록 합니다.

현재 시스템 목표 시스템

학과별 맞춤형 답변 기능

각 학과별 정보에 특화된 LLM 모델을 통해, 정확한 학과 관련 정보를 제공할 수 있도록 합니다.

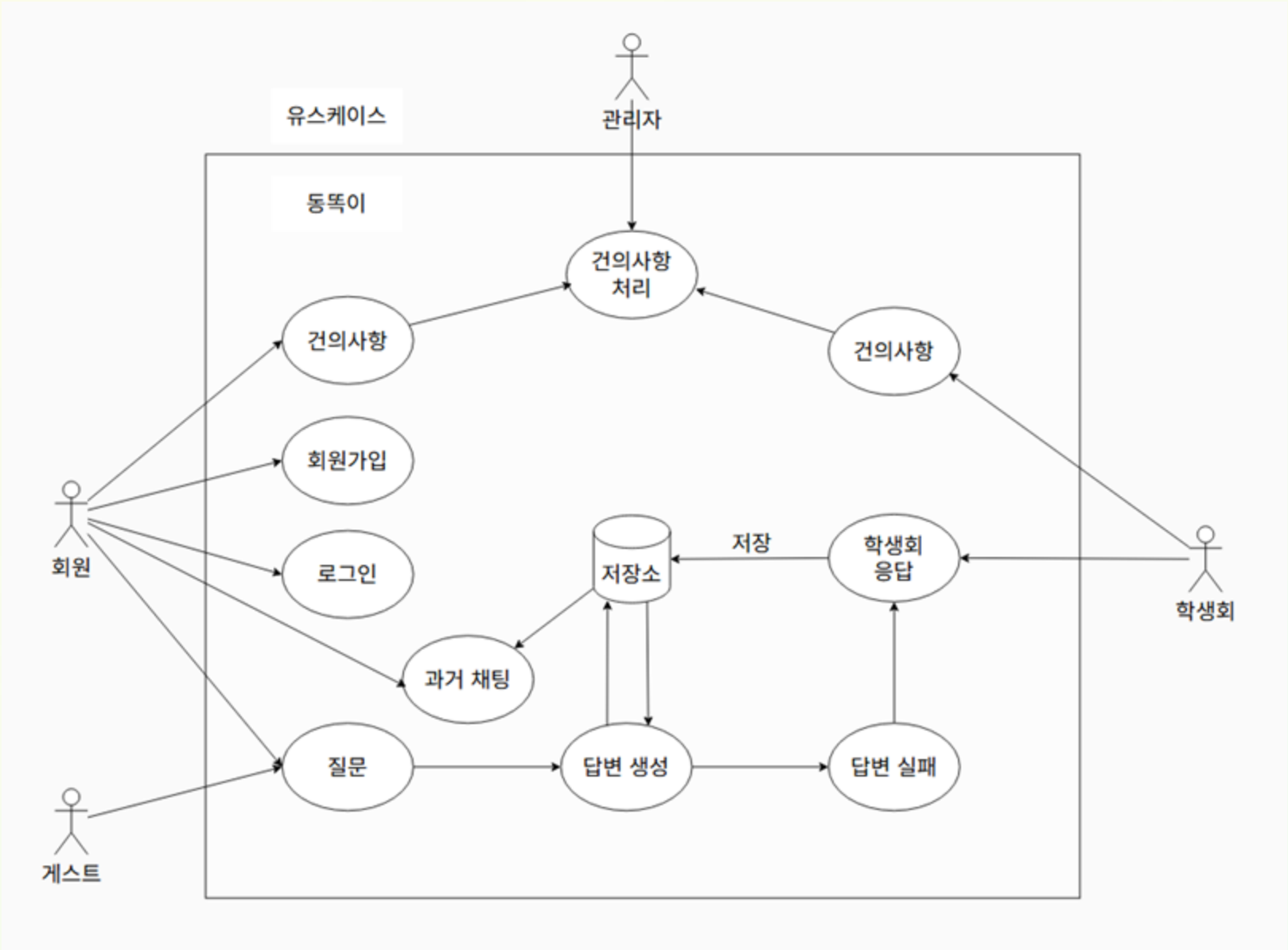
현재 시스템 목표 시스템

최신 정보 제공 시스템

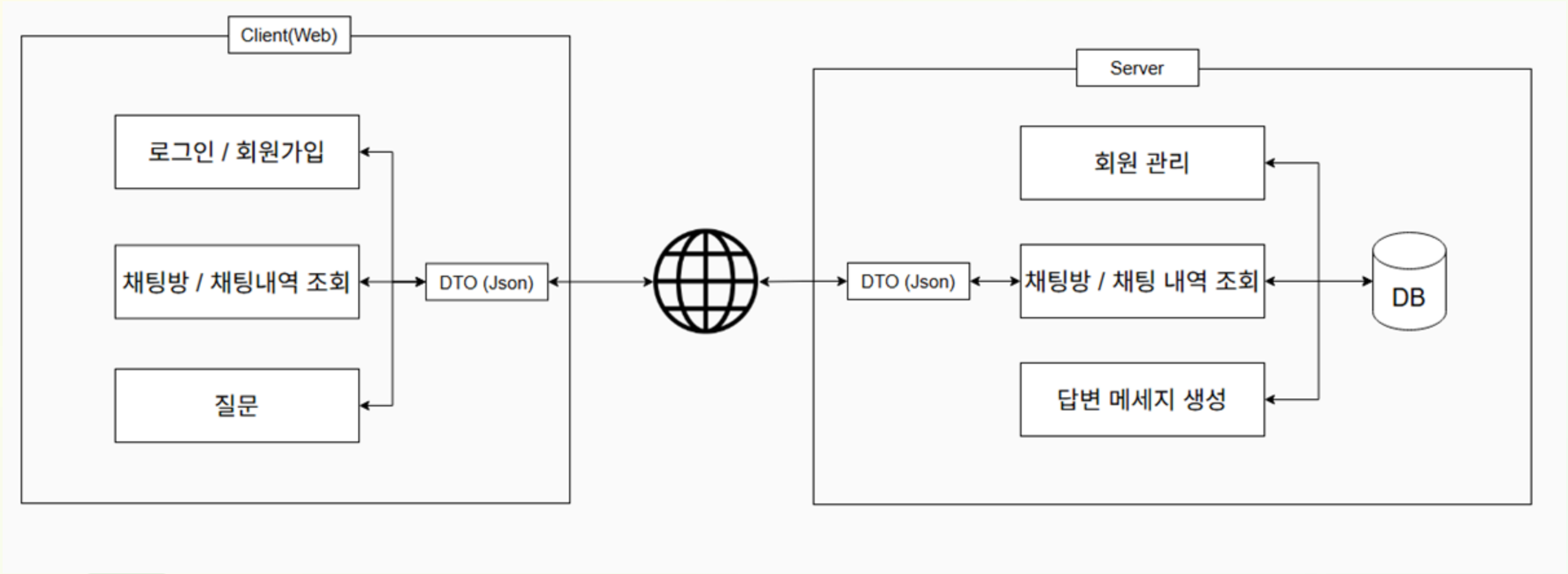
학교 공지사항을 주기적으로 수집/정제하여, 사용자가 최신 정보에 접근할 수 있도록 합니다.

현재 시스템 목표 시스템

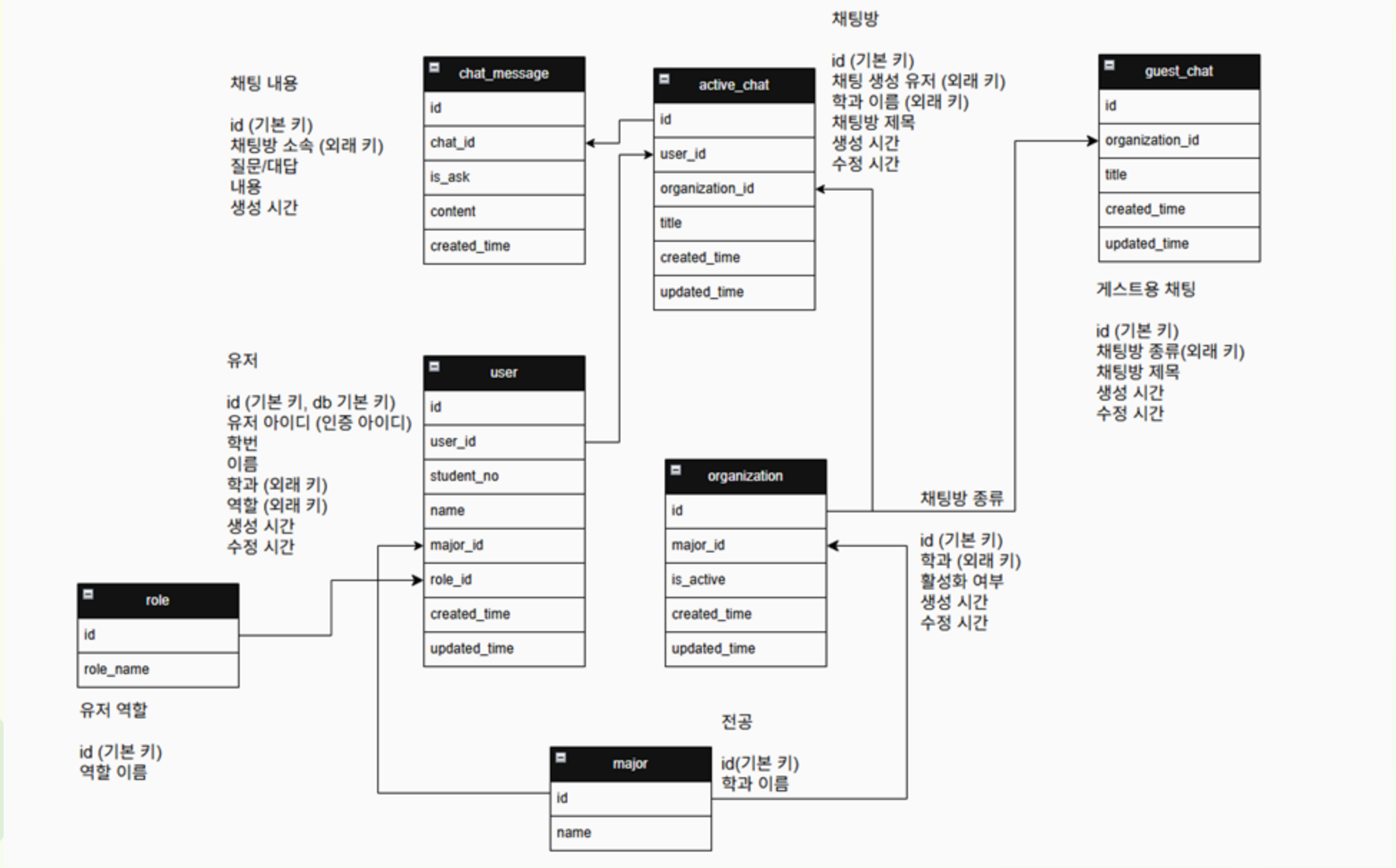
유스케이스


유스케이스
관리자
동똑이
건의사항 처리
건의사항
건의사항
회원가입
로그인
저장
저장소
학생회 응답
회원
학생회
과거 채팅
질문
답변 생성
답변 실패
게스트

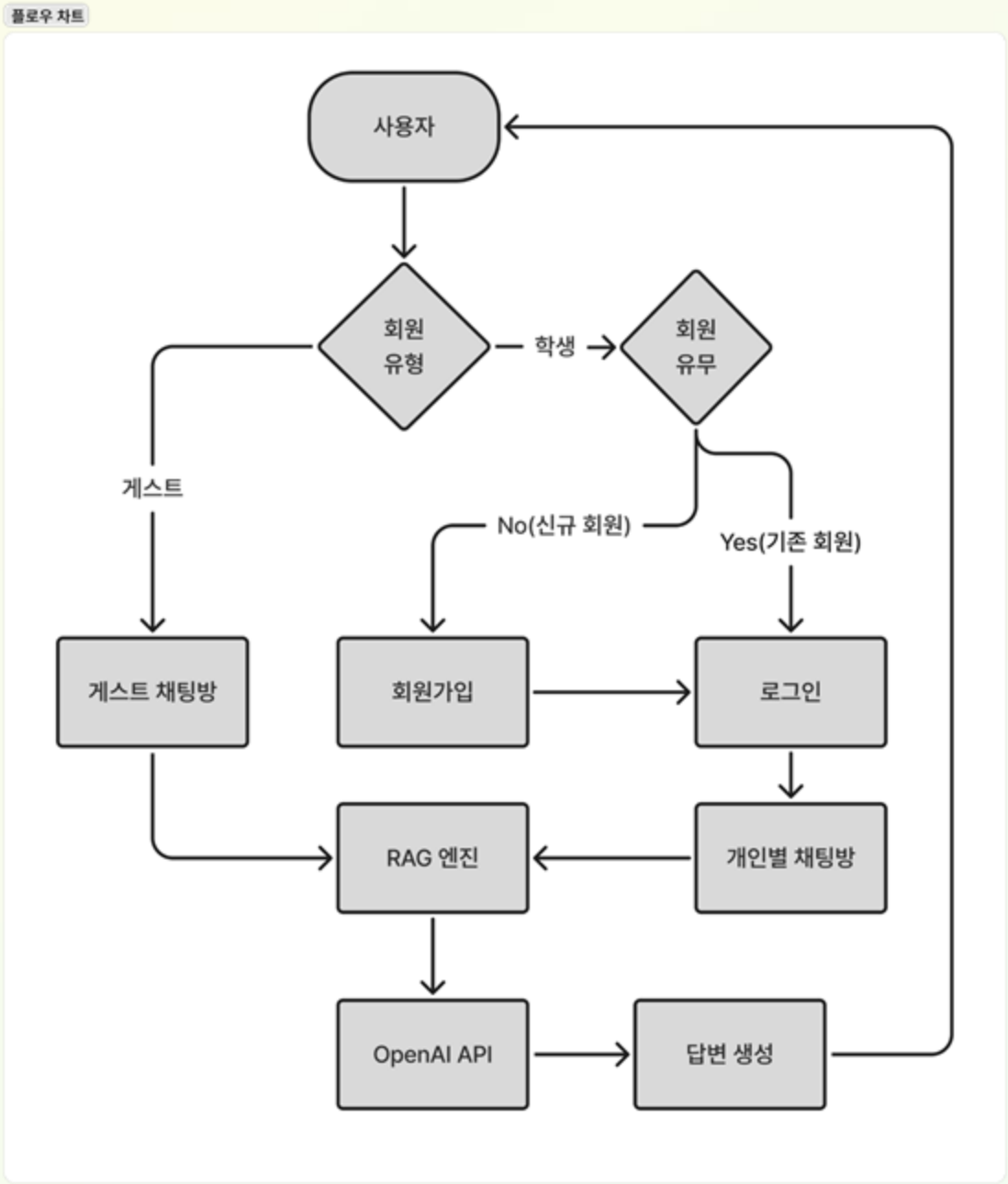
시스템 블록 다이어그램


Client(Web)
로그인 / 회원가입
채팅방 / 채팅내역 조회
질문
DTO (Json)
Server
회원 관리
DTO (Json)
채팅방 / 채팅 내역 조회
DB
답변 메세지 생성

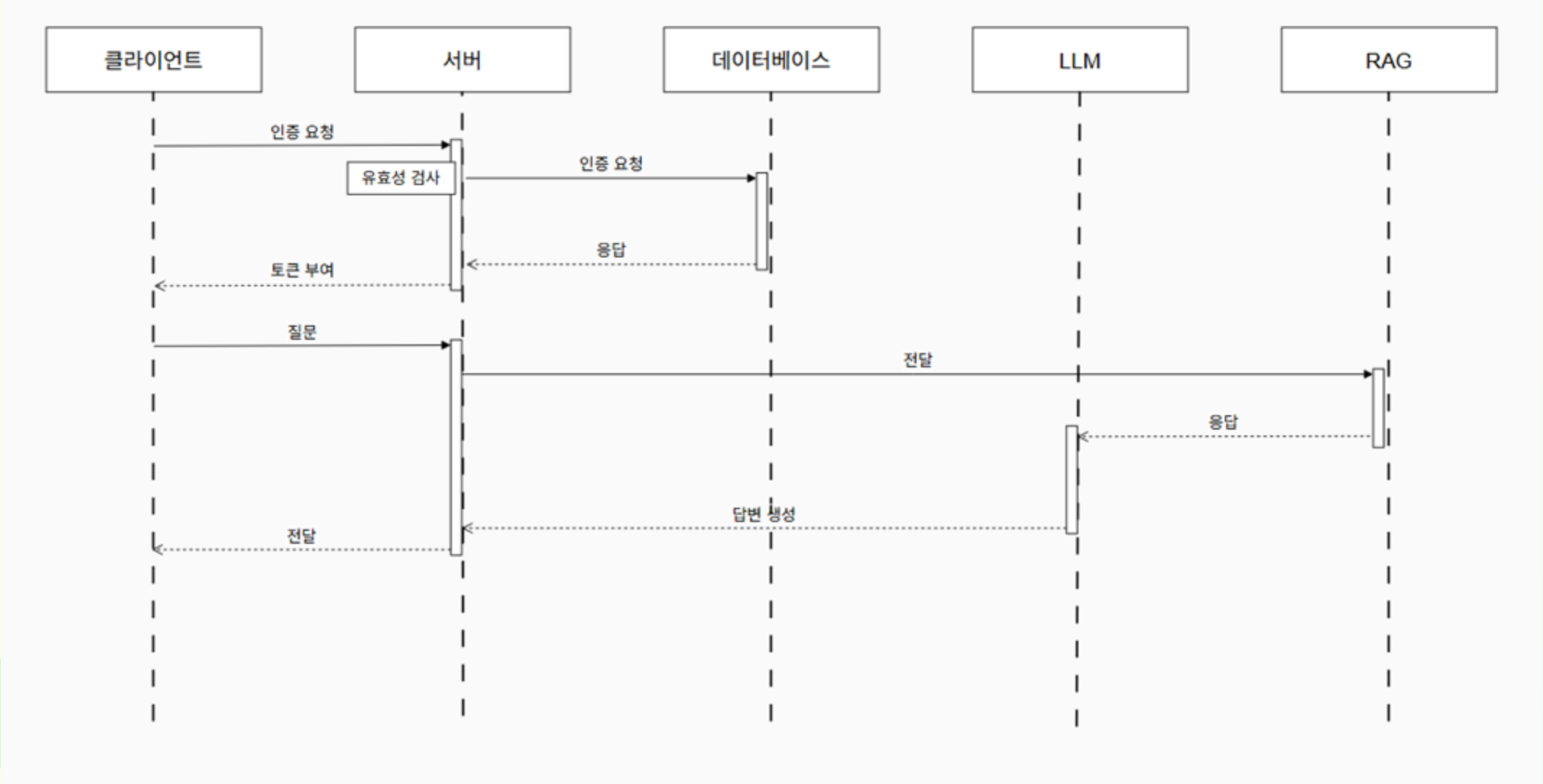
ERD


채팅방
id (기본 키)
채팅 생성 유저 (외래 키)
학과 이름 (외래 키)
채팅방 제목
생성 시간
수정 시간
chat_message
id
chat_id
is_ask
content
created_time
채팅 내용
id (기본 키)
채팅방 소속 (외래 키)
질문/대답
내용
생성 시간
active_chat
id
user_id
organization_id
title
created_time
updated_time
guest_chat
id
organization_id
title
created_time
updated_time
게스트용 채팅
id (기본 키)
채팅방 종류(외래 키)
채팅방 제목
생성 시간
수정 시간
유저
id (기본 키, db 기본 키)
유저 아이디 (인증 아이디)
학번
이름
학과 (외래 키)
역할 (외래 키)
생성 시간
수정 시간
user
id
user_id
student_no
name
major_id
role_id
created_time
updated_time
organization
id
major_id
is_active
created_time
updated_time
채팅방 종류
id (기본 키)
학과 (외래 키)
활성화 여부
생성 시간
수정 시간
role
id
role_name
유저 역할
id (기본 키)
역할 이름
전공
major
id
name
id(기본 키)
학과 이름

Flow Chart


플로우 차트
사용자
회원 유형
학생
회원 유무
게스트
No(신규 회원)
Yes(기존 회원)
게스트 채팅방
회원가입
로그인
RAG 엔진
개인별 채팅방
OpenAI API
답변 생성

시퀀스 다이어그램


클라이언트
서버
데이터베이스
LLM
RAG
인증 요청
유효성 검사
인증 요청
응답
토큰 부여
질문
전달
응답
답변 생성
전달

핵심 기능 소개

동똑이는 학생들의 정보 탐색을 원활하게 하기 위해 다음과 같은 핵심 기능들을 제공합니다. 각 기능은 사용자 경험을 향상시키고, 정확한 정보를 빠르게 제공하기 위해 설계되었습니다.

학교 맞춤 답변 생성

동국대학교의 내규, 학과별 정보, 부서별 업무 내용 등을 기반으로 한 정확한 답변을 생성합니다. 사용자가 자연어로 질문하면, 학교의 공식 정보를 바탕으로 적절한 답변을 제공합니다.

- 출처 인용 및 검증 가능

자연어 처리

템플릿 기반 응답 구조가 아닌, 고급 자연어 처리 기능으로 복잡한 질문이나 맥락 있는 대화에도 적절히 응답합니다. 사용자의 의도를 정확히 파악하여 유연한 대화를 가능하게 합니다.

- 맥락 기반 상호작용

최신 정보 제공

학교 공지사항을 주기적으로 수집/정제하여 최신 정보를 빠르게 제공합니다. 답변 생성 시점에서 가장 최근 데이터를 기반으로 하여 정확성을 보장합니다.

- 1일 4회 자동 업데이트

데이터 수집 → 임베딩 처리 → 하이브리드 검색 → LLM 응답 생성

데이터 수집 및 전처리

동국대학교 공지사항을 주기적으로 크롤링하여 정제된 텍스트로 변환함으로써, 학생들이 최신이고 정확한 정보에 접근할 수 있도록 합니다.

데이터 수집

동국대학교 공지사항을 주기적으로 크롤링합니다.

공지 제목, 게시일, 게시판 종류 메타데이터 확보

불용어 처리

HTML 태그와 조사 등을 처리합니다.

보안 및 데이터 정화

문단 분할

원문을 문단 단위로 분할하여 의미 단위로 나눕니다.

의미 있는 청크 단위로 분할

최종 텍스트

정제된 텍스트로 변환되어 모델 입력으로 사용됩니다.

의미 있는 텍스트만 유지

전처리 과정의 중요성

- 원하지 않는 HTML 태그, 스타일, 스크립트 등이 제거되어 데이터 정화
- 문단 단위 분할로 의미 있는 청크로 나누어 검색 정확도 향상
- 모델이 처리하기 쉬운 형태로 텍스트 정제

사용되는 기술

- Python Requests & BeautifulSoup을 활용한 크롤링
- 한국어 불용어 처리 라이브러리
- Huggingface/KURE-v1 모델을 활용한 임베딩

하이브리드 검색 시스템

하이브리드 검색 시스템이란?

의미 기반 검색과 키워드 기반 검색의 장점을 결합하여, **보다 정확하고 관련도 높은** 검색 결과를 제공하기 위해 KURE-v1 임베딩과 TF-IDF 기반 검색 시스템을 통합하였습니다.

임베딩 기반 검색

HuggingFace의 nlpai-lab/KURE-v1 모델을 이용해 문단을 벡터화하여 의미적 유사성을 기반으로 검색

TF-IDF 키워드 검색

문서 내 키워드의 중요도를 가중치로 표현하여 키워드 기반으로 검색 수행

하이브리드 결합 로직

임베딩과 TF-IDF 결과에 α 가중치를 적용하여 상위 K개의 문단을 반환

구현 방식

- 각 문단은 KURE-v1 모델을 통해 벡터화
- 동일 문서에 대해 TF-IDF 행렬 생성
- α 가중치 기반으로 상위 K개 문단 반환

하이브리드 접근의 장점

- 의미적 이해와 키워드 매칭의 결합
- 복잡한 자연어 질문에도 높은 정확도
- 출처 인용 및 검증을 위한 메타데이터 포함

질문 분류 및 라우팅

동국대학교 챗봇은 사용자의 질문에 가장 적절한 정보를 제공하기 위해 **질문 분류 및 라우팅 시스템**을 구축했습니다. 이 시스템은 질문의 내용을 분석하여 적절한 데이터 소스로 라우팅함으로써 정확하고 빠른 정보 제공을 가능하게 합니다.

질문 입력
사용자로부터 자연어 질문을 입력받습니다

→

TF-IDF 벡터화
질문 텍스트를 TF-IDF 벡터로 변환합니다

→

로지스틱 회귀
벡터화된 질문을 분류하여 적절한 데이터셋과 게시판을 선택합니다

→

라우팅
선택된 데이터소스로부터 질문에 적합한 답변을 생성합니다

데이터셋 분류 모델

- 모든 데이터셋의 제목과 본문을 합친 텍스트를 입력으로 사용
- TF-IDF + 로지스틱 회귀 모델을 기반으로 참조할 데이터를 선정
- 질문의 내용과 가장 유사한 데이터셋을 자동으로 분류

게시판 분류 모델

- 청크 텍스트와 게시판명 데이터를 이용하여 임베딩
- 로지스틱 회귀 모델을 기반으로 참조할 게시판을 선정
- 질문 유형에 맞춘 적절한 부서별 정보 제공

응답 생성 및 인용 처리

응답 생성 흐름

사용자 질문 → 관련 내용 추출 → GPT-4o-mini API → 자연어 응답

본문 내용 기반 답변 생성

프롬프트 템플릿에 질문/컨텍스트/출력 포맷이 포함된

- 선택된 문단을 컨텍스트로 삽입하여 자연어 답변 생성
- 답변 포맷은 일정, 문의처, 참조 링크 등 행정정보 항목 중심
- 맥락을 이해하여 이어지는 대화에서도 일관된 답변 제공

출처 자동 인용 처리

자동 인용 시스템으로 출처 신뢰성 보장

- 문단 메타데이터를 활용해 출력
- 출처는 주제, 게시일, 게시판 URL 포함한 구조화된 정보

기술적 제약사항

API 비용 제약

- ChatGPT API 사용에 따른 비용 부담이 존재
- 모델 호출 빈도와 데이터 사용량이 증가할 수록 비용 증가

해결 방안

학교 측의 지원금을 최대한 활용

서버 자원 제약

- 벡터 임베딩 및 대규모 문서 저장 공간 부족
- 학교나 개인 서버 환경을 고려한 저사양 환경에서도 동작 가능해야 함

해결 방안

벡터 DB Chroma는 일정 기간별로 갱신/압축 관리가 필요

데이터 수집 제약

- 학교 홈페이지 구조 변경 시 크롤러가 오작동하거나 데이터 수집이 중단될 수 있음

해결 방안

구조 개편 사항 인지 후 그에 맞게 유동적으로 크롤링 코드 변경

기대 효과

재학생

- 신입생의 학교 적응 및 커리큘럼 로드맵 구성에 용이
- 일관된 정보 습득으로 정보 혼란 방지
- 정보 획득 시간 감소

학교

- 대학교 위상 상승
- 단순 문의 전화 감소로 인한 행정 효율 상승

동국대학교 재학생 및 교직원의 일 효율성 및 편의성 상승

구현 계획 및 일정

작업	담당	9월		10월				11월			
		4주차	5주차	1주차	2주차	3주차	4주차	1주차	2주차	3주차	4주차
아이디어 기획											
프로세스 정의	조준용										
사용자 흐름 구체화	신원철, 육심호										
인프라 설계											
데이터베이스 구축, 오픈소스 API 조사	조준용										
화면 UI/UX 설계	육심호										
와이어프레임 작성	신원철										
기능 구현											
백엔드 개발	조준용										
프론트엔드 개발	육심호										
RAG 개발	신원철										
마무리											
기능별 테스트	신원철, 육심호										
웹 배포	조준용										

성과 창출 계획

동똑이 프로젝트의 결과물을 공개하고 향후 확장을 통해 더 많은 학생들과 교육 현장에서 가치를 제공하기 위해 다음과 같은 계획을 세우고 있습니다.

GitHub 공개

- 예상 마감일: 12/2
- 소스 코드 공개
- 오픈소스 커뮤니티 기여

소프트웨어 등록

- 예상 마감일: 12/2
- SW 등록 저널에 등록
- 품질 인증 획득

향후 확장 계획

- **기능 확장**
  캘린더 연동, 파일 첨부 기능 추가 등
- **앱스토어 등록**
  모바일 환경에서도 편리한 이용 가능
- **데이터베이스 확대**
  교내 모든 정보 시스템 연동
- **특허 출원**
  独创적 기술에 대한 특허 보호

동국대학교 학생들을 위한 지속적인 정보 제공 서비스를 목표로 합니다

감사합니다