**프로젝트명:** 시스템 아키텍처

**버전:** v0.9

**작성일:** 2024년 12월 30일

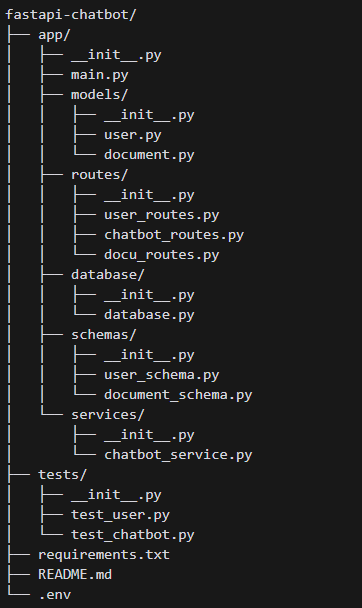
**작성자:** 김문수

1. 개요

본 문서는 LLM 기반 문서 검색 시스템의 아키텍처를 설명합니다. 이 시스템은 토큰 기반 인증, 동적 메뉴, 내부 문서 벡터화, LangChain 및 LangGraph를 활용한 챗봇, 사용자 관리 기능 등을 포함합니다.

2. 시스템 구성

2.1 시스템 구조도



2.1 프론트엔드

- 기술 스택: jquery.js

- 주요 기능:

- 사용자 인터페이스 제공

- 토큰 기반 로그인/로그아웃

- 동적 메뉴 렌더링

- 챗봇 인터페이스

2.2 백엔드

- 기술 스택: Python (FastAPI)

- 주요 컴포넌트:

1. 인증 서비스

- JWT 토큰 발급 및 검증

- 로그인/로그아웃 처리

2. 사용자 관리 서비스

- 사용자 CRUD 작업

- 권한 관리

3. 문서 처리 서비스

- 내부 문서 벡터화

- Vector DB 연동

4. 챗봇 엔진

- LangChain 및 LangGraph 활용

- 외부 자원 연동 Agent 구현

2.3 데이터베이스

1. 관계형 DB (PostgreSQL)

- 사용자 정보 저장

- 시스템 설정 관리

2. Vector DB (Pinecone)

- 문서 벡터 저장

- 유사도 기반 검색 지원

2.4 외부 서비스

- LLM API (예: OpenAI GPT)

- 외부 자원 API (예: 기업 내부 시스템, 공개 API 등)

3. 주요 기능 상세

3.1 인증 시스템

- JWT 기반 토큰 인증

- 토큰 발급, 갱신, 폐기 로직

- 보안 강화를 위한 짧은 만료 시간 및 리프레시 토큰 사용

3.2 동적 메뉴

- 사용자 권한에 따른 메뉴 구성

- 백엔드에서 메뉴 구조 및 권한 정보 제공

- 프론트엔드에서 동적 렌더링

3.3 문서 처리 및 검색

1. 문서 처리 파이프라인:

- 문서 업로드 → 텍스트 추출 → 청크 분할 → 벡터화 → Vector DB 저장

2. 검색 프로세스:

- 쿼리 벡터화 → Vector DB 유사도 검색 → 관련 문서 추출

3.4 LLM 기반 챗봇

- LangChain을 활용한 대화 처리 파이프라인 구축

- LangGraph로 복잡한 대화 흐름 관리

- RAG(Retrieval-Augmented Generation) 기법 적용

3.5 사용자 관리

- 사용자 등록, 조회, 수정, 삭제 기능

- 역할 기반 접근 제어(RBAC) 구현

- 관리자 전용 대시보드 제공

4. 보안 고려사항

- 모든 통신에 HTTPS 적용

- API 요청에 대한 인증 및 권한 검증

- SQL 인젝션, XSS 등 웹 취약점 대비

- 정기적인 보안 감사 및 업데이트

5. 배포 전략

- Docker 컨테이너화