홍승완

www.hosugator.com

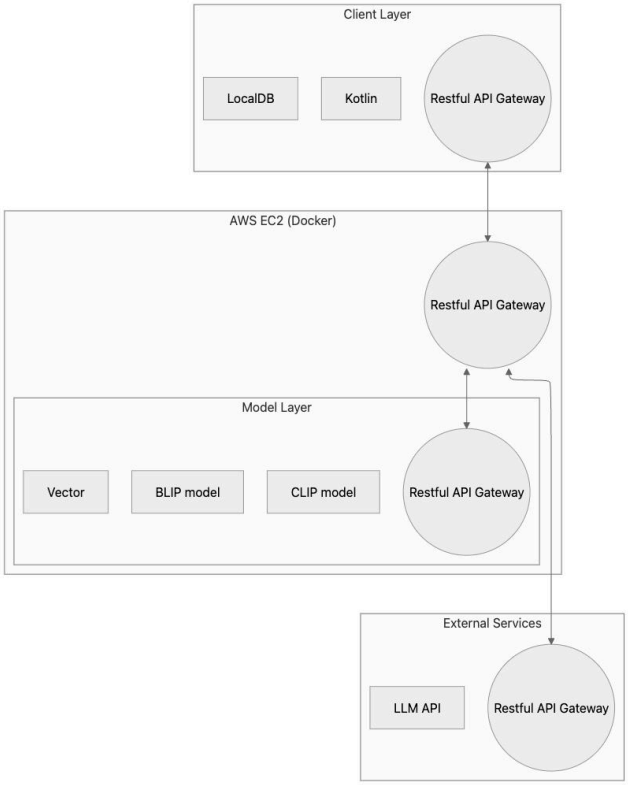
소담일기

시각 장애인을 위한 사용자 음성 기반 사진 해설 다이어리

(2025 한국장애인해커톤 본선 진행 중)

시스템 아키텍처

초기 MVP 개발 및 서비스 효율성을 위해 모놀리식 아키텍처 채택
Django 프레임워크를 FastAPI의 비동기 라우터로 포팅하도록 고도화
Docker를 사용해 통합 환경을 구축 및 AWS EC2 인스턴스로 최종 배포



핵심 기능

다중 모델 파이프라인 기반 이미지 분석
BLIP 모델을 사용하여 이미지에 대한 구체적이고 사실적인 묘사
CLIP 모델과 30여개의 사전 정의된 분위기 텍스트를 활용하여 유사도 기반으로 이미지 분위기 Top-3 추출 및 할당

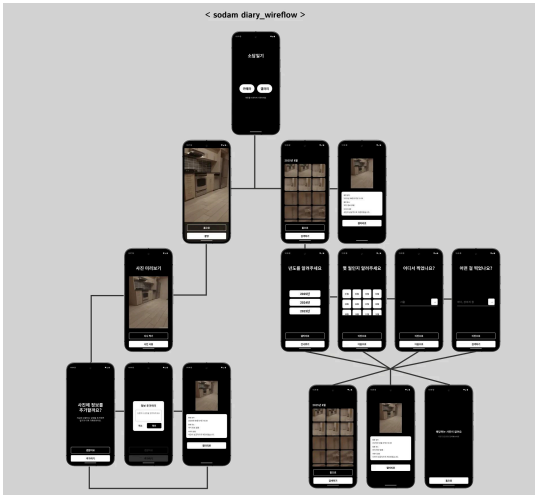
RAG 기반 개인화 해설
STT 기반 사용자 음성 입력을 받아 핵심 키워드 추출
추출된 정보 기반 LLM 프롬프트 동적 구성하는 Prompt Engineering
RAG 시스템 적용하여 객체를 데이터로 태깅한 자연스러운 문장으로 반환

비동기 처리 및 추론 최적화
다중 모델 추론 과정을 병렬 스레드로 분리 처리
다중 모델 로드 시 양자화 적용하여 메모리/응답시간 최적화
지연 최소화 및 성능 향상 위해, OpenVINO 적용하여 추론 최적화

성능 분석

비용 절감
LLM (GPT-4V) 단독 사용 시 발생하는 높은 비용(5만 건 응답 기준 월 약 1,300,000원) 발생
이를 오픈소스 모델을 활용한 다중 모델 구조를 채택하여 총 비용을 30% 이상 절감하였으며,
LLM 호출이 증가할수록 비용 절감 효과 극대화

성능 최적화
LLM 단독 사용 시 평균 30초에 달하던 응답 시간을 다중 모델 처리 구조로 개선하여 응답 시간을 20% 단축
이를 다시 비동기 처리 로직으로 개선하여 응답 시간을 추가 30% 단축
다양한 모델들을 테스트하고 최종적으로 성능/비용 측면에서 최적의 양자화 모델 채택하여 응답 시간 최적화 및 서버 비용 절감



회고

시각 장애인 사용자 계층을 대상으로 하는 공익 서비스 개발에 참여하며, 실제 스타트업 환경처럼 '사용자 경험'과 '사업성(비용 효율성)'이라는 두 가지 목표를 동시에 달성하기 위한 도전을 주도했습니다. 특히, 높은 비용 문제를 인식하고 다중 모델 구조를 제안하고 구현하여 비즈니스적 문제를 기술적으로 해결한 경험이 가장 의미 있었습니다.
향후 고도화 전략은 아래와 같습니다.
1. 배포 환경 안정화: AWS EC2 서버 연동 및 Docker 기반의 안정적인 프로덕션 환경 구축(완료)
2. 개인화 시스템 고도화: DB 구조를 개선하여 이전에 입력된 사용자 정보와 유사한 객체(강아지, 사람)가 탐지될 경우 이름을 추천하는 개인화 태깅 추천 시스템 구현 (진행 중)
3. 성능 최적화 완료: OpenVINO 통합을 통해 모델 추론 성능을 극대화하여 시각 장애인의 음성 안내 지연 시간을 최소화 (진행 중)
최종적으로 고도화를 완료하여 해커톤 우승 및 배포를 통한 장애인 분들의 사진에 대한 접근성 향상 및 피드백을 통한 개선을 목표로 하고 진행 중에 있습니다.

Figma Jetpack Compose Hugging Face django Kotlin asyncio POSTMAN