실증적SW개발프로젝트 주간보고 (10주차)

작성일: 2025/05/11 팀명: chAt & me

| 팀 활동  보고 | 활동일시 | 2025.05.11 |
| --- | --- | --- |
| 장 소 | 온라인 |
| 참석자 | 이채영, 김나영, 박수빈, 원민 |
| 특이사항 | - |
| 이번주  진행사항 | 1. 개발내용   * TAG RAG 개발 시작 * TAG 팀: 1) Nougat를 활용한 PDF 테이블 추출 및 LaTeX 포맷 파싱   2) 추출된 테이블을 LLM 기반 요약 후 Vector DB에 저장 시도  3) Ollama 기반 LangChain RAG 파이프라인으로 질의 응답 구현  4) PDF 테이블 자동 추출 및 CSV 저장 기능 추가   * RAG 팀: 1) Llama Parse를 활용한 문서 파싱 코드 구현   2) OpenAI api 사용해 pdf 파싱 코드 구현 시도(실패) -> Hugging Face 사용  3) 파싱 결과 기반으로 Llama Index에 저장해 RAG 시스템 구성  4) LangChain활용 Vector DB 구축  2. 팀원별 활동내용   * 이채영(팀장) : 개인 학습, Llama Parse 활용한 pdf 파싱 코드 작성 * 김나영(팀원, 프론트엔드담당) : 개인 학습, RAG+TAG 통합 파이프라인 구현 코드 작성 * 박수빈(팀원, 백엔드 담당) : 개인 학습, Nougat를 활용한 table RAG 실습 코드 작성 * 원 민(팀원, 인공지능 담당) : 개인 학습, Vector DB와 RAG 파이프라인 구축(미완성) | |
| 개발계획 대비 진행현황 |  | |
| 다음주  계획 | * TAG 팀:  1. pdf -> csv 변환시 가독성 문제 해결 2. 코드 최적화 3. 읽은 csv 파일 LLM이 학습할 수 있는 형태로 만들기  * RAG 팀:  1. Vector DB와 RAG 파이프라인 구축 완성시키기 2. LLM 응답 포맷 불안정 문제 해결을 위한 출력 정제 로직 적용 3. 코드 최적화 | |
| 주요 결과물 | pipeline\_TAG.py  Nougat\_TAG.py  LlamaParse\_RAG.py  VectorDB\_pipeline.py | |