

실증적SW개발프로젝트 주간보고 (6주차)

작성일: 2025/04/14 팀명: CivilWord

팀 활동 보고	활 동 일 시	2025-04-07 ~ 2025-04-14
	장 소	비대면 진행
	참 석 자	양한나, 배준영, 임세희, 임지은, 최지은
	특 이 사 항	없음
이번주 진행사항	1. 백엔드(양한나)	
	1) 로그인 구현 내역 <ul style="list-style-type: none">- 인증 방식: OAuth2 Password Flow (폼 기반 인증)- 사용자 입력: Swagger Authorize에서 username, password 입력- 비밀번호 처리: bcrypt로 해시 저장 및 passlib로 검증- JWT 토큰 발급: 사용자 인증 후 access_token 반환- Swagger 연동: OAuth2PasswordRequestForm 도입으로 Swagger Authorize 버튼 연동- 응답 스키마: Token 스키마 사용 - access_token + token_type	
	2) 라우터 구현 내역 <ul style="list-style-type: none">- /register: 신규 사용자 등록, 회원가입 시 유효성 검사 및 DB 저장 처리- /login: OAuth2 기반 로그인 처리 및 JWT 토큰 발급- /user_info: 사용자 정보 입력, 수정, 조회 기능- /complaint: 민원 등록 기능(사용자 ID와 제목/내용 기반 POST 처리), LLM 응답 기능 (더미 테스트), 유사답변 검색 기능(테스트 구현)	
	2. 데이터 크롤링(임세희, 최지은)	
	1) 사하구 민원 데이터 전체 크롤링 <ul style="list-style-type: none">- 약 15000개 데이터 수집 완료- checkpoint와 backup을 통해 코드 강화- xlsx 파일을 csv와 json파일로 생성	
	2) 데이터 크롤링 코드 통합 및 속도 개선 <ul style="list-style-type: none">- 각자 작성한 데이터 크롤링 코드 테스트 후 대면 회의에서 속도 및 오류 등 결과 점검.- 'headless', 'webdriver_manager'와 'sleep' 줄이기 등 코드 정리를 통한 속도 개선- 결과: 사하구 공개 민원 크롤링 시간 8시간 -> 3시간 단축 - 속도 향상을 위해 request, BeautifulSoup 사용을 고려하였지만 페이지네이션이 URL 파라미터로 되어있지 않아 selenium 사용을 결정.- 멀티 스레딩 등 다른 방법들도 강구하였지만 개발환경의 RAM용량이 작아 더 이상 속도 향상 불가능.- 부산시 소재의 다른 구들의 민원은 사하구정보다 적기 때문에 이 정도로 타협.	
	3) PostgreSQL, WSL 환경 공부 <ul style="list-style-type: none">- PostgreSQL 패키지 설치 및 기본 지식 공부- 소량의 데이터로 우선 연결	
	4) 데이터 전처리 형식 및 DB 테이블 설정 <ul style="list-style-type: none">- LLM팀의 테스트 후 회의 필요.	
	3. LLM(배준영, 임지은)	
	- Mistral AI 파인튜닝 가이드 페이지로 파인튜닝 이론 학습	
	- Mistral 7B 파인튜닝 예제 검색 후 공부	
	- 데이터 크롤링된 csv 파일을 jsonl 파일로 변환 후 파인 튜닝 학습 시도 -> 메모리 부족으로 실패	

다음주
계획

1. 백엔드(양한나)

- 현재 사용중인 DB(SQLite) -> PostgreSQL로 교체
- 전체 라우터 기능 점검
- 관리자 전용 라우터 개발

2. 데이터 크롤링 및 엔지니어링(임세희, 최지은)

- LLM팀의 테스트 후 데이터 전처리
- DB 테이블 설정 및 연결
- 다른 구 데이터 크롤링
- 기능명세서를 따라 우선순위에 맞게 프론트 개발
- 중간 발표 준비

3. LLM(배준영, 임지은)

- S03-209호 or Colab Pro 버전을 사용하여 파인 튜닝 및 학습