

실증적SW개발프로젝트 주간보고 (1주차)

작성일: 2025/05/04 팀명: CivilWord

팀 활동 보고	활동 일시	2025.04.28 ~ 2025.05.04
	장소	2025.04.28: S06, 비대면 진행
	참석자	양한나, 최지은, 임지은, 배준영, 임세희
	특이사항	없음
<p>0. 팀 공통 진행 사항</p> <ol style="list-style-type: none">백엔드 연동 환경설정 설명사하구청 인터뷰 질문 정리학습 방향 관련 회의 <p>1. 백엔드(양한나)</p> <ol style="list-style-type: none">진행 상황<ul style="list-style-type: none">모델 연동: Polyglot 기반 LLM (polyglot_base)을 로컬에 다운로드하고, FastAPI 백엔드에 통합FastAPI API 구현: /generate_answer 엔드포인트 구현 (POST 요청으로 민원 텍스트 입력 시 LLM 응답 반환)모델 경량화 경로 설계: Tokenizer 및 Model을 글로벌 변수로 캐싱하고 최초 요청 시 로딩되도록 처리Swagger 연동 확인: localhost:8000/docs에서 Swagger UI로 테스트 가능하도록 설정 완료주요 이슈<ul style="list-style-type: none">모델 로딩 시 느림/무한 대기: CPU 환경(WLS2)에서 모델 로딩 및 추론이 매우 느림 -> GPU 사용해야 함 <p>2. 프론트엔드(임세희, 최지은)</p> <ol style="list-style-type: none">백엔드 연동 환경설정<ul style="list-style-type: none">WSL 환경에서 가상환경 생성 및 PostgreSQL 설치GITHUB README 참고와이어프레임 수정<ul style="list-style-type: none">민원목록에서 선택 기능 추가엑셀로 취합해서 다운받기 기능 추가데이터 공개 여부 삭제텍스트 input 삭제엑셀 파일 양식 제공 추가프론트 기본 화면 개발 파트 배분		

3. 데이터 크롤링(임세희, 최지은)

1) 사하구청 대면 인터뷰

- 자원환경과 민원 담당자 인터뷰
- 민원여권과 민원 분류 담당자 인터뷰

2) 크롤러 성능 검토 및 데이터 상태 분석

- 2시간 (~20분)
- 두 팀원의 상이한 결과 데이터 확인
- 페이지마다 처리한 데이터 개수 확인하는 기능 추가 후 데이터 누락 여부 확인
- 사하구청의 새울전자민원창구의 공개 데이터와 대조
- 접수: 94, 취하: 21, [관리자에 의해 삭제]: 36, 다부처병렬: 163 => 314(일일이 계산한 결과라서 오차 있을 수 있음.)
- 총 민원 데이터 15147건 - 누락된 데이터(답변완료, 이송이첩 외의 데이터) 314 = 14833 => 크롤링 데이터 14834로 계산 오차 감안하면 크롤링 잘 된듯함.
- "다부처병렬"이라는 예상 외의 데이터 확인 => 처리 방법 회의 필요.

4. LLM(배준영, 임지은)

- 실습실 컴퓨터로 mistral 7B와 kllum3 모델 계속해서 돌리기(mistral 7B 컴퓨터 업데이트로 인해 또 꺼진 것 같음. 업데이트도 늦췄기에 이제는 꺼지면 안됨. 만약 꺼지면 그냥 아예 다른 컴퓨터에서 할 예정)

```
OUTPUT DEBUG CONSOLE TERMINAL PORTS PROBLEMS

{'loss': 0.0757, 'grad_norm': nan, 'learning_rate': 0.00015489021956087827, 'epoch': 0.84}
{'loss': 2.3825, 'grad_norm': nan, 'learning_rate': 0.00015489021956087827, 'epoch': 0.85}
{'loss': 51.0067, 'grad_norm': nan, 'learning_rate': 0.00015489021956087827, 'epoch': 0.86}
{'loss': 0.1288, 'grad_norm': nan, 'learning_rate': 0.00015489021956087827, 'epoch': 0.86}
{'loss': 8.9634, 'grad_norm': nan, 'learning_rate': 0.00015489021956087827, 'epoch': 0.87}
{'loss': 0.0, 'grad_norm': nan, 'learning_rate': 0.00015489021956087827, 'epoch': 0.88}
{'loss': 2.421, 'grad_norm': nan, 'learning_rate': 0.00015489021956087827, 'epoch': 0.88}
{'loss': 1.5959, 'grad_norm': nan, 'learning_rate': 0.00015489021956087827, 'epoch': 0.89}
{'loss': 0.2618, 'grad_norm': nan, 'learning_rate': 0.00015489021956087827, 'epoch': 0.9}
{'loss': 0.3393, 'grad_norm': nan, 'learning_rate': 0.00015489021956087827, 'epoch': 0.9}
{'loss': 9.906, 'grad_norm': nan, 'learning_rate': 0.00015489021956087827, 'epoch': 0.91}
{'loss': 0.1566, 'grad_norm': nan, 'learning_rate': 0.00015489021956087827, 'epoch': 0.92}
31% | 1384/4509 [243:02:23<588:14:19, 677.65s/it]
```

<그림 1> kllum3 학습 진행 상황

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

{'loss': 1.6614, 'grad_norm': 0.8478355407714844, 'learning_rate': 0.0002, 'num_tokens': 569324.0, 'mean_token_accuracy': 0.5989007840553919, 'epoch': 0.08}
{'loss': 1.6375, 'grad_norm': 1.1019433736801147, 'learning_rate': 0.0002, 'num_tokens': 621346.0, 'mean_token_accuracy': 0.6064088682333628, 'epoch': 0.09}
{'loss': 1.6347, 'grad_norm': 1.2851381301879883, 'learning_rate': 0.0002, 'num_tokens': 674130.0, 'mean_token_accuracy': 0.6064990232388179, 'epoch': 0.1}
{'loss': 1.6238, 'grad_norm': 1.027258038520813, 'learning_rate': 0.0002, 'num_tokens': 725573.0, 'mean_token_accuracy': 0.6078069428602855, 'epoch': 0.1}
{'loss': 1.5882, 'grad_norm': 0.70557200008660089, 'learning_rate': 0.0002, 'num_tokens': 780744.0, 'mean_token_accuracy': 0.6153690656026204, 'epoch': 0.11}
{'loss': 1.6462, 'grad_norm': 0.9708796739578247, 'learning_rate': 0.0002, 'num_tokens': 834566.0, 'mean_token_accuracy': 0.6025013665358225, 'epoch': 0.12}
6% | 456/7516 [3:31:16<54:17:40, 27.69s/it]
```

<그림 2> mistral 7B 학습 진행 상황

- polyglot 모델을 위한 데이터셋 전처리 -> 파인튜닝한 polyglot 모델의 답변을 보니 부족한 게 많았음 => 민원내용과 답변내용의 핵심만 남기기로 결정

(현재 진행 상황: 링크, 첨부파일 및 붙임 제거)

답변 내용의 핵심만 남기려 했으나 전처리 하기에는 다양한 케이스의 답변이 있어서 다음주 내로 LLM 팀 회의 예정

<그림 3> dataset 전처리 파일 사진

0. 팀 공통 진행계획

- 프론트 팀과 백엔드 팀의 연동 회의
 - 사하구청 대면 인터뷰 기반으로 팀 전체 회의

1. 백엔드(양한나)

- 응답 JSON 정형화: 현재는 단순 문자열 → { "answer": "...", "elapsed": "...초", 등 }
 - GPU 환경에서 서버 실행 테스트

다음주 계획

2. 프론트엔드(임세희, 최지운)

- 프론트 구현 -> 메인 파일 삽입, 답변 생성 페이지

3. 데이터 크롤링 및 엔지니어링(임세희, 최지은)

- #### - '다부처 병렬' 데이터 처리 방법 논의

4. LLM(배준영, 임지은)

- dataset 답변 내용 서식 정규화
 - 정규화된 답변 양식으로 다시 polyglot fine tuning
 - mistral 7B, kullm3 계속해서 학습