



2025학년도 1학기 SW 캡스톤디자인 경진대회

LLM 기반 통계 분석 및 보고서 생성 서비스

오픈소스 URL : <https://github.com/Ddaaan/capston>

팀명 단호박

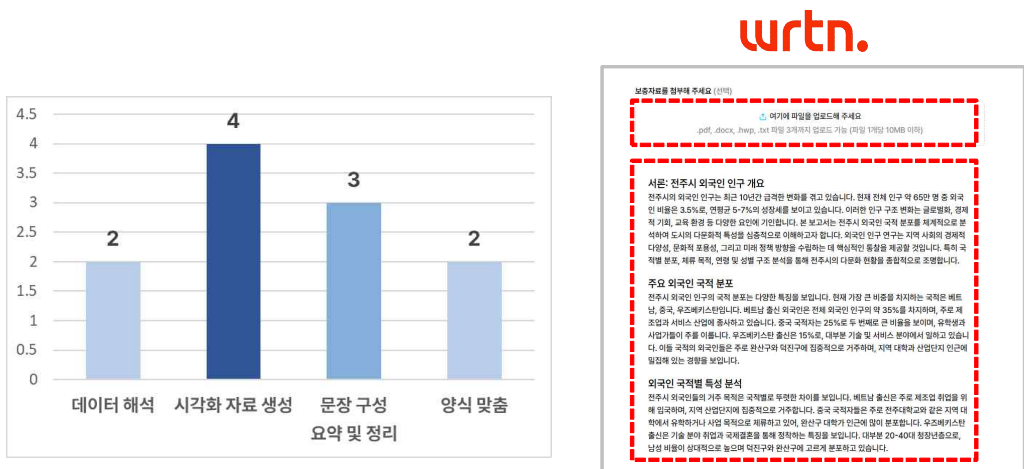
지도교수 김윤경

팀원 김현수(통계학과 4학년), 이재연(IT정보공학과, 4학년), 최다은(IT정보공학과, 4학년)

산업체 (주) 선도소프트

개발 동기 및 목적

개발 동기



- 분석 보고서의 경우 데이터 분석 및 시각화 등 많은 시간 소요
- 분석 보고서 작성 시 시각화 자료 생성과 문장 구성 요약 등에 가장 많은 어려움을 호소
- 기존 자동화 도구의 경우 한글(HWPX) 환경에서 활용이 어려움
- 텍스트 출력 위주, 표, 그래프 깨짐 등의 한계 존재

개발 목적

- LLM 기반 데이터 분석부터 HWPX 보고서까지 자동 생성하는 서비스를 개발하여 전문 지식 없이도 간단 요청만으로 한글 보고서 생성이 가능하도록 구현

주요 기술

전체 시스템은 사용자 흐름에 따라 크게 네 단계로 구성되며, 각 단계에는 다양한 기술이 적용되어 자동화와 정확도를 높였다.

〈활용 데이터 검색〉

- KURE-v1 임베딩 모델 사용으로 한국어 데이터에 최적화
- FAISS를 이용한 유사도 기반 검색으로 오탈자도 검색 가능
- 정규화 처리로 특수문자/띄어쓰기 상관없이 검색 가능
- 1차 후보군 생성 → 2차 재검색 방식으로 정확한 검색 결과 제공

〈데이터 전처리 결과〉

- 엑셀 파일 업로드 시 자동 전처리
- 병합 셀 보정 및 다중 헤더 분리
- 전처리 후 CSV 파일로 자동 저장되어 모델 분석 최적화

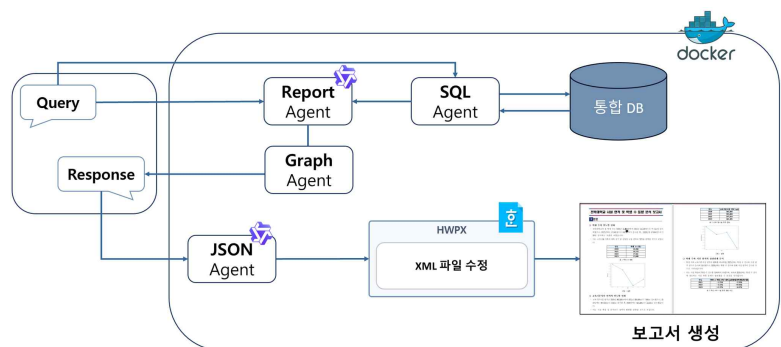
〈보고서 분석 결과〉

- LLM이 자동으로 SQL 쿼리 생성해 필요한 데이터 추출
- 추출한 데이터를 기반으로 분석 보고서 작성
- 표와 그래프 자동 생성, 그래프 유형은 가이드 기준으로 선택

〈한글보고서 생성〉

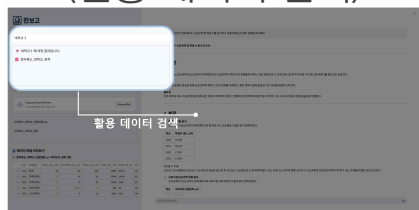
- 분석 결과를 JSON 구조로 변환
- JSON을 기반으로 HWPX(XML) 구조 수정
- 표, 그래프, 스타일이 자동 적용된 정형 한글 보고서 생성
- 양식 선택 기능 제공 (일반/전북대 전용)

개발 내용



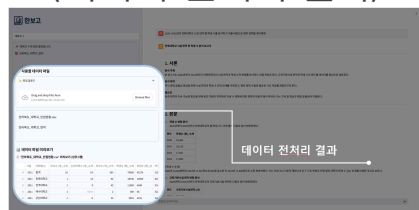
전체 시스템의 아키텍처는 위 그림과 같다.
사용자 흐름과 기능에 따라 네 가지 주요 단계로 구분할 수 있다.

〈활용 데이터 검색〉



1. 사용자는 웹 플랫폼을 통해 활용 데이터 검색 가능

〈데이터 전처리 결과〉



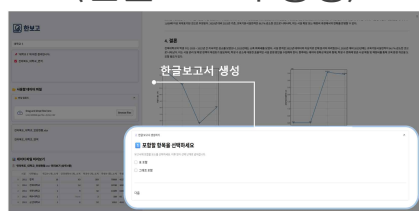
2. 업로드 된 엑셀(xlsx) 파일 자동 전처리를 통해 CSV 형태로 변환

〈보고서 분석 결과〉



3. 사용자의 요청에 따라 데이터 분석 결과 자동 생성

〈한글보고서 생성〉



4. 분석된 내용을 기반으로 한글(HWPX) 형식의 정형화된 보고서 파일 제공

결과 및 분석

LLM 기반 통계 분석 및 자동 보고서 생성 플랫폼인 ‘한보고’ 서비스를 개발하였다.

프로젝트 결과



- 데이터 검색 → 분석 → 결과 시각화 → 보고서 생성 일괄 처리
- 양식에 맞는 내용 요약 및 삽입을 통한 완성도 높은 보고서 제공

활용방안

- 보고서 자동 생성에 따른 업무 효율 증대
- 기관별 맞춤형 양식 지원 가능
- 로컬 LLM 기반으로 내부망 및 폐쇄망 환경에서도 실행 가능