

의약품 기본 정보 수집 및 데이터베이스 적재

- 식약처 공공 의약품 API를 활용하여 의약품 기본 정보를 수집한다.
- API 응답 데이터를 데이터베이스 스키마에 맞게 컬럼 단위로 매핑한다.
- 약품 고유 코드(code)를 기준으로 중복 데이터를 제거한다.
- API에서 제공되지 않는 항목은 엑셀 데이터를 활용하여 보완한다.
- 기존 데이터베이스에 저장된 약품 코드와 비교하여 신규 데이터만 선별적으로 저장한다.
- 정제된 데이터를 drug 테이블에 적재한다.

이미지 CSV 기반 의약품 이미지 정보 갱신

- 의약품 이미지 CSV 파일에서 품목명, 제조사, 이미지 URL 정보를 추출한다.
- 데이터베이스에 저장된 약품 정보와 품목명 및 제조사를 기준으로 매칭한다.
- 이미지 URL이 존재하는 데이터만 대상으로 하여 갱신 작업을 수행한다.
- 매칭된 데이터에 대해 drug 테이블의 image_url 컬럼을 업데이트한다.
- 이미지 데이터 처리를 기본 정보 수집 로직과 분리하여 관리한다.

의약품 상세 정보 수집을 위한 웹 크롤링 모듈

- 식약처 의약품 상세 정보 페이지를 대상으로 웹 크롤링을 수행한다.
- 약품 코드를 기준으로 상세 페이지에 접근한다.
- 페이지 내 텍스트 정보를 분석하여 성상, 첨가제, 원료약품 및 분량 정보를 추출한다.
- 크롤링 기능을 함수 단위로 구현하여 재사용 가능하도록 구성한다.
- 공공 API에서 제공되지 않는 상세 정보를 보완하기 위한 목적으로 사용한다.

구조 설계

- 데이터 수집, 저장, 보완 단계를 명확히 분리하여 설계
- 약품 코드 중심의 데이터 관리로 중복 및 불일치를 방지
- 재실행 시에도 데이터 무결성이 유지되도록 구성