

job-finder

소개

- 프로젝트 명: job-finder

구인 게시글이 플랫폼별로 다르게 올라옴

취업하려는 사람이 한 눈에 원하는 조건에 맞는 정보를 전부 볼 수 없음

→ 크롤링과 파싱을 통해 등록된 여러 플랫폼의 구직 게시글을 하나로 통합하여 파일로 저장
(로켓펀치, 잡플래닛 등)

개발 주요 기능

구직 게시글 크롤링 및 파싱

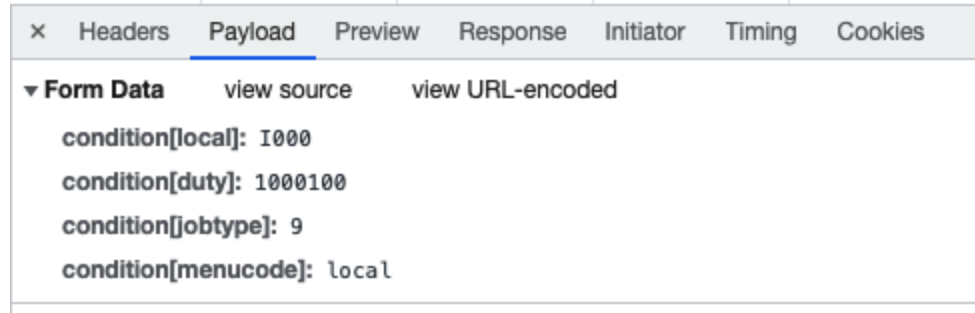
- Json으로 상세 검색 데이터 받아서 다운받는 방식
 - 크롤링 request 보내면 RestController로 받아 response
 - response 방식은 다운 여부에 따라 달라질듯
 - 로컬로 파일 받는 경우 성공 여부
 - 데이터 전송시 데이터
- Jsoup을 이용한 html 크롤링
 - 왜 Jsoup?

- html파싱에 가장 유명함(오라클에서도 Jsoup을 html 파싱 라이브러리로 소개함 [link](#))
 - 이전에 사용해본 경험이 있음
 - 기본적으로 공개 게시판 들어가서 내용 크롤링
 - 잡코리아
 - requestBody에 상세 검색 데이터 있음
 - 상세 내용 매핑 정보 따라함 → 종류가 플랫폼마다 다르고 너무 많기 때문에 공통적인 부분으로 최대한 간추려서 크롤링
 - 근무지역의 경우 구 단위 까지만 크롤링 → 더 자세하게 가면 너무 많아짐(서울만 12만개)
 - 학력, 기업형태x
 - 직무 플랫폼별 통일화 필요
 - <https://www.jobkorea.co.kr/recruit/joblist?menucode=local&localorder=1>
body:
condition%5Blocal%5D=l000%2CK000&condition%5Bmenucode%5D=local
 - 예시

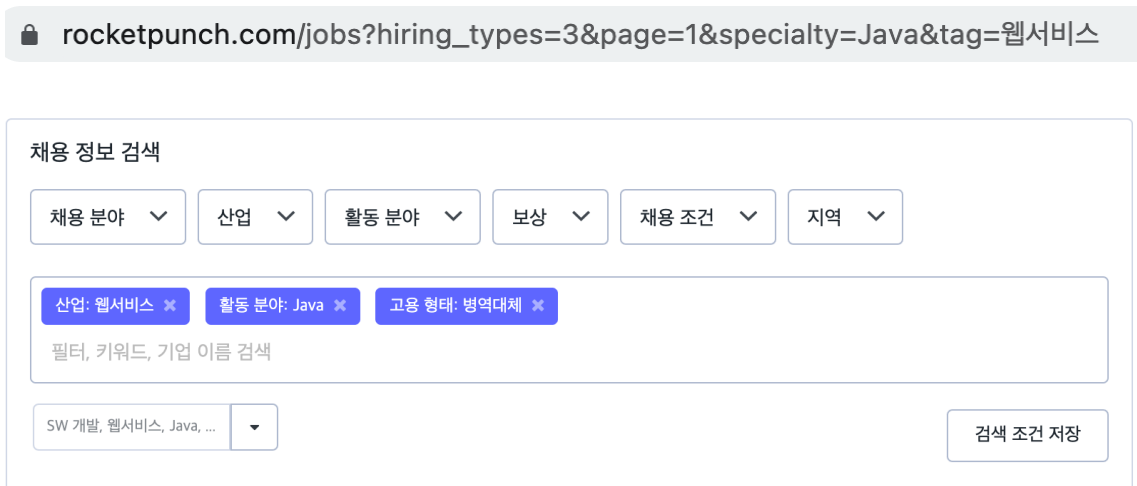
채용공고상세검색 최근 검색한 조건 2건 ▾

직무	근무지역	경력	학력	기업형태	고용형태
IT-인터넷	서울특별시	신입	대학교졸업(4년)	대기업	정규직
디자인	경기도	1~3년	대학교졸업(2,3년)	30대그룹사	도급
무역/유통	인천광역시	4~6년	대학원 석사졸업	매출1000대기업	프리랜서
영업-고객상담	대전광역시	7~9년	대학원 박사졸업	중견기업	아르바이트
서비스	세종특별자치시	10~15년	고등학교졸업	강소기업	연수생/교육생
연구개발-설계	충청남도	16~20년	학력무관	외국계기업	병역특례
					위촉직/개인사업자
산업	직급/직책/급여	우대전공	자격증	우대조건	복리후생

서울 ✕ 웹프로그래머 ✕ 병역특례 ✕
초기화 ☆ 조건저장
선택된 조건 검색하기



- 로켓펀치
 - url에 상세 검색 내용 들어가있음



- 추후 추가

csv 전환

전환 방식

1. BufferedWriter 사용
 - a. 장점: 외부 라이브러리에 의존하지 않음
 - b. 단점: CSV 형식을 스스로 맞춰줘야함
2. 오픈소스(OpenCSV) 사용

- a. 장점: 1의 단점이 사라짐
- b. 단점: 라이브러리 사용 방법을 익혀야함, 외부 라이브러리 의존도가 높아짐

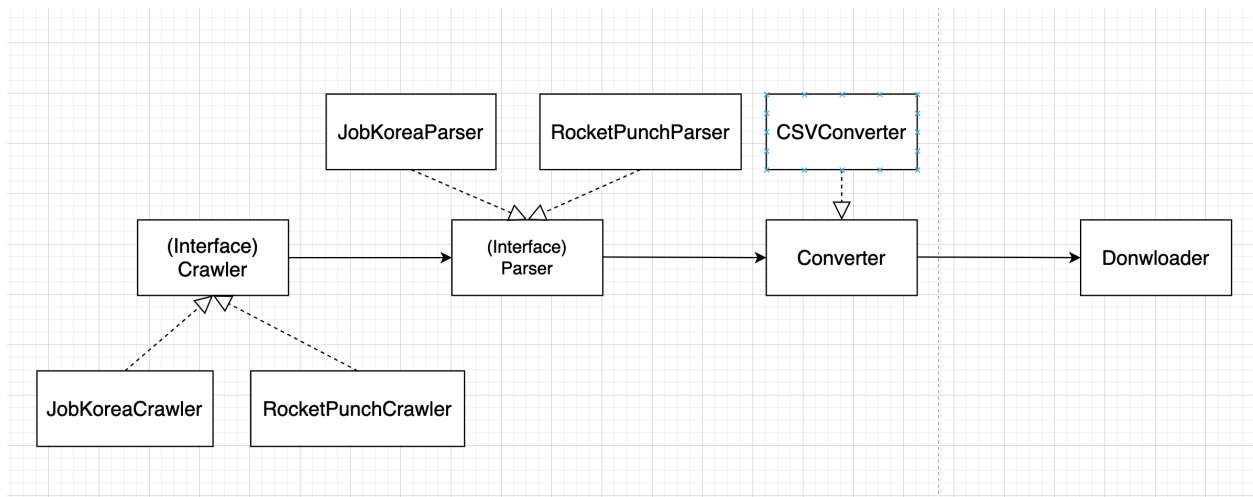
내용

- 게시글 제목
- url
- 상세 정보
 - 근무지역
 - 연봉
 - 기술 스택
 - 고용형태
 - 경력
 - 직무
 - etc

파일 다운로드

- Java Files 이용(로컬 이용시)
 - Files, nio에 대한 공부 필요

클래스 다이어그램



- crawler는 추가 가능

개발 요구사항

- IDE: IntelliJ
- JDK 17
- framework: spring boot 2.6.5
- DB: 미정 or 필요 x
- OOP관점으로 확장에 열린 코드
- 크롤링 라이브러리: Jsoup
- 테스트: JUnit4 or 5 ?
- linux 서버 세팅
- ci / cd

사용 라이브러리

- JUnit(4, 5 미정)
- lombok
- Jsoup
- OpenCSV(미정)

? 생각해볼 점