

채용 정보 분석 플랫폼 - 데이터 출처 정리표

출처 플랫폼	URL	수집 방법	대상 데이터 항목
사람인 (Saramin)	https://www.saramin.co.kr	웹 크롤링 (BeautifulSoup) + 정규표현식	- 포지션명 - 회사명 - 지역 - 연봉 - 경력 조건 - 기술 키워드(본문 기반 추출) - 공고 등록일/마감일
원티드 (Wanted)	https://www.wanted.co.kr	웹 크롤링 (Selenium) 또는 일부 REST API 활용	- 직무명 - 회사명 - 근무 지역 - 연봉 정보 (있을 경우) - 경력 요건 - 요구 기술 스택 (태그 기반) - 공고 게시일

잡코리아 (Jobkorea) <https://www.jobkorea.co.kr>

웹 크롤링 (Selenium)

- 직무명

+ HTML 파싱

- 회사명

- 연봉 범위

- 경력 구분

- 근무지역

- 공고 내용 내
기술 키워드

로켓펀치 (Rocketpunch) <https://www.rocketpunch.com>

웹 크롤링 (BeautifulSoup)

- 포지션명

+ 태그 분석

- 회사명

- 고용형태

- 기술 스택 (태그
구조 명확)

- 지역

- 공고 등록일

수집 항목 요약 분류

공통 기본 항목

- 공고 ID / URL
- 포지션명 (직무명)
- 회사명
- 근무 지역
- 고용형태 (정규직, 계약직 등)
- 연봉 (범위 or 평균)
- 경력 조건 (신입/경력)
- 공고 등록일 / 마감일

분석용 항목

- 기술 키워드 (태그 or 본문 분석 기반)

- 기술 키워드 빈도
- 직무 카테고리
- 산업군 (로켓펀치 기준, 스타트업 중심)
- 지역별 채용 분포
- 연봉 통계

참고사항

- 모든 데이터는 공개된 채용 공고 기반으로 수집
- **robots.txt** 및 플랫폼 정책 준수
- 공고 내용에서 기술 스택은 정규화 및 텍스트 분석을 통해 추출
- API 제공 시 법적으로 안정적이며, 우선 활용 권장