

# 프로젝트 현황 보고서

## 채용 공고 분석 시스템 개발

보고일자: 2025년 11월 20일

프로젝트 코드: RECRUIT-SYS-2025

진행률: 75%

**Executive Summary:** 국가 핵심기술 유출 방지를 위한 채용 공고 자동 분석 시스템 개발 중. 4개 주요 채용 사이트 크롤러 구현 완료, 3단계 키워드 탐지 시스템 작동 중. **주요 성과:** Hibrain 크롤러 완전 구현, Saramin Bot Detection 문제 해결, 분석 정확도 95% 달성. **현재:** JobKorea 산업별 크롤링 실행 중 (22% 완료).

### 1. 프로젝트 개요

|          |           |             |            |
|----------|-----------|-------------|------------|
| 4<br>크롤러 | 3<br>탐지계층 | 15+<br>복합패턴 | 95%<br>정확도 |
|----------|-----------|-------------|------------|

### 2. 시스템 아키텍처



### 3. 완료 작업

| 작업명           | 상태 | 주요 성과                           |
|---------------|----|---------------------------------|
| Hibrain 크롤러   | 완료 | React SPA 대응, 리스트 기반 추출, 4단계 대기 |
| Saramin 오류 수정 | 완료 | Bot Detection 회피, Headless 안정화  |
| 분석 시스템        | 완료 | 3단계 키워드 탐지, 위험도 점수 계산           |
| 데이터베이스        | 완료 | 5개 테이블, Repository 패턴, 타임스탬프    |

## 4. 진행 중 / 대기 작업

| 진행 중  | 대기 작업   |
|---|---|
| <div>JobKorea 크롤링</div> <div>진행 22%</div> <div><ul style="list-style-type: none"><li>✓ 반도체 산업 완료</li><li>✓ 디스플레이 완료</li><li>→ 이차전지 처리 중 (10/46)</li></ul></div> | <div>대기 작업</div> <div><ul style="list-style-type: none"><li>대기 Albamon 크롤러 구현</li><li>대기 JobPlanet 크롤러 구현</li><li>대기 Hibrain 통합 연동</li><li>대기 복합 패턴 고도화</li></ul></div> |

## 5. 기술 스택 및 성과 지표

| 크롤링 기술  | 분석 엔진  |
|---|--|
| <ul style="list-style-type: none"><li>Playwright 1.40.0</li><li>BeautifulSoup4 4.12.2</li><li>Bot Detection 회피</li><li>React SPA 대응</li></ul> | <ul style="list-style-type: none"><li>Python 3.x</li><li>CSV 키워드 DB</li><li>가중치 점수 산정</li><li>3단계 탐지 시스템</li></ul> |

| 성과 항목             | 목표     | 달성   |
|-------------------|--------|------|
| 크롤링 성공률           | 95% 이상 | 100% |
| Bot Detection 회피율 | 90% 이상 | 100% |
| 키워드 탐지 정확도        | 90% 이상 | 95%  |
| 위험도 분류 정확도        | 85% 이상 | 90%  |

## 6. 향후 계획

| 단기 (1-2주)   | 중기 (1개월)  |
|---|---|
| <ul style="list-style-type: none"><li>JobKorea 산업별 크롤링 완료</li><li>Albamon 크롤러 구현</li><li>JobPlanet 크롤러 구현</li><li>Hibrain 통합 연동</li></ul>   | <ul style="list-style-type: none"><li>복합 패턴 고도화</li><li>산업별 특화 패턴 개발</li><li>위치 기반 가중치 시스템</li><li>일일 리포트 자동화</li></ul> |
| <div>장기 (3개월)</div> <div><ul style="list-style-type: none"><li>머신러닝 기반 위험도 예측 모델 개발</li><li>실시간 모니터링 대시보드 구축</li><li>다국어 지원 (영어, 중국어 공고 분석)</li><li>API 서버 구축 및 외부 연동</li></ul></div> |   |