

ICT Internship 개발 과제

인턴 과제 안내

목적

본 과제는 인턴 기간에 실제 진행 할 업무인, OCR 텍스트 파싱과 동적 크롤링에 대한 문제 해결력과 코드 품질을 확인하기 위한 사전 과제입니다.

1) 과제 공통 안내

- 과제는 총 2가지 문제로 구성되어 있습니다.
- 아래 2개 과제 중 **1개만 선택하여 제출해도 됩니다.**
- **2개 모두 제출 시 우대하지만, 완성도 높은 1개를 더 선호합니다.**
- AI 도구 사용은 허용 및 권장합니다. 단, 제출 시점에 본인이 작성한 코드의 동작 원리, 설계 의도, 트레이드오프를 명확히 설명할 수 있어야 합니다.
- 구조적으로 갖춘 코드(모듈화/리팩토링/테스트/문서화)를 우대합니다.
 - 작성 후, 유지보수가 꾸준히 이루어져야 할 프로젝트로 가정하고 리팩토링을 진행하세요.
 - 테스트 코드 작성을 권장합니다.
 - 간단한 설계 문서를 작성을 권장합니다.
 - 프로세스 진행과정, 결과 등 로그 작성을 권장합니다.
 - 문서는 마크다운 형태로 작성을 권장합니다.

2) 과제 안내

과제A. OCR → Parsing (계근지/영수증 텍스트 파싱)

배경

계근지(차량의 총중량/공차/실중량 등을 포함한 영수증 형태)의 OCR 결과 텍스트를 입력으로 받아, 업무에 필요한 필드를 정확하고 견고하게 파싱하여 구조화합니다.

핵심 요구

- OCR 텍스트의 노이즈(띄어쓰기/오탈자/순서 변경/라벨 누락/숫자 포맷 불규칙 등)를 고려한 파싱 설계
 - 직접 파싱 로직을 구현해도 좋습니다.
 - 기존에 작성된 오픈소스 파싱 엔진을 활용해도 좋습니다.

- 예시1. spacy, <https://github.com/explosion/spaCy>
- 무게/날짜 등 주요 값의 정규화 및 검증 로직 포함
- 출력은 JSON/CSV 등 구조화 형태로 저장

참조

샘플 OCR 데이터를 파일로 첨부합니다.

과제B. 크롤링 (입찰공고 수집)

배경

동적으로 렌더링되는 입찰공고 페이지에서 목록 및 상세 정보를 수집하여 표준 스키마로 저장합니다.

핵심 요구

- 크롤링 대상 - 누리장터
 - **접근 가능 브라우저 안내**
 - 누리장터 > 입찰공고 > 입찰공고 목록
- 목록 → 상세 진입 후, 핵심 필드 추출
 - 가능한 범위 내 모든 필드
- 안정적인 동적 크롤링 코드 작성
 - 대기 전략, 예외 처리, 재시도, 중복 방지, 키워드 처리 등
 - “제품 수준” 관점의 구현을 우대
 - 예시1. 오류로 중단된 뒤 재실행 시 마지막 지점부터 이어서 수집
 - 예시2. interval/cron 등 실행 모드 구성제공

3) 제출물(공통)

아래 항목을 포함해 제출해 주세요.

- 저장소 링크(GitHub 등) 또는 압축 파일(zip)
- README.md 필수
 - 실행 방법(로컬 재현 가능하도록)
 - 의존성/환경(버전 포함)
 - 설계 및 주요 가정
 - 한계 및 개선 아이디어
- 결과 산출물 예시 파일(JSON/CSV 등)

4) 평가 기준(공통)

- 정확성: 요구 필드가 정확하게 추출/정규화되는가
 - 견고함: 예외/노이즈/변형 케이스에 강한가
 - 재현성: 누구나 쉽게 실행하여 동일 결과를 얻을 수 있는가
 - 코드 품질: 구조화(모듈화), 가독성, 리팩토링, 확장성
 - 문서/테스트: 실행 가이드와 최소한의 검증 체계가 있는가
-

5) 일정 및 제출 안내

- **제출 기한:** 2026.02.09(월) (EOD, 23:59까지)
 - **제출 방식:** 이메일 제출
 - 제출처(e-mail): dev.interview+ict@recokr.com
 - 메일 제목: [2026 ICT] 리코 과제 제출_성함
 - 메일 내용/첨부: 3) 제출물에 해당하는 결과물
-

6) 문의 / 기타

- 과제 관련 질문은 dev.interview+ict@recokr.com 으로 메일 주시면 확인 후 답변드리겠습니다.
- 기한 내 과제 제출하신 분들 중, 면접 대상자에 한하여 인터뷰 일정 등 진행 예정입니다.
(위치: 삼성역 부근, 대면 인터뷰 예정)