

SK네트웍스 Family AI과정 5기

데이터 수집 보고서

□ 개요

- 산출물 단계 : 데이터 수집 및 저장
- 평가 산출물 : 데이터 수집 보고서
- 제출 일자 : 2024-12-27
- 깃허브 경로 : <https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN05-FINAL-4TEAM.git>
- 작성 팀원 : 김지연

데이터 수집 목적	<p>본 데이터 수집의 목적은 AI 모의면접 시스템을 통해 실제 면접과 유사한 환경에서 연습할 수 있도록 다양한 직무와 기술 수준에 맞춘 인터뷰 질문을 확보하는 데 있다.</p> <p>수집된 데이터는 기초적인 질문부터 심화형, 해결형 질문까지 포괄하며, 이를 활용해 사용자 맞춤형 질문 생성과 답변 분석을 지원하며 면접 준비 과정의 효율성을 극대화한다.</p>
데이터 수집 방법	<p>1.1 기본 인터뷰 질문 (웹 크롤링)</p> <ul style="list-style-type: none">- 특정 웹사이트(monster.com)에서 제공하는 면접 질문을 크롤링하여 수집- 행동 면접 질문, 문제 해결 면접 질문, 소프트 스킬 면접 질문, 역량 기반 면접 질문 <p>1.2 기술 면접 질문(웹 크롤링, pdf, kaggle)</p> <ul style="list-style-type: none">- 다양한 기술 분야에 대한 면접 질문 수집- 특정 웹사이트 , kaggle 공개 데이터 셋, PDF 형식 자료 활용 <p>1.3 cs 면접 질문 (웹 크롤링)</p> <ul style="list-style-type: none">- 컴퓨터 과학(cs) 기본 지식을 평가하는 질문 수집- cs 관련 블로그 활용 <p>1.4 사용 기술 및 저장 포맷</p> <ul style="list-style-type: none">- 웹 크롤링: Python의 BeautifulSoup, Selenium, Webdriver 사용- pdf 데이터 추출 : PyPDF2, PDFplumber 사용- 수집된 데이터를 CSV 또는 JSON 포맷으로 저장

수집 데이터(요약)

2.1 기본 인터뷰 질문

- 주요 키워드 : 팀워크, 리더십, 문제해결
- 특정 상황에서의 대응 및 문제 해결 역량 확인을 위한 질문
- 조직 문화 적응력과 의사소통 능력을 평가하는 데 활용

2.2 기술 면접 질문

- python, java, aws, sql, LLM, machine learning 분야 관련 질문 내용
- 다양한 기술 스택에 대한 LLM 학습용 데이터로 활용 예정
- 면접 질문 생성 알고리즘의 기반 데이터로 사용 예정
- 기술 스택별로 실무 역량을 검증하기 위한 질문 포함

2.3 cs 면접 질문

- 자료구조, 알고리즘, 운영체제, 네트워크
- cs 기본기를 검증하는 질문 및 답변 생성 데이터로 활용 예정

2.4 수집 데이터 개요

- 총 데이터 양 : 11개 파일(json, csv 파일 기준)
 - 기본 인터뷰: 1개
 - 기술 면접: 9개
 - cs 면접: 1개