



Code States AIB(AI Bootcamp)

프로젝트 기획 보고서

수강생 이름 : 우성호
시작일 : 2022.11.4.
프로젝트 : 프로젝트1

1. 프로젝트 주제

코드스테이츠 수료생의 취업연계를 위한 구인 정보 탐색

2. 프로젝트 목적

코드스테이츠는 수강생들에게 IT에 관련된 양질의 교육을 제공함으로써 커리어 전환 및 취업을 도모하고 있다. 그럼에도 교육을 수료한 학생들에게 구직활동에 어려움을 겪는 경우가 존재해왔다. 이를 위해 코드스테이츠에서 수료생 취업을 위한 잡서칭의 일환으로 구인 공고를 가져와 수료생 및 수료 예정 교육생을 위해 취업 정보를 가져와 교육생과 수료생에게 추천해주는 서비스를 제공하고자 한다.

3. 프로젝트 구현에 필요한 기술

- 웹사이트 크롤링.(BeautifulSoup4, Selenium)
 - Open API 존재할 경우 제공되는 API 활용(Open API)
- JSON의 형태로 NoSQL에 저장(MongoDB, NoSQL)
- JSON을 정제하여 LocalDB에 저장. 최종적으로 DS가 LocalDB를 이용하게끔 조치(mysql)
- 이상의 형식을 7일 주기로 수행하게끔 자동화 구현하여 Docker 컨테이너 형태로 제작(Docker)

4. 프로젝트에 필요한 강의 콘텐츠

- Section 3 Notes
 - N314: DB API
 - N322: Web Scraping - BeautifulSoup4
 - N323: OpenAPI
 - N324: NoSQL - MongoDB
 - N331: Docker
 - N334:시간과 부호화
- Selenium(<https://www.selenium.dev/documentation/>)

5. 프로젝트 디자인

1. 워크넷, 잡코리아, 사람인 등 -(크롤링 및 Open API 수령) -> MongoDB(JSON) -(데이터정제)-> LocalDB(mysql)
2. 1을 7일 주기로 수행하는 Batch파일 구성. 이를 Docker 컨테이너에 업로드 및 배포

6. 프로젝트 일정

- 11.7. 구인 사이트 데이터 크롤링
- 11.8. 크롤링 데이터 정제 및 로컬 데이터 베이스에 적재
- 11.9. 크롤링 자동화 구현
- 11.10. 도커 이미지 제작(가능하다면 이때 최소기능제품 구현)
- 11.11. 발표 보고서 작성 및 제출

7. 프로젝트 활용 데이터

- 워크넷 OpenAPI
 - 데이터 설명 : 정부고용포털 워크넷에서 제공하는 OpenAPI
 - 데이터 수집 방법 : 공공 데이터(<http://openapi.work.go.kr/opi/opi/opia/wantedApi.do>)
 - 데이터 타입(종류) : XML
 - 데이터 크기 : 미정
 - 데이터 활용 방안(예정) : 받아온 API를 정제하여 NoSQL에 저장 후 LocalDB에 저장 예정
- 잡코리아 채용정보
 - 데이터 설명 : 특정 키워드에 따른 잡코리아에 업로드된 채용정보
 - 데이터 수집 방법 : 크롤링
 - 데이터 타입(종류) : JSON
 - 데이터 크기 : 미정
 - 데이터 활용 방안(예정) : 크롤링한 데이터를 JSON으로 변환 및 정제하여 NoSQL에 저장 후 LocalDB에 저장 예정

8. 팀 및 개인 Github 링크 작성

- 우성호: <https://www.github.com/e080hsm>
-