|  |  |
| --- | --- |
| **[2019-데이터크롤링] 기말 과제 계획서** | |
| **작 성 자** | **소프트웨어학과 2014301001 강민승** |
| **프로젝트 명** | **고용률에 따른 취준생들의 만족감 유추하기** |
| 프로젝트 목표 | 취업 커뮤니티에서 활동하는 취준생들의 게시글에서 말투의 분위기를 유추하여 고용률과의 상관관계를 알아내기   * 고용률이 높은 해의 취준생들은 말투의 분위기가 고용률이 낮은 해의 말투보다 밝다. |
| 프로젝트 수행을 위한 학습내용 | 1. BeautifulSoup 모듈을 이용한 DataCrawling  2. selenium 라이브러리의 webdriver 모듈을 이용한 Web-Crawling  3. csv파일의 필요 정보만을 정제하여 조합  4. matplotlib 라이브러리를 이용하여 데이터 시각화 |
| 데이터 수집  계획 | 1. BeautifulSoup 모듈을 이용한 DataCrawling  2. selenium 라이브러리의 webdriver 모듈을 이용한 Web-Crawling  3. 통계청 오픈 데이터를 통해 연도별 고용률 현황 조사 |
| 데이터 준비 및 탐색 작업 계획 | 1. 데이터 정제: 필요 정보를 제외한 정보를 정제.  2. 데이터 조합: 취업 커뮤니티 게시판마다 얻은 게시글의 정보를 연도별로 조합 |
| 분석 모델 구축  계획 | 1. webdriver 모듈을 이용하여 원하는 웹 페이지에 접근  2. BeautifulSoup 모듈을 이용하여 필요 정보만을 모아 csv파일로 저장 |
| 결과 시각화  계획 | matplotlib 라이브러리를 이용하여 얻어낸 csv 파일을 막대 그래프 형태로 표현 |