**[서비스 산업 데이터를 활용한 빅데이터 분석]**

**비정형 데이터 활용 분석 프로젝트 기획안**

기획안 작성일자 : 2022년 6월 14일

|  |  |
| --- | --- |
| 조 명 | 부들부들 |
| 프로젝트 주제 및 개요 | 미리보기와 태그를 활용한 맞춤형 웹 소설 추천 시스템 |
| 프로젝트 수행 방향 | 소설 사이트 종류 파악 및 미리보기와 태그 수집  > yes24에는 리뷰와 평점도 존재하는데 이 데이터를 활용  >> yes 24 e book 미리보기 = 텍스트 데이터 (다른 곳은 이미지, 뷰어)  소설을 태그에 따라 분류  > 알고리즘: 클러스터링  > 형태소 분석 후 군집 알고리즘 사용    (개인취향) 선호하는 소설 작품과 유사한 소설을 추천  특정 소설을 선택시, 선택한 소설과 비슷한 소설을 추천함  작품 태그가 정확히 반영 안된 소설들에 자동으로 태그생성  ● 수행도구, 데이터소개  - BeautifulSoup  - Selenium  - NLP(코사인 유사도)  - Clustering  - React, Flask |
| 프로젝트 조직  (구성원 및 역할) | ● 역할분담  조장: 전승훈 : 데이터 전처리, 모델 튜닝  조원1: 강창훈 : 데이터 크롤링, ppt  조원2: 김규형 : 데이터 크롤링, 웹앱  조원3: 김연지 : 데이터 크롤링, ppt  조원4: 김지원 : 데이터 전처리, 웹앱  조원5: 정성원 : 데이터 전처리, 모델 튜닝 |
| 프로젝트 추진 일정 | ● 일정  6/13 ~ 6/14: 주제 선정 및 일정 수립  6/15 ~ 6/17: 데이터 크롤링, 데이터 전처리  6/20 ~ 6/22: 데이터 전처리, 모델 학습, 모델 튜닝, 웹앱 제작  6/23 ~ 6/24: 최종 마무리  6/13 ~ 6/24: ppt 작성 및 발표 |