**기업요구사항 기반의 문제해결 프로젝트 기획안**

**2021년 8월 27일**

| **프로젝트 조** | **2조** | |
| --- | --- | --- |
| **프로젝트 팀원** | 팀명: 녀석들  팀장: 김민규  팀원: (사이언스) 김영성, 장우창, 이우재 (엔지니어링) 강민규, 김민규 | |
| **프로젝트 주제** | 웹툰 현황 분석 및 텍스트 분석과 이미지 유사도를 이용한 추천 시스템 개발 | |
| **프로젝트 수행 방향**  **(주요 기능 설명)** | **사이언스** | **엔지니어링** |
| 웹툰 추천 알고리즘 개발  1. 썸네일 이미지 유사도 기반  2. 웹툰 설명 유사도 기반  - 작가, 장르 선택 등 옵션 기능 추가 (또는 가중치) | - 데이터 수집 및 전처리  - 웹스크래핑 후 mySQL 등 DB에 저장  - 웹툰 제목, 작가, 장르, 소개, 평균 별점 : 네이버, 카카오페이지, 레진코믹스 등 다양한 플랫폼  - 웹 서비스 구현 |
| **프로젝트 수행 도구** | **사이언스** | **엔지니어링** |
| **(공통) pandas, matplotlib, seaborn** | |
| scikit-learn  tensorflow | mySQL  BeautifulSoup |

| **프로젝트 목적** | **특정 웹툰을 좋아하는 유저에게 그와 비슷한 다른 웹툰을 추천하여 잠재적인 결제 고객 유치** | | |
| --- | --- | --- | --- |
| **필수 기능** | **사이언스** | **엔지니어링** |  |
| **(공통) EDA, 데이터 시각화** | |  |
| 장르와 소개에 기반하여 유저가 원하는 주제의 웹툰 추천  이미지 유사도를 측정하여 유저가 원하는 그림체의 웹툰 추천 | 웹페이지 구현  웹툰 검색  검색 웹툰과 유사한 웹툰 추천기능   * 동일 작가 * 장르 * 웹툰 설명 * 그림체 |  |
| **포함 기술** | **사이언스** | **엔지니어링** |  |
| TF-IDF 등을 이용한 텍스트 유사도 측정  CNN 등 딥러닝 모델 이용  - tensorflow hub를 이용한 feature extraction을 통해 군집화  - 그림 작가별로 label을 붙여 이미지 학습 후 softmax의 2순위 웹툰 추천 | BeautifulSoup, Selenium 등을 이용한 웹툰 정보 및 썸네일 스크래핑  Django 등을 이용한 웹 서비스  데이터 적재, 처리, 추출, 관리  네이버 API를 이용한 사용자 수집  서비스 단계에서 사용자 정보를 입력받는 데이터 베이스 구축 |  |

**□ WBS**

[2조\_WBS\_녀석들.xlsx](https://docs.google.com/spreadsheets/d/112qr_Z4Axji9X_GqXZ9gnCbkOlKz1e3j/)