

01) Python

1. 목표

- 기초 Python에 대한 이해
- Python을 통한 데이터 수집 및 파일 저장
- 웹 크롤러 및 API 활용을 통해 데이터를 수집하고 내가 원하는 형태로 가공한다.
- 영화평점사이트(예- watcha)에 필요한 데이터를 프로그래밍을 통해 수집한다.

2. 준비 사항

1. Python 환경 설정
 - python 3.6 이상
2. 필수 라이브러리 활용
 - requests
3. 필수 API
 - [영화진흥위원회 오픈 API](#)
 - 주간/주말 박스오피스 API 서비스
 - 영화 상세정보 API 서비스
 - [네이버 영화 검색 API](#)

주의!!! API키는 반드시 환경변수로 지정해주세요. 절대 소스코드에 노출시키지 마세요.

3. 요구 사항

- 영화평점서비스(예- watcha)를 만들기 위한 데이터 수집 단계로, 영화 데이터베이스 구축을 위한 csv 파일을 만든다. 아래 기술된 사항은 필수적으로 구축해야 하는 내용이며, 이외에 자유롭게 추가 데이터를 수집하는 것도 가능하다.

1. 영화진흥위원회 오픈 API(주간/주말 박스오피스 데이터)

- 최근 10주간 데이터 중에 주간 박스오피스 TOP10데이터를 수집합니다. 해당 데이터는 향후 영화평점서비스에서 기본으로 제공되는 영화 목록으로 사용될 예정입니다.
 - 요청 조건
 1. 주간(월~일)까지 기간의 데이터를 조회합니다.
 2. 조회 기간은 총 10주이며, 기준일(마지막 일자)은 2019년 1월 13일입니다.
 3. 다양성 영화/상업 영화를 모두 포함하여야 합니다.
 4. 한국/외국 영화를 모두 포함하여야 합니다.
 5. 모든 상영지역을 포함하여야 합니다.
 - 결과
 - 수집된 데이터에서 `영화 대표코드`, `영화명`, `해당일 누적관객수`, `해당일` 을 기록합니다.
 - `해당일 누적관객수` 는 중복시 최신 정보를 반영하여야 합니다. 예) 영화 아쿠아맨이 20190113 기

준 50,000명이고, 20190106 기준 5,000명이면 50,000명이 저장되어야 합니다.

- 해당 결과를 **boxoffice.csv** 에 저장합니다.

2. 영화진흥위원회 오픈 API(영화 상세정보)

- 위에서 수집한 영화 대표코드를 활용하여 상세 정보를 수집합니다. 해당 데이터는 향후 영화평점서비스에서 영화 정보로 활용될 것입니다.

- 결과

- 영화별로 다음과 같은 내용을 저장합니다.

영화 대표코드, 영화명(국문), 영화명(영문), 영화명(원문), 개봉연도, 상영시간, 장르, 감독명, 관람등급, 배우1, 배우2, 배우3

- 배우의 경우 최대 3명입니다. 영화에 따라 1~2명일 수도 있습니다.

- 해당 결과를 **movie.csv**에 저장합니다.

3. 네이버 영화 검색 API

- 앞서 영진위에서 얻은 영화명(국문)을 바탕으로 네이버 영화 검색 API를 통해 추가적인 데이터를 수집합니다. 해당 데이터는 향후 영화평점서비스에서 기준 평점 및 영화 포스터 썸네일로 활용될 것입니다.

- 요청

- 영화명을 통해 요청합니다.

- 응답

- 영화별로 다음과 같은 내용을 저장합니다. 영진위 영화 대표코드, 영화 썸네일 이미지의 URL, 하퍼텍스트 link, 유저 평점

- 해당 결과를 **movie_naver.csv**에 저장합니다.

4. 영화 포스터 이미지 저장

- 앞서 네이버 영화 검색 API를 통해 얻은 이미지 URL에 요청을 보내 실제 이미지 파일로 저장합니다. 해당 데이터는 향후 영화 목록에서 포스터 이미지로 사용될 것입니다.

- 요청

- 이미지 URL

- 응답

- 응답받은 결과를 파일로 저장합니다. 저장시 반드시 **wb** 옵션으로 저장하시기 바랍니다.

- 저장되는 파일명은 images 폴더 내에 영진위 영화 대표코드.jpg 입니다.

4. 결과 예시

위에 명시된 사항은 최소 조건이며, 추가적인 정보를 수집하여도 됩니다.

결과물은 반드시 [README.md](#) 으로 활용하였던 API 정보를 정리하고, 결과로 저장된 csv 파일에 대한 설명을 기록해야 합니다.

movie/

README.md

*.py

boxoffice.csv

movie.csv

movie_naver.csv

images/

*.jpg