01 - 파이썬을 활용한 데이터 수집 I

1. 목표

- 기초 Python에 대한 이해
- Python을 통한 데이터 수집 및 파일 저장
- Python 조건/반복문 및 다양한 자료구조 조작
- 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 3.7 이상
    - `Visual Studio Code` 혹은 `Jupyter notebook` 활용 할 것.
2. 필수 라이브러리 활용
    - requests
3. 필수 API
    - 영화진흥위원회 오픈 API
        - API 활용시 요청URL 작성시 `json` 을 활용하세요.
        - 주간/주말 박스오피스 API 서비스
        - 영화 상세정보 API 서비스

주의!!! APIKEY는 반드시 환경변수로 지정 하세요. 절대 소스 코드에 노출시키지 마세요.

3. 요구 사항

- 영화평점서비스(예- watcha)을 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.

주의!!! 각 문제 별로 별도의 파일을 만들어서 제출 해주세요.

1. **영화진흥위원회 오픈 API(주간/주말 박스오피스 데이터) - `01.py` 혹은 `01.ipynb`**

- 최근 50주간 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 50주이며, 기준일(마지막 일자)은 2019년 7월 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`를 기록합니다.
  - `해당일 누적관객수`는 중복시 최신 정보를 반영하여야 합니다.
    예) 영화 엄복동이 20190713 기준 50,000명이고, 20190106 기준 5,000명이면 50,000명이 저장되어야 합니다.
  - 해당 결과를 **boxoffice.csv** 에 저장합니다.

2. **영화진흥위원회 오픈 API(영화 상세정보) - `02.py` 혹은 `02.ipynb`**

- 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)`, `관람등급`, `개봉연도`, `상영시간`, `장르`, `감독명`
    - 해당 결과를 **movie.csv**에 저장합니다.
    - (선택) 배우 정보, 배급사 정보 등을 추가적으로 수집할 수 있습니다.

3. 영화진흥위원회 오픈 API(영화인 정보) - `03.py` 혹은 `03.ipynb`

- 위에서 수집한 영화 감독정보를 활용하여 상세 정보를 수집합니다. 해당 데이터는 향후 영화평점서비스에서 감독 정보로 활용될 것입니다.
  - 요청 조건
    - `영화인명` 으로 조회합니다.
  - 결과
  - 영화인별로 다음과 같은 내용을 저장합니다.

    `영화인 코드`, `영화인명`, `분야`, `필모리스트`
  - 해당 결과를 **director.csv**에 저장합니다.

  단, 만약 검색 결과가 없으면 저장할 필요 없습니다.

4. 결과 예시

위에 명시된 사항은 최소 조건이며, 해당 API를 활용하여 추가적인 정보를 수집하여도 됩니다.

혹은 다른 곳에서 제공하는 영화 관련 API를 확인 해보세요.

- KMDB
- TMDB
- 네이버 영화검색

결과물은 반드시 `README.md` 으로 활용하였던 API 정보를 정리하고, 결과로 저장된 csv 파일에 대한 설명을 기록해야 합니다.

pjt01/

README.md

*.py 또는 *.ipynb : 문제별 소스코드

boxoffice.csv

movie.csv

director.csv