

02 - 파이썬을 활용한 데이터 수집 II

1. 목표

- API 요청을 통한 데이터 수집 및 파일 저장
- Python을 통한 JSON 및 이미지 파일 조작
- API 활용을 통해 데이터를 수집하고 내가 원하는 형태로 가공한다.
- 영화평점서비스(예- watcha)에 필요한 추가 데이터를 프로그래밍을 통해 수집한다.

2. 준비 사항

1. Python 환경 설정

- python 3.7 이상
- Visual Studio Code 혹은 Jupyter notebook 활용 할 것.
 - Jupyter notebook 활용하는 경우는 .py 로 변환하여 제출 바랍니다.

2. 필수 라이브러리 활용

- requests

3. 필수 API

- [네이버 영화 검색 API](#)
 - API 활용시 요청URL 작성시 json 을 활용하세요.

주의!!! API키는 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.

1. 네이버 영화 검색 API - movie_naver.py

- 지난 프로젝트(파이썬을 활용한 데이터 수집 I)에서 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, 하이퍼텍스트 link, 영화 썸네일 이미지의 URL, 유저 평점
 - 영화 썸네일 이미지의 URL 이 없는 경우 저장하지 않습니다.
 - 해당 결과를 movie_naver.csv에 저장합니다.

2. 영화 포스터 이미지 저장 - `movie_image.py`

- 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
 - 영화 썸네일 이미지의 URL
 - 응답
 - 응답 받은 결과를 파일로 저장합니다. 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.

추가적인 영화 관련 API를 확인 해보세요.

- [KMDB](#)
- [TMDB](#)
- [네이버 영화검색](#)

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

주의!!! 각 문제 별로 별도의 파일을 만들어서 제출 해주세요.

```
project02/  
  README.md  
  movie_naver.py  
  movie_naver.csv  
  movie_image.py  
  images/  
  ...
```