

웹개발 종합반 정리 -3

- 3주차: Python 패키지 requests, bs4(BeautifulSoup) / 크롤링(scraping)

링크

- ※ 스파르타코딩클럽 <https://online.spartacodingclub.kr/classroom>
- ※ 웹개발 강의 doc <https://teamsparta.notion.site/1-da6d8595945d4b868e5fcfa933c56011>

- ※ Python 패키지: 외부 라이브러리 // Python: <https://www.python.org/>
- ※ 가상환경 Virtual Environment이란? 프로젝트별로 패키지를 담을 도구함, .venv폴더 // ex) node_modules
- ※ 패키지 설치: 파일 > 설정 > 프로젝트: ~~ > Python 인터프리터 > +클릭 > 검색

requests

Python 패키지, HTTP request

```
import requests          # requests를 import

r = requests.get('~~')   # url에 get리퀘스트를 보냄
rjson = r.json()         # response, json()

print(rjson)             # json출력

rows=rjson['RealtimeCityAir']['row'] # rows설정
for row in rows:         # forEach
    gu_name = row['MSRSTE_NM']
    gu_mise = row['IDEX_MVL']
    if gu_mise < 60:      # 들여쓰기 + 조건문
        print(gu_name, ":", gu_mise) # 들여쓰기2 + 프린트
```

beautifulsoup

Screen-scraping library, 크롤링, 스크린 스크래핑 // bs4로 설치!

- ① bs4 설치
- ② url 설정

select_one(): 하나 추출
요소.text

- select(): 여러개 추출
for ~~ 요소.text
- ① 요소 잡기
 - ② 요소 박스 잡기

이름, 랭크, 점수

```
import requests          # requests를 import
from bs4 import BeautifulSoup # bs4에서 BeautifulSoup를 import

headers = {'User-Agent' : 'Mozilla/5.0 (Windows NT 10.0; Win64; x64)AppleWebKit/537.36
(KHTML, like Gecko) Chrome/73.0.3683.86 Safari/537.36'}
data= requests.get('~~',headers=headers)

soup = BeautifulSoup(data.text, 'html.parser')

title = soup.select_one(' 복사한 selector ')
print(title.text) # text만 추출
print(title['href']) # href만 추출

# TD, 리스트 요소까지 잡았을 경우
movies = soup.select(' 복사한 selector ')
for movie in movies:
    print(movie.text)

# TR, 리스트의 목록 요소를 잡아서 예외처리
movies = soup.select(' 복사한 selector ')
for movie in movies:
    a = movie.select_one(' td.title > a ')
    if a is not None:
        print(a.text)

for movie in movies:
    a = movie.select_one("td > div > a")
    if a is not None:
        name = a.text
        rank = movie.select_one("td:nth-child(1) > img")["alt"] // ().attr["alt"] = ()["alt"]
        star = movie.select_one("td.point").text
        print(rank, name, star)
```

예러A: Ctrl+V Python에서 코드 복사해서 한글로 복사 파일 멈춤/응답없음 오류

예러B: 파일명 attributeError: partially initialized module 'requests' has no attribute 'get' (most likely due to a circular import)
파일명과 모듈명이 같을 경우 발생 ex) requests.py 안 됨!