



네이버 영화 크롤링

목차

1. Requests
2. 구현하기
3. 웹페이지 정보 가져오기
4. 파이썬과 BeautifulSoup를 이용한 웹 크롤러 실습하기

Requests 기본 사용

- 웹 상에 있는 HTML 파일을 읽어올때 requests 라이브러리가 필요하다.
- `pip install requests`

예시)

```
import requests
URL = 'http://naver.com/'
response = requests.get(URL)
print(response.text)
```

`https://www.naver.com/` 의 HTML소스를 String으로 변환시켜서 변수에 저장시켜주는 역할

구현하기

Step 1) 웹 페이지 정보 가져오기

<https://movie.naver.com/movie/bi/mi/review.nhn?code=191633>

- pip install bs4
- pip install lxml

```
import requests

from bs4 import BeautifulSoup

# '트롤'의 네이버 영화 리뷰 링크

url = "https://movie.naver.com/movie/bi/mi/review.nhn?code=191633"

# html 소스 가져오기

res = requests.get(url)
```

구현하기

Step 2) HTML 소스를 파싱(=분석)하여 원하는 정보얻기

```
# html 파싱
soup = BeautifulSoup(res.text, 'lxml')

# 리뷰 리스트
ul = soup.find('ul', class_="rvw_list_area")
lis = ul.find_all('li')

# 리뷰 제목 출력
count=0

for li in lis:
    count += 1
    print(f"[{count}th] ", li.a.string)
```

구현하기

출력결과>

[1th] [리뷰] 트롤 : 월드투어
[2th] 다른 의견을 들어주고 다른 것을 인정하는 자세를 귀엽게도 그린 트롤 월드 20
[3th] K-POP 좋아해? '트롤: 월드 투어' // 2020.04.30 (수) 관람
[4th] 트롤 : 월드 투어
[5th] 트롤:월드투어
[6th] [영화라는 이름의 꿈] 화려한 콘서트의 현장-트롤:월드투어
[7th] 트롤 : 월드 투어(우리말녹음) 부평 대한극장에서 봤어요!!
[8th] 트롤: 월드 투어
[9th] 트롤: 월드 투어
[10th] [영화] 트롤2: 월드 투어 {줄거리&결말포함/스포주의}

구현하기

Step 3) 전체 코드

```
import requests

from bs4 import BeautifulSoup

url = "https://movie.naver.com/movie/bi/mi/review.nhn?code=191633"
res = requests.get(url)

soup = BeautifulSoup(res.text, 'lxml')

ul = soup.find('ul', class_="rvw_list_area")

lis = ul.find_all('li')

count=0

for li in lis:

    count += 1

    print(f"[{count}th] ", li.a.string)
```

참고자료

- `print(f"[{count}th] ", li.a.string)`
- 시작 인용 부호 또는 삼중 인용 부호 앞에 `f` 또는 `F` 를 붙여 문자열을 시작하십시오. 이 문자열 안에서, `{` 및 `}` 문자 사이에, 변수 또는 리터럴 값을 참조할 수 있는 파이썬 표현식을 작성할 수 있습니다.

예시)

```
>>> year = 2016
>>> event = 'Referendum'
>>> f'Results of the {year} {event}'
'Results of the 2016 Referendum'
```

- <https://docs.python.org/ko/3/tutorial/inputoutput.html>

Thank you
