

# 크롤링 패턴 코드

다음 코드를 우선 그대로 쓰고, 두 부분만 수정!

```
import requests
from bs4 import BeautifulSoup
```

① 크롤링할 페이지 주소 넣기



```
res = requests.get('http://v.media.daum.net/v/20170615203441266')
```

```
soup = BeautifulSoup(res.content, 'html.parser')
```

```
mydata = soup.find('title') ← ② 필요한 데이터 추출하는 코드 넣기
```

```
print(mydata.get_text()) ← 추출한 데이터를 변수에 넣은 후, 원하는 프로그래밍
```

# HTML/CSS 언어 이해를 기반으로 크롤링해보기

```
import requests
from bs4 import BeautifulSoup

# 1) requests 라이브러리를 활용한 HTML 페이지 요청
# 1-1) res 객체에 HTML 데이터가 저장되고, res.content로 데이터를 추출할 수 있음
res = requests.get('http://v.media.daum.net/v/20170615203441266')

# print(res.content)
# 2) HTML 페이지 파싱 BeautifulSoup(HTML 데이터, 파싱방법)
# 2-1) BeautifulSoup 파싱방법
soup = BeautifulSoup(res.content, 'html.parser')

# 3) 필요한 데이터 검색
title = soup.find('h3', 'tit_view')

# 4) 필요한 데이터 추출
print(title.get_text())
```

# HTML/CSS 언어 이해를 기반으로 크롤링해보기

```
data = soup.find('h3', 'tit_view')  
print(data.get_text())
```

```
data = soup.find('span', 'txt_info')  
print(data.get_text())
```

```
data = soup.find_all('span', 'txt_info')  
for item in data:  
    print(item.get_text())
```

```
data = soup.find_all('span', 'txt_info')  
print(data[1].get_text())
```

```
data = soup.find('div', 'layer_util layer_summary')  
print(data.get_text())
```