

크롤링 강력한 팁1 - 크롬 브라우저 활용하기

- 오픈 크롬 브라우저
- 오픈 크롬 개발자 모드
 - Command + Alt + i (맥)
 - Ctrl + Shift + i 또는 F12 (윈도우)
- 마우스로 원하는 부분 선택

예제1

```
import requests
from bs4 import BeautifulSoup

res = requests.get('https://davelee-fun.github.io/blog/crawl_test')
soup = BeautifulSoup(res.content, 'html.parser')

titles = soup.find_all('li', 'course')
for title in titles:
    print(title.get_text())
```

크롤링 강력한 팁2 - 추출한 것에서 또 추출하기

1. find()로 더 크게 감싸는 HTML 태그로 추출하고
2. 다시 추출된 데이터에서 find_all()로 원하는 부분을 추출

추출된 데이터는 객체(object)입니다.

예제2

```
import requests
from bs4 import BeautifulSoup

res = requests.get('https://davelee-fun.github.io/blog/crawl_test')
soup = BeautifulSoup(res.content, 'html.parser')

section = soup.find('ul', id='hobby_course_list')
titles = section.find_all('li', 'course')
for title in titles:
    print(title.get_text())
```

크롤링 강력한 팁3 - 파이썬 문자열 함수와 함께 쓰기

데이터 전처리

1. strip() 함수 사용해보기
2. split() 함수 사용해보기

```
import requests
from bs4 import BeautifulSoup

res = requests.get('https://davelee-fun.github.io/blog/crawl_test')
soup = BeautifulSoup(res.content, 'html.parser')

section = soup.find('ul', id='dev_course_list')
titles = section.find_all('li', 'course')
for index, title in enumerate(titles):
    print(str(index + 1) + '.', title.get_text().split('-')[-1].split('[')[0].st
```