

# 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

# 실전 크롤링1: 다음 실시간 검색어 크롤링

다음 실시간 검색어를 크롤링하고 싶어요!

<http://www.daum.net>

다양한 크롤링 기법을 익히고, 케이스별 적절히 사용

```
import requests
from bs4 import BeautifulSoup

res = requests.get('http://www.daum.net')
soup = BeautifulSoup(res.content, 'html.parser')

link_title = soup.find_all('a', 'link_issue')
for title in link_title:
    print(title.get_text())
```

다양한 크롤링 기법을 익히고, 케이스별 적절히 사용

```
import requests
from bs4 import BeautifulSoup

res = requests.get('http://www.daum.net')
soup = BeautifulSoup(res.content, 'html.parser')

link_title = soup.find_all('a', attrs = {'class': 'link_issue', 'tabindex': '-1'})
for title in link_title:
    print(title.get_text())
```

# 실전 크롤링2: 네이버 실시간 검색어 크롤링

다양한 크롤링 기법을 익히고, 케이스별 적절히 사용

```
import requests
from bs4 import BeautifulSoup

res = requests.get('http://www.naver.com')
soup = BeautifulSoup(res.content, 'html.parser')

link_title = soup.find_all('li', 'ah_item')
for title in link_title:
    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.

```
import requests
from bs4 import BeautifulSoup
res = requests.get('http://www.naver.com')
soup = BeautifulSoup(res.content, 'html.parser')

keyword_div = soup.find('div', 'ah_roll_area PM_CL_realtimeKeyword_rolling')
keywords = keyword_div.find_all('span', 'ah_k')
for keyword in keywords:
    print(keyword.get_text())
```

## 실전 크롤링3: 네이버 실시간 검색어와 등수 크롤링

```
import requests
from bs4 import BeautifulSoup
res = requests.get('http://www.naver.com')
soup = BeautifulSoup(res.content, 'html.parser')

keyword_div = soup.find('div', 'ah_roll_area PM_CL_realtimeKeyword_rolling')
keywords = keyword_div.find_all('li', 'ah_item')
for keyword in keywords:
    print(keyword.get_text())
```

# 크롤링 강력한 팁3 - 파이썬 문자열 함수와 함께 쓰기

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

```
import requests
from bs4 import BeautifulSoup
res = requests.get('http://www.naver.com')
soup = BeautifulSoup(res.content, 'html.parser')

keyword_div = soup.find('div', 'ah_roll_area PM_CL_realtimeKeyword_rolling')
keywords = keyword_div.find_all('li', 'ah_item')
for keyword in keywords:
    print(keyword.get_text().split()[0] + "위: " + keyword.get_text().split()[1])
```