

동일한 기능을 하는 다양한 코드와 문법이 있습니다

다양한 크롤링 추출 방법: CSS Selector 사용하기

어느 것이 더 나으냐는 없습니다. CSS Selector 도 꽤 유용합니다.

CSS Selector 사용법

- select() 안에 태그 또는 CSS class 이름등을 넣어주면 됨
- 결과값은 리스트 형태로 반환됨
 - 매칭되는 첫번째 데이터만 얻고자 할 때는 select_one(), 이 때는 해당 아이템 객체가 리턴

초간단 예제1 (태그 선택)

사용법을 익힐 때는 정말 간단한 예제로 익혀야 사용법 이해에 집중할 수 있습니다.

```
import requests
from bs4 import BeautifulSoup

res = requests.get('https://davelee-fun.github.io/blog/crawl_test_css.html')
soup = BeautifulSoup(res.content, 'html.parser')
items = soup.select('li')
for item in items:
    print(item.get_text())
```

초간단 예제2-1 (하위 태그선택)

```
import requests
from bs4 import BeautifulSoup

res = requests.get('https://davelee-fun.github.io/blog/crawl_test_css.html')
soup = BeautifulSoup(res.content, 'html.parser')
items = soup.select('html body h1')
for item in items:
    print(item.get_text())
```

초간단 예제2-2 (하위 태그선택)

- '>'는 바로 아래 태그인 경우를 검색

```
import requests
from bs4 import BeautifulSoup

res = requests.get('https://davelee-fun.github.io/blog/crawl_test_css.html')
soup = BeautifulSoup(res.content, 'html.parser')
items = soup.select('ul > li')
for item in items:
    print(item.get_text())
```

초간단 예제2-3 (css class 이름으로 검색)

- '.클래스 이름' 으로 검색 가능

```
import requests
from bs4 import BeautifulSoup

res = requests.get('https://davelee-fun.github.io/blog/crawl_test_css.html')
soup = BeautifulSoup(res.content, 'html.parser')
items = soup.select('.course')
for item in items:
    print(item.get_text())
```

초간단 예제2-4 (id 이름으로 검색)

- '#id 이름' 으로 검색 가능

```
import requests
from bs4 import BeautifulSoup

res = requests.get('https://davelee-fun.github.io/blog/crawl_test_css.html')
soup = BeautifulSoup(res.content, 'html.parser')
items = soup.select('#start')
for item in items:
    print(item.get_text())
```

초간단 예제2-5 (태그.클래스이름1.클래스이름2. ...)

```
import requests
from bs4 import BeautifulSoup

res = requests.get('https://davelee-fun.github.io/blog/crawl_test_css.html')
soup = BeautifulSoup(res.content, 'html.parser')
items = soup.select('li.course.paid')
for item in items:
    print(item.get_text())
```

초간단 예제2-6 (복합 예제)

```
import requests
from bs4 import BeautifulSoup

res = requests.get('https://davelee-fun.github.io/blog/crawl_test_css.html')
soup = BeautifulSoup(res.content, 'html.parser')
items = soup.select('ul#hobby_course_list li.course')
for item in items:
    print(item.get_text())
```

초간단 예제2-7 (select_one)

```
from bs4 import BeautifulSoup
import requests

res = requests.get('https://davelee-fun.github.io/blog/crawl_test_css.html')
soup = BeautifulSoup(res.content, 'html.parser')
item = soup.select_one('ul#dev_course_list > li.course.paid')
print(item.get_text())
```

추출한 것에서 추출하기 (find()와 select() 비교)

- find()/select()로 가져온 객체에는 find()/select() 함수를 사용 가능

```
from bs4 import BeautifulSoup
import requests

res = requests.get('https://davelee-fun.github.io/blog/crawl_html_css.html')
soup = BeautifulSoup(res.content, 'html.parser')
items = soup.select('tr')
for item in items:
    columns = item.select('td')
    row_str = ''
    for column in columns:
        row_str += ', ' + column.get_text()
print (row_str[2:])
```