

네이버 연관 검색어 수집

- 정적(static) 웹페이지 데이터 수집
- BeautifulSoup을 이용하여 HTML 문자열 데이터 parsing

```
In [3]: import pandas as pd
import requests
from bs4 import BeautifulSoup
```

1. 웹페이지 분석 : URL

```
In [6]: query= '삼성전자'
url = f'https://search.naver.com/search.naver?query={query}'
url
```

```
Out[6]: 'https://search.naver.com/search.naver?query=삼성전자'
```

2. request(URL) > response : str(html)

```
In [7]: response = requests.get(url)
response
```

```
Out[7]: <Response [200]>
```

```
In [8]: response.text[:300]
```

```
Out[8]: '<!doctype html> <html lang="ko"><head> <meta charset="utf-8"> <meta name="referrer" content="always"> <meta name="format-detection" content="telephone=no,address=no,email=no"> <meta property="og:title" content="삼성전자 : 네이버 통합검색"/> <meta property="og:image" content="https://ssl.pstatic.net/sstatic/se'
```

3. str(html) > bs object

```
In [9]: dom = BeautifulSoup(response.text, 'html.parser')
type(dom)
```

```
Out[9]: bs4.BeautifulSoup
```

4. bs object > .select(css-selector), .select_one(css-selector) > str(text)

```
In [10]: # select() : element 여러개 선택
# select_one() : element 한개 선택
```

```
In [11]: # 10개 선택 # 크롬
selector = '#nx_footer_related_keywords > div > div.related_srch > ul > li'
elements = dom.select(selector)
len(elements), elements[0]
```

```
Out[11]: (10,
<li class="item"> <a class="keyword" href="?where=nexearch&ssc=tab.nx.all&query=%EC%82%BC%EC%84%B1%EC%A0%84%EC%9E%90%EC%A3%BC%EA%B0%80&sm=tab_she&qdt=0" onclick="return goOtherCR(this, 'a=rsk_btm*q.list1&r=1&u=' + urlencode(this.href))"> <div class="tit">삼성전자주가</div> </a> </li>)
```

```
In [12]: elements[0].text.strip()
```

```
Out[12]: '삼성전자주가'
```

```
In [15]: keywords = [element.text.strip() for element in elements]
keywords
```

```
Out[15]: ['삼성전자주가',
'삼성전자주식',
'삼성전자 배당금',
'삼성전자채용',
'삼성전자주식가격',
'삼성전자서비스',
'오늘 삼성전자 주가',
'삼성전자주가지수',
'삼성전자서비스센터',
'삼성 전자레인지']
```

5. str(text) > DataFrame

```
In [16]: pd.DataFrame(keywords)
```

```
Out[16]:
```

	0
0	삼성전자주가
1	삼성전자주식
2	삼성전자 배당금
3	삼성전자채용
4	삼성전자주식가격
5	삼성전자서비스
6	오늘 삼성전자 주가
7	삼성전자주가지수
8	삼성전자서비스센터
9	삼성 전자레인지

```
In [ ]:
```