

xpath

- html element 선택하는 방법
- scrapy 에서 기본적으로 사용하는 selector
- scrapy : 파이썬 코드로 웹사이트의 데이터를 수집하는 프레임워크

In [3]:

```
# !pip install scrapy
```

In [1]:

```
import scrapy, requests
from scrapy.http import TextResponse
```

In [2]:

```
# 네이버 연관검색어 수집
```

In [10]:

```
query = "삼성전자"
url = f"https://search.naver.com/search.naver?query={query}"
url
```

Out[10]:

```
'https://search.naver.com/search.naver?query=삼성전자'
```

In [11]:

```
response = requests.get(url)
dom = TextResponse(response.url, body=response.text, encoding="utf-8")
```

In [12]:

```
# xpath syntax
# //*[@id="nx_footer_related_keywords"]/div/div[2]/ul/li[1]/a/div
# // : 최상위 엘리먼트
# * : 모든 하위 엘리먼트 : css selector(.wrap p)
# [@id="nx_footer_related_keywords"] : 속성값으로 엘리먼트 선택
# / : 한단계 하위 엘리먼트 : css selector(.wrap > p)
# [n] : n 번째 엘리먼트
```

In [18]:

```
dom.xpath('//*[@id="nx_right_related_keywords"]/div/div[2]/ul/li/a/div')[2]
```

Out[18]:

```
[<Selector xpath='//*[@id="nx_right_related_keywords"]/div/div[2]/ul/li/a/div' data='<div class="tit">삼성전자추가</div>'>,
 <Selector xpath='//*[@id="nx_right_related_keywords"]/div/div[2]/ul/li/a/div' data='<div class="tit">삼성전자 배당금</div>'>]
```

In [19]:

```
# text 수집
keywords = dom.xpath('//*[@id="nx_right_related_keywords"]/div/div[2]/ul/li/a/div/text()')
print(keywords)
```

```
['삼성전자주가', '삼성전자 배당금', '삼성전자서비스', '삼성전자주식', '오늘 삼성전자 주가',  
'삼성전자 배당금 지급일', '삼성전자물', '삼성 전자레인지', '삼성전자 채용', '삼성전자 고객센터']
```

In [9]:

```
# selenium 에서의 xpath는 for 을 사용해서 데이터를 수집
```