xpath

- html element 선택하는 방법
- scrapy 에서는 기본적으로 xpath를 사용

In [1]:
```
# install scrapy
# !pip install scrapy
```

In [2]:
```
import scrapy
import requests
from scrapy.http import TextResponse
```

In [3]:
```
url = 'https://www.gmarket.co.kr/n/best'
headers = {
    'user-agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebK
    'cookie': 'cguid=1169283480505400728200000; pguid=216928348050540072820
    'accept-language': 'ko-KR,ko;q=0.9,en-US;q=0.8,en;q=0.7',
    'sec-ch-ua-platform': 'macOS',
}
response = requests.get(url, headers=headers)
response
```

Out[3]:
```
<Response [200]>
```

In [4]:
```
text_response = TextResponse(
    response.url, body=response.text, encoding='utf-8')
text_response
```

Out[4]:
```
<200 https://www.gmarket.co.kr/n/best>
```

In [5]:
```
text_response.text[:500]
```

Out[5]:
```
'<!DOCTYPE html><html lang="ko" class="no-js"><head><meta name="viewport" c
ontent="width=980"/><meta charSet="utf-8"/><meta http-equiv="X-UA-Compatibl
e" content="ie=edge"/><meta name="description" content="인터넷쇼핑,오픈마켓,패션/
뷰티,디지털,식품/유아,스포츠/자동차,생활용품,도서/DVD,여행/항공권,e쿠폰/티켓,만화/게임,공동구매,
경매,중고,글로벌쇼핑,브랜드샵,베스트셀러,방문쇼핑몰,G스탬프,할인쿠폰,동영상,이벤트 등 G마켓"/><me
ta name="keywords" content="베스트100,베스트셀러,경매,할인쿠폰,베스트셀러,공동구매,컴퓨터/
핸드폰,에어컨/TV/디카,MP3/게임,패션/명품/브랜드,여성의류/속옷,남성의류/정장/빅사이즈,분유/기저귀/
식품/생리대/임부복,유아동/장난감,쌀/과일/한우/생선,건강식품/음료,화장품/'
```

1-1. xpath Selector

- `//*[@id="container"]/div[2]/ul/li[1]`
  - `//` : 최상위 엘리먼트
  - `*` : 모든 하위 엘리먼트 : css selector의 한칸띄우기와 같다.
  - `[@id="value"]` : 속성값 선택
  - `/` : 한단계 하위 엘리먼트 : css selector의 >와 같다.
  - `[n]` : nth-child(n)

In [6]:
```
items = text_response.xpath('//*[@id="container"]/div[2]/ul/li')
items[:2]
```

Out[6]:
```
[<Selector xpath='//*[@id="container"]/div[2]/ul/li' data='<li class="list-
item"><a href="http:/...'>,
 <Selector xpath='//*[@id="container"]/div[2]/ul/li' data='<li class="list-
item"><a href="http:/...'>]
```

1-2. Printing Titles Using Loops

In [7]:
```
# //*[@id="container"]/div[2]/ul/li[1]/a/div[2]/p
item = items[0]
item.xpath('a/div[2]/p/text()').extract()[0]
```

Out[7]:
```
'(추석전배송) 9/12폴햄 단하루 남여공용) 레이어드 반팔티 2팩 WWW (LIVE방송 30%쿠폰)'
```

In [8]:
```
for item in items[:2]:
    title = item.xpath('a/div[2]/p/text()').extract()[0]
    print(title)
```

```
(추석전배송) 9/12폴햄 단하루 남여공용) 레이어드 반팔티 2팩 WWW (LIVE방송 30%쿠폰)
초코하임 142g 4팩+화이트하임 142g 4팩
```

1-3. Printing Titles Without Loops

In [9]:
```
# 반복문 없이 출력
titles = text_response.xpath(
    '//*[@id="container"]/div[2]/ul/li/a/div[2]/p/text()').extract()
len(titles), titles[:2]
```

Out[9]:
```
(200,
 ['(추석전배송) 9/12폴햄 단하루 남여공용) 레이어드 반팔티 2팩 WWW (LIVE방송 30%쿠폰)',
  '초코하임 142g 4팩+화이트하임 142g 4팩'])
```