1-1. 오픈소스 패키지

1-2. 데이터 수집기 완성하기

2. 안티크롤링과 회피법

3. 선택자 심화와 수집기 완성하기

4. url과 요청값 이해하기

==========

1. 오픈소스 패키지

==========

1-2. 데이터 수집기 완성하기

- Requests를 활용한 코드

week3_1.py

```
1  import requests
2  from bs4 import BeautifulSoup
3
4  raw = requests.get("https://tv.naver.com/r")
5  print(raw)
```

1-2. 데이터 수집기 완성하기

- BeautifulSoup4을 활용한 코드

week3_1.py

```
import requests
from bs4 import BeautifulSoup

raw = requests.get("https://tv.naver.com/r")
# print(raw.text)
# 소스코드 출력부분을 주석처리해줍시다!

html = BeautifulSoup(raw.text, "html.parser")
print(html)
```

#1 컨테이너 수집 → #2 영상 별 데이터 수집 → #3 수집 반복

week3_1.py

```
import requests
from bs4 import BeautifulSoup

raw = requests.get("https://tv.naver.com/r")
html = BeautifulSoup(raw.text, "html.parser")

clips = html.select("div.inner")
# print(clips[0])
# 컨테이너 소스코드는 주석처리해줍니다.

for cl in clips:
    title = cl.select_one("dt.title")
    chn = cl.select_one("dd.chn")
    hit = cl.select_one("span.hit")
    like = cl.select_one("span.like")

    print(title.text.strip())
    print(chn.text.strip())
    print(hit.text.strip())
    print(like.text.strip())
```

==========

2. 안티 크롤링과 회피법

week3_2.py

```
import requests
from bs4 import BeautifulSoup

raw = requests.get("https://search.naver.com/search.naver?where=news&query=코알라",
                   headers={'User-Agent':'Mozilla/5.0'})
html = BeautifulSoup(raw.text, "html.parser")
```

==========

3. 네이버 뉴스 기사 데이터 수집

#1 컨테이너 수집 → #2 기사 별 기사제목과 언론사 데이터 수집 → #3 수집 반복

week3_2.py

```
import requests
from bs4 import BeautifulSoup

raw = requests.get("https://search.naver.com/search.naver?where=news&query=코알라",
                   headers={'User-Agent':'Mozilla/5.0'})
html = BeautifulSoup(raw.text, "html.parser")

articles = html.select("ul.type01 > li")

# 리스트를 사용한 반복문으로 모든 기사에 대해서 제목/언론사 출
for ar in articles:
    title = ar.select_one("a._sp_each_title").text
    source = ar.select_one("span._sp_each_source").text

    print(title, source)
```

4. URL과 요청값 이해하기

URL 요청값의 기본구조

==========

where=카테고리	query=검색어	start=기사번호
해당하는 카테고리의 검색결과를 보여줍니다.	해당하는 검색어의 검색결과를 보여줍니다.	해당하는 기사번호의 기사부터 보여줍니다.
where=news	*query=코알라*	*start=11*

==========

```
import requests
from bs4 import BeautifulSoup

for n in range(1, 100, 10):
    raw = requests.get("https://search.naver.com/search.naver?where=news&query=코알라&start="+str(n), headers={'User-Agent':'Mozilla/5.0'})
    html = BeautifulSoup(raw.text, "html.parser")
    articles = html.select("ul.type01 > li"

    for ar in articles:
        title = ar.select_one("a._sp_each_title").text
        source = ar.select_one("span._sp_each_source").text
        print(title, source)
```