

네이버 연관 검색어 수집

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

In [1]:

```
import pandas as pd
import requests
from bs4 import BeautifulSoup
```

In [2]:

```
# 1. 웹페이지 분석 : URL
```

In [3]:

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

<https://search.naver.com/search.naver?query=삼성전자> (<https://search.naver.com/search.naver?query=삼성전자>)

In [4]:

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

In [5]:

```
response = requests.get(url)
response
```

Out[5]:

<Response [200]>

In [6]:

```
response.text[:250]
```

Out[6]:

```
'<!doctype html> <html lang="ko"> <head> <meta charset="utf-8"> <meta name="referrer" content="always"> <meta name="format-de
tection" content="telephone=no,address=no,email=no"> <meta name="viewport" content="width=device-width,initial-scale=1.0,maxi'
```

In [7]:

```
# 3. str(html) > bs object
```

In [8]:

```
dom = BeautifulSoup(response.text, "html.parser")
type(dom)
```

Out[8]:

bs4.BeautifulSoup

In [9]:

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

In [10]:

```
# 10개의 엘리먼트들 선택
# select() : list(Tag, Tag)
# select_one() : Tag
elements = dom.select(".lst_related_srch > .item")
len(elements)
```

Out[10]:

10

In [11]:

```
element = elements[0]
keyword = element.select_one(".tit").text
keyword
```

Out[11]:

'삼성전자추가'

In [12]:

```
link = element.select_one("a").get("href")
link
```

Out[12]:

'?where=nexearch&query=%EC%82%BC%EC%84%B1%EC%A0%84%EC%9E%90%EC%A3%BC%EA%B0%80&ie=utf8&sm=tab_she&qdt=0'

In [13]:

```
element.text.strip()
```

Out[13]:

'삼성전자추가'

In [14]:

```
# 각각의 엘리먼트에서 text 데이터 수집
```

In [15]:

```
keywords = []
for element in elements:
    keyword = element.text.strip()
    keywords.append(keyword)
print(keywords)
```

['삼성전자추가', '삼성전자 배당금', '삼성전자 성과급', '삼성전자 주식', '삼성전자 세일 페스타', '삼성전자서비스', '오늘 삼성전자 추가', '삼성 전자레인지', '삼성전자 배당금 지급일', '삼성전자 opi']

In [16]:

```
keywords = [element.text.strip() for element in elements]
print(keywords)
```

['삼성전자추가', '삼성전자 배당금', '삼성전자 성과급', '삼성전자 주식', '삼성전자 세일 페스타', '삼성전자서비스', '오늘 삼성전자 추가', '삼성 전자레인지', '삼성전자 배당금 지급일', '삼성전자 opi']

In [17]:

```
# 5. str(text) > DataFrame
```

In [18]:

```
df = pd.DataFrame({"keywords": keywords})
df["query"] = query
df.tail(2)
```

Out[18]:

	keywords	query
8	삼성전자 배당금 지급일	삼성전자
9	삼성전자 opi	삼성전자

In [19]:

```
# 현재 시간 데이터 추가
```

In [20]:

```
from datetime import datetime
```

In [21]:

```
now = datetime.now()
now = now.strftime("%Y-%m-%d %H:%M")
now
```

Out[21]:

```
'2023-02-05 23:02'
```

In [22]:

```
df["date_time"] = now
df.tail(2)
```

Out[22]:

	keywords	query	date_time
8	삼성전자 배당금 지급일	삼성전자	2023-02-05 23:02
9	삼성전자 opi	삼성전자	2023-02-05 23:02

In [23]:

```
# query를 입력하면 데이터 프레임을 출력하는 함수
def naver_relate_keyword(query):

    url = f"https://search.naver.com/search.naver?query={query}"
    response = requests.get(url)
    dom = BeautifulSoup(response.text, "html.parser")
    elements = dom.select(".lst_related_srch > .item")
    keywords = [element.text.strip() for element in elements]

    df = pd.DataFrame({"keywords": keywords})
    df["query"] = query

    now = datetime.now()
    now = now.strftime("%Y-%m-%d %H:%M")
    df["date_time"] = now

    return df
```

In [24]:

```
query = "삼성전자"
df = naver_relate_keyword(query)
df.tail(2)
```

Out[24]:

	keywors	query	date_time
8	삼성전자 배당금 지급일	삼성전자	2023-02-05 23:02
9	삼성전자 opi	삼성전자	2023-02-05 23:02

In [25]:

```
dfs = []
queries = ["삼성전자", "LG전자"]

for query in queries:
    print(query, end=" ")
    df = naver_relate_keyword(query)
    dfs.append(df)

result = pd.concat(dfs, ignore_index=True)
result
```

삼성전자 LG전자

Out[25]:

	keywors	query	date_time
0	삼성전자주가	삼성전자	2023-02-05 23:02
1	삼성전자 배당금	삼성전자	2023-02-05 23:02
2	삼성전자 성과급	삼성전자	2023-02-05 23:02
3	삼성전자 주식	삼성전자	2023-02-05 23:02
4	삼성전자 세일 페스타	삼성전자	2023-02-05 23:02
5	삼성전자서비스	삼성전자	2023-02-05 23:02
6	오늘 삼성전자 주가	삼성전자	2023-02-05 23:02
7	삼성 전자레인지	삼성전자	2023-02-05 23:02
8	삼성전자 배당금 지급일	삼성전자	2023-02-05 23:02
9	삼성전자 opi	삼성전자	2023-02-05 23:02
10	lg전자 서비스센터	LG전자	2023-02-05 23:02
11	lg전자 주가	LG전자	2023-02-05 23:02
12	lg전자 베스트샵	LG전자	2023-02-05 23:02
13	lg전자 성과급	LG전자	2023-02-05 23:02
14	lg 전자레인지	LG전자	2023-02-05 23:02
15	lg전자 서비스센터 전화번호	LG전자	2023-02-05 23:02
16	lg전자 고객센터	LG전자	2023-02-05 23:02
17	lg전자렌지	LG전자	2023-02-05 23:02
18	lg전자 실적	LG전자	2023-02-05 23:02
19	lg전자 렌탈	LG전자	2023-02-05 23:02