

## 네이버 주식 정보 가져와 보기-실습2

### 정보 가져오기에 앞서 페이지 확인해 보기

```
https://finance.naver.com/robots.txt
```

```
User-agent: *
Disallow: /
User-agent: yeti
Disallow: /
Allow: /sise/
Allow: /research/
Allow: /marketindex/
Allow: /fund/
Disallow: /fund/news/
Disallow: /marketindex/news/
```

In [7]:

```
from urllib.request import urlopen
from bs4 import BeautifulSoup
```

### 관심 종목의 주식 정보를 for문을 이용하여 가져와 보기

- 예외 케이스
  - 태그가 보이지 화면상에서는 보이지 않을 때
  - 소스 코드 보기를 이용하여 찾는 방법

In [15]:

```
## 삼성전자 : 005930
## 카카오 : 035720
## 유한양행 : 000100
```

In [16]:

```
my_stock = ['삼성전자', '카카오', '유한양행', 'NAVER']
my_stock_code = ['005930', '035720', '000100', '035420']
```

In [17]:

```
chk_url = 'https://finance.naver.com/item/main.nhn?code=000100'
page = urlopen(chk_url)
soup = BeautifulSoup(page, 'lxml')
```

### 제목 가져오기

In [18]:

```
soup.title
```

Out[18]:

```
<title>유한양행 : 네이버 금융</title>
```

- 먼저 정보가 포함되는 부분에 대한 부분을 찾는다.
- 태그가 보이지 않아, 웹 페이지 소스 보기를 통해 원하는 정보를 찾는다.
- 정보가 포함된 부분의 class 'no\_today'을 검색
  - 해당 되는 부분에 우리가 원하는 정보를 가져올 수 있는지 확인한다.

In [19]:

```
soup.find('p', class_='no_today').find('span', class_='blind').text
```

Out[19]:

```
'65,000'
```

In [20]:

```
base_url = 'https://finance.naver.com/item/main.nhn?code='
```

In [21]:

```
com = []    # 회사명
price_c = [] # 가격

for code in my_stock_code:
    cnt = 0    # 회사명 선택을 위한 count
    all_url = base_url + code
    page = urlopen(all_url)
    soup = BeautifulSoup(page, 'xml')

    ## 현재가
    tmp = soup.find('p', class_='no_today')
    price = tmp.find('span', class_='blind').text

    com.append(my_stock[cnt])
    price_c.append(price)

    cnt += 1

print(com)
print(price_c)
```

```
['삼성전자', '삼성전자', '삼성전자', '삼성전자']
['60,400', '371,500', '65,000', '292,000']
```

In [22]:

```
import pandas as pd
```

In [23]:

```
dic_dat = {'회사명':com, '주가':price_c}  
dat = pd.DataFrame( dic_dat )  
dat.to_csv("date_0118.xlsx", index=False)
```

In [25]:

```
# 파일 생성 확인 (리눅스 환경 !ls, 윈도우 !dir로 확인 가능)  
!dir *date*
```

C 드라이브의 볼륨에는 이름이 없습니다.  
볼륨 일련 번호: 2C14-882B

C:\Users\WHP\Documents\GitHub\AI\_innovation\part02\_library 디렉터리

2020-10-12	오후 09:28	112	date_0118.xlsx
	1개 파일		112 바이트
	0개 디렉터리	78,452,830,208	바이트 남음

In [ ]: