

네이버 데이터 랩에서 키워드 가져오기 ¶

- 파일명 : datalab_word.xlsx
- 시트명은 입력을 받음.
- 데이터 랩에서 10대부터 전체 목록에 대한 급상승 검색어를 가지고 온다.

In [1]:

```
filename = "datalab_word.xlsx"
```

In [2]:

```
# 해당 시트명 입력
sheetname = input("시트명을 입력해 주세요:")
```

시트명을 입력해 주세요:toto

In [3]:

```
# -*- coding: utf-8 -*-
import os
os.getcwd()

from selenium import webdriver
from bs4 import BeautifulSoup

driver = webdriver.Chrome('chromedriver_85')

## http://www.opinet.co.kr/searRgSelect.do 접속
driver.get('https://datalab.naver.com/keyword/realtimeList.naver')
```

In [4]:

```
## 10대  //*[@id="content"]/div/div[2]/div[1]/div[2]/div/div/div/ul/li[1]/a
## 20대  //*[@id="content"]/div/div[2]/div[1]/div[2]/div/div/div/ul/li[2]/a
## 30대  //*[@id="content"]/div/div[2]/div[1]/div[2]/div/div/div/ul/li[3]/a
# ...
## 전체  //*[@id="content"]/div/div[2]/div[1]/div[2]/div/div/div/ul/li[6]/a
```

In [5]:

```
import time
```

In [6]:

```
sel_10 = driver.find_element_by_xpath('//*[@id="content"]/div/div[2]/div[1]/div[2]/div/div/div/ul/li[1]/a')
sel_20 = driver.find_element_by_xpath('//*[@id="content"]/div/div[2]/div[1]/div[2]/div/div/div/ul/li[2]/a')
sel_30 = driver.find_element_by_xpath('//*[@id="content"]/div/div[2]/div[1]/div[2]/div/div/div/ul/li[3]/a')
sel_40 = driver.find_element_by_xpath('//*[@id="content"]/div/div[2]/div[1]/div[2]/div/div/div/ul/li[4]/a')
sel_50 = driver.find_element_by_xpath('//*[@id="content"]/div/div[2]/div[1]/div[2]/div/div/div/ul/li[5]/a')
sel_60 = driver.find_element_by_xpath('//*[@id="content"]/div/div[2]/div[1]/div[2]/div/div/div/ul/li[6]/a')
```

In [7]:

```
sel_10.click()
```

In [8]:

```
soup = BeautifulSoup(driver.page_source, 'lxml')
age_10 = soup.find_all("span", class_ = "item_title")
age_10
```

Out[8]:

```
[<span class="item_title">코로나 1단계</span>,
 <span class="item_title">자가진단</span>,
 <span class="item_title">사회적 거리두기 1단계</span>,
 <span class="item_title">잉글랜드 벨기에</span>,
 <span class="item_title">프랑스 포르투갈</span>,
 <span class="item_title">피시방 청소년 출입</span>,
 <span class="item_title">UEFA 네이션스리그</span>,
 <span class="item_title">등교</span>,
 <span class="item_title">경기도 학생 자가진단</span>,
 <span class="item_title">학생 자가진단</span>,
 <span class="item_title">ebs 온라인 클래스</span>,
 <span class="item_title">이비에스 온라인 클래스</span>,
 <span class="item_title">청주 날씨</span>,
 <span class="item_title">대전 날씨</span>,
 <span class="item_title">2020 10월 모의고사</span>,
 <span class="item_title">월드컵 8강</span>,
 <span class="item_title">수만휘</span>,
 <span class="item_title">유웨어플라이</span>,
 <span class="item_title">학교 등교</span>,
 <span class="item_title">인생존망</span>]
```

In [9]:



```
pop_10 = []
for i in age_10:
    text = i.text
    pop_10.append(text)

pop_10
```

Out[9]:

```
['코로나 1단계',
 '자가진단',
 '사회적 거리두기 1단계',
 '잉글랜드 벨기에',
 '프랑스 포르투갈',
 '피시방 청소년 출입',
 'UEFA 네이션스리그',
 '등교',
 '경기도 학생 자가진단',
 '학생 자가진단',
 'ebs 온라인 클래스',
 '이비에스 온라인 클래스',
 '청주 날씨',
 '대전 날씨',
 '2020 10월 모의고사',
 '월드컵 8강',
 '수만취',
 '유웨이어플라이',
 '학교 등교',
 '인생존망']
```

전체를 가져와 보기

In [10]:



```
import time
```

In [11]:

```

base_xpath1 = '//*[@id="content"]/div/div[2]/div[1]/div[2]/div/div/div/ul/li['
base_xpath2 = ']/a'

all_word = []
for i in range(1,7):
    all_xpath = base_xpath1 + str(i) + base_xpath2
    sel_btn = driver.find_element_by_xpath(all_xpath)
    sel_btn.click()

    soup = BeautifulSoup(driver.page_source, 'lxml')
    age_data = soup.find_all("span", class_ = "item_title")

    pop_10 = []
    for i in age_data:
        text = i.text
        pop_10.append(text)
    time.sleep(2)
    all_word.append(pop_10)
    print(pop_10)

all_word

```

['코로나 1단계', '자가진단', '사회적 거리두기 1단계', '잉글랜드 벨기에', '프랑스 포르투갈', '피시방 청소년 출입', 'UEFA 네이션스리그', '등교', '경기도 학생 자가진단', '학생 자가진단', 'ebs 온라인 클래스', '이비에스 온라인 클래스', '청주 날씨', '대전 날씨', '2020 10월 모의고사', '월드컵 8강', '수만취', '유웨이어플라이', '학교 등교', '인생존망']

['사회적 거리두기 1단계', '코로나 1단계', '대전 날씨', '청주 날씨', '국가비', '똥죽똥죽', '서울 날씨', '똥죽', '오늘날씨', '코로나 1단계 기준', '프랑스 포르투갈', '환불원정대', '1호가 될 순 없어', '가짜사나이 2기', 'UEFA 네이션스리그', '족발의달인', '사회적 거리두기 1단계 결혼식', '노래방', 'savage', '미세먼지']

['사회적 거리두기 1단계', '국가비', '코로나 1단계', '똥죽똥죽', '똥죽', '똥죽 프랜차이즈', '1호가 될 순 없어', '족발의달인', '환불원정대', '구미날씨', '대전 날씨', '구해줘 홈즈', '엄기준', '복지로', '나달', '미세먼지', '사회적 거리두기 1단계 결혼식', '류호정', '코로나 1단계 기준', '돼지열병']

['사회적 거리두기 1단계', '국가비', '똥죽 프랜차이즈', '코로나 1단계', '똥죽', '똥죽똥죽', '1호가 될 순 없어', '구해줘 홈즈', '족발의달인', '코로나 1단계 기준', '학생 자가진단', '돼지열병 관련주', '환불원정대', '복지로', '엄기준', '날씨예보', '미씽 그들이있었다', '오늘의날씨', '등교', '환불원정대 돈터치미']

['사회적 거리두기 1단계', '코로나 1단계', '국가비', '코로나 1단계 기준', 'lpga', '날씨예보', '똥죽 프랜차이즈', '돼지열병 관련주', '라파엘 나달', '오늘 서울 날씨',

In [12]:

```

dict_val = {"10대":all_word[0], "20대":all_word[1], "30대":all_word[2],
            "40대":all_word[3], "50대":all_word[4], "전체":all_word[5]}

```

In [13]:



```
import pandas as pd
dat = pd.DataFrame(dict_val)
dat
```

Out[13]:

	10대	20대	30대	40대	50대	전체
0	코로나 1단계	사회적 거리두기 1단계	사회적 거리두기 1단계	사회적 거리두기 1단계	사회적 거리두기 1단계	사회적 거리두기 1단계
1	자가진단	코로나 1단계	국가비	국가비	코로나 1단계	코로나 1단계
2	사회적 거리두기 1단계	대전 날씨	코로나 1단계	뎡죽 프랜차이즈	국가비	국가비
3	잉글랜드 벨기에	청주 날씨	뎡죽뎡죽	코로나 1단계	코로나 1단계 기준	뎡죽뎡죽
4	프랑스 포르투갈	국가비	뎡죽	뎡죽	lpga	뎡죽
5	피시방 청소 년 출입	뎡죽뎡죽	뎡죽 프랜차이즈	뎡죽뎡죽	날씨예보	뎡죽 프랜차이즈
6	UEFA 네이션스리그	서울 날씨	1호가 될 순 없어	1호가 될 순 없어	뎡죽 프랜차이즈	대전 날씨
7	등교	뎡죽	죽발의달인	구해줘 홈즈	돼지열병 관련주	1호가 될 순 없어
8	경기도 학생 자가진단	오늘날씨	환불원정대	죽발의달인	라파엘 나달	코로나 1단계 기준
9	학생 자가진단	코로나 1단계 기준	구미날씨	코로나 1단계 기준	오늘 서울 날씨	청주 날씨
10	ebs 온라인 클래스	프랑스 포르투갈	대전 날씨	학생 자가진단	오늘의날씨	죽발의달인
11	이비에스 온라인 클래스	환불원정대	구해줘 홈즈	돼지열병 관련주	뎡죽	구해줘 홈즈
12	청주 날씨	1호가 될 순 없어	엄기준	환불원정대	정태욱	환불원정대
13	대전 날씨	가짜사나이 2기	복지로	복지로	구해줘 홈즈	오늘의날씨
14	2020 10월 모 의고사	UEFA 네이션스리그	나달	엄기준	아프리카 돼지열병관련주	프랑스 포르투갈
15	월드컵 8강	죽발의달인	미세먼지	날씨예보	인간극장	나달
16	수만휘	사회적 거리두기 1단계 결혼식	사회적 거리두기 1단계 결혼식	미씽 그들이 있었다	옵티머스	엄기준
17	유웨이어플라이	노래방	류호정	오늘의날씨	서울제약	UEFA 네이션스리그
18	학교 등교	savage	코로나 1단계 기준	등교	1호가 될 순 없어	잉글랜드 벨기에
19	인생존망	미세먼지	돼지열병	환불원정대 돈터치미	2차 재난지원금 신청	복지로

In [14]:



```
# dat.to_excel("output.xlsx", sheet_name='Sheet_name_1', index=False)
dat.to_excel(filename, sheet_name=sheetname, index=False)
```

Copyright 2020 LIM DJ(예영Edu Co.) all rights reserved.

교육용으로 작성된 것으로 배포 및 복제시에 사전 허가가 필요합니다.

In []:

