

230228 네이버웹툰 크롤링

지니뮤직 크롤링

```
from selenium import webdriver          # 웹 브라우저 컨트롤 (크롬)
from bs4 import BeautifulSoup as bs    # 데이터 분석을 용이하게 정제
import pandas as pd                    # 데이터 분석 관련 모듈

driver = webdriver.Chrome('chromedriver.exe') # 버전 주의
url = 'https://www.genie.co.kr/chart/top200'
driver.get(url)

txt = driver.page_source                # 이때 읽어온 데이터는 그냥 글자
html = bs(txt)

songs = html.select('tbody > tr')

song_data = []
rank = 1

for song in songs:
    title = song.select('a.title')[0].text.strip()
    singer = song.select('a.artist')[0].text.strip()

    song_data.append(['Genie', rank, title, singer])
    rank += 1

url = 'https://www.genie.co.kr/chart/top200?ditc=D&ynd=20230228&hh=14&rtm=Y&pg=2'
driver.get(url)

txt = driver.page_source                # 이때 읽어온 데이터는 그냥 글자
html = bs(txt)

songs = html.select('tbody > tr')

for song in songs:
    title = song.select('a.title')[0].text.strip()
    singer = song.select('a.artist')[0].text.strip()

    song_data.append(['Genie', rank, title, singer])
    rank += 1

df = pd.DataFrame(song_data, columns = ['서비스', '순위', '타이틀', '가수'])
df
```

네이버 현재 온도 크롤링

```
## 네이버 현재 날씨 읽어오기

import requests
txt = requests.get('https://weather.naver.com/')
html = bs(txt.text)

html.select('strong.current')[0].text.strip()
temp = html.select('strong.current')[0].text.strip()
temp[5:]
```

네이버 웹툰 썸네일

```
# 사진 다운로드
from urllib.request import urlretrieve
import re

li_list = []

for data in tlist:
    # 제목과 썸네일 영역 추출
    li_list.extend(data.findAll('li')) # 해당 부분을 찾아 병합

# pprint(li_list)

# 썸네일과 제목 추출
for li in li_list:
    img = li.find('img')
    title = img['title']
    img_src = img['src']
    #print(title, img_src)
    #특수 문자 제거
    title = re.sub('[^0-9a-zA-Zㄱ-힣]', '', title)

    urlretrieve(img_src, title+'.jpg')
```