

# Crawling Day 1 230227

## Crawling Basic

```
import requests # 웹 페이지 읽어오는 모듈
from bs4 import BeautifulSoup as bs # 읽어온 웹페이지 정제
```

본인이 사용하는 크롬 버전을 확인할것. 아래 파일은 **110.0.5481.77 버전용 파**

[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/f67170d5-6c77-4714-b53b-529691532c27/chromedriver\\_win32.zip](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/f67170d5-6c77-4714-b53b-529691532c27/chromedriver_win32.zip)

## - Melon

```
#####
#                               Melon 차트 읽어오기                               #
#####

# url = "https://www.melon.com/chart/"
# txt = requests.get(url) # 웹페이지 그냥 글자로 읽어오기
# html = bs(txt.text) # 읽어온 그냥 글자를 의미가 있는 html 로 변환 bs

# print(html)

# 멜론도 위에서의 벅스처럼 크롤링을 시도하였으나 사이트에서 크롤링을 막고 있다.
# 따라서 다른 방식으로 크롤링을 시도해야 한다.

from selenium import webdriver # 위에서의 request와는 다른 방식

driver = webdriver.Chrome('chromedriver.exe')
driver.get("https://www.melon.com/chart/index.htm")

txt = driver.page_source
html = bs(txt) # 벅스 문법과 약간 다름
```

```
print(html)
```

`.select()` 와 `.find()`

두가지 모두 태그를 찾아주는 메서드이다. 약간 다른점이 있다.

`.select()` 괄호 안의 조건에 해당하는 태그를 모두 추출

`.find()_all` 괄호 안의 조건에 해당하는 태그를 모두 추출

`.select_one()` 괄호 안의 조건에 해당하는 태그를 하나만 추출

`.find()` 괄호 안의 조건에 해당하는 태그를 하나만 추출

**가장 큰 차이점은 `.find()`는 참/거짓 조건을 넣어서 필터링 할수 있다.**