

# 웹 크롤링 - selenium 기초

⌚ 작성 일시	@2023년 2월 28일 오후 8:26
▼ 강 의 번 호	Python
▼ 유 형	
📎 자 료	
☑ 복 습	<input type="checkbox"/>
≡ 학 습 소 스 출 처 1	<a href="https://www.youtube.com/watch?v=yQ20jZwDjTE&amp;list=PLMsa_0kAjjrd8hYYCwbAuDsXZmHpqHvIV&amp;index=4">https://www.youtube.com/watch?v=yQ20jZwDjTE&amp;list=PLMsa_0kAjjrd8hYYCwbAuDsXZmHpqHvIV&amp;index=4</a>
≡ 학 습 소 스 출 처 2	<a href="https://greeksharifa.github.io/references/2020/10/30/python-selenium-usage/#xpath로-요소-찾기">https://greeksharifa.github.io/references/2020/10/30/python-selenium-usage/#xpath로-요소-찾기</a>
📅	

## 파이썬 selenium 이란?

파이썬 코드로 브라우저를 마치 사용자가 직접 조작하는 것처럼 사용할 수 있게 하는 라이브러리이다. requests 라이브러리는 Postman을 이용한 API 통신과는 다르게 selenium은 사용자가 브라우저에서 클릭을 통한 조작을 하는듯이 사용할 수 있다.

## 왜 파이썬 selenium을 사용하였는가?

requests 라이브러리를 임포트하여 크롤링을 할 당시, 크롤링하려던 사이트에서 403에러를 응답받아서 사용할 수 없었다. 그러므로 다른 환경에서 접근할 필요가 있어서 selenium을 선택하였다.

## selenium 시작해보기

### selenium 설치

```
pip install selenium
```

### driver 설치

셀레니움을 사용하기 위해선 라이브러리 이외에 드라이버를 받아야 한다.

먼저 현재 자신의 브라우저 버전을 확인한다.

나는 크롬을 사용하고 있기 때문에 브라우저에

```
chrome://version/
```


를 치고 보면

Chrome: 110.0.5481.177 공식 빌드) (64비트) (cohort: M110\_Stable\_RampUp) 📄

이에 해당하는

Downloads - ChromeDriver - WebDriver for Chrome

WebDriver for Chrome

 <https://sites.google.com/a/chromium.org/chromedriver/downloads>

크롬 드라이버

다운 받은 exe 파일을 원하는 위치에 둔다( 이왕이면 코드를 사용하려는 python 파일과 같은 위치를 추천)

## 코드에서 사용

```
import selenium

URL = 'https://여기에크롤링하려는경로추가'

# 현재 파이썬 파일과 같은 위치에 설치받은 파일이 있다면 아래와같이
# 다르다면 경로를 맞춰줘야한다.
# 이 코드를 이용하면 빈 크롬브라우저를 불러온다
driver = webdriver.Chrome('./chromedriver.exe')

# 현재 브라우저를 url 경로로 이동시킨다.
driver.get(url=URL)
```

## 현재 브라우저의 원하는 요소 받아오기

먼저 요소를 찾기 위한 라이브러리를 임포트받아온다.

```
from selenium.webdriver.common.by import By

# 아이디 값으로 찾기
elem = driver.find_element(By.ID, '찾으려는 아이디 값')
```

<b>element(단일) , elements(복수)</b>	<b>설명</b>
By.ID	태그의 id값으로 추출
By.NAME	태그의 name값으로 추출
By.XPATH	태그의 경로로 추출
By.LINK_TEXT	링크 텍스트값으로 추출
By.PARTIAL_LINK_TEXT	링크 텍스트의 자식 텍스트 값을 추출
By.TAG_NAME	태그 이름으로 추출
By.CLASS_NAME	태그의 클래스명으로 추출
By.CSS_SELECTOR	css선택자로 추출

출처 : <https://pythonblog.co.kr/coding/23/>