**파이썬 셀레늄/뷰티풀숲 웹크롤러 정리**

**0. 크롬창 띄우기**

driver = webdriver.Chrome("chromedriver.exe") # 크롬 드라이버 로드

driver.get("https://www.instagram.com(url)") # url 입력

**1. 엘리먼트 찾기**

# 엘리먼트(element) : 홈페이지를 이루고 있는 다양한 요소들을 부르는 말.

# 엘리먼트 찾기 : 오른쪽 화면의 상단의 네모칸 화살표 있는 버튼을 누릅니다.

● id, name, class 있을 때는 이 3개로 찾는다.

element = driver.find\_element\_by\_id(‘id’)

element = driver.find\_element\_by\_name(‘name’)

element = driver.find\_element\_by\_css\_selector(‘#tsf .gLFyf.gsfi’)

● id, name, class없을 때 xpath 이용한다.

element = driver.find\_element\_by\_xpath("""//\*[@id="snb"]/div/ul/li[1]/a""").click( ) # 정렬 버튼의 xpath 클릭

● 기타

element = driver.find\_element\_by\_class\_name(‘some\_class\_name’)

element = driver.find\_element\_by\_tag\_name(‘h1’)

**2. 액션**

● driver 컨트롤하는 명령어들

driver.back() # 뒤로 가기

driver.foward() # 앞으로 가기

driver.switch\_to.window(driver.window\_handles[-1]) # 최근 열린 탭으로 전환

driver.close() # 현재 탭 닫기

driver.switch\_to.window(driver.window\_handles[0]) # 맨 처음 탭으로 변경(0번 탭)

alert = driver.switch\_to.alert # alert창-“확인” 누르기

alert.accept() #

time.sleep(1) #

● 버튼 클릭

|  |
| --- |
| button = driver.find\_element\_by\_css\_selector(ele)  AC(driver).move\_to\_element(button).click().perform() # 버튼 클릭 |

driver.find\_element\_by\_id("\_search\_option\_btn").click( ) # 서치 옵션 버튼 클릭

● 인풋 받기

input('사용자에게 받을 내용 : ') # 인풋 받아서 변수로 사용 가능

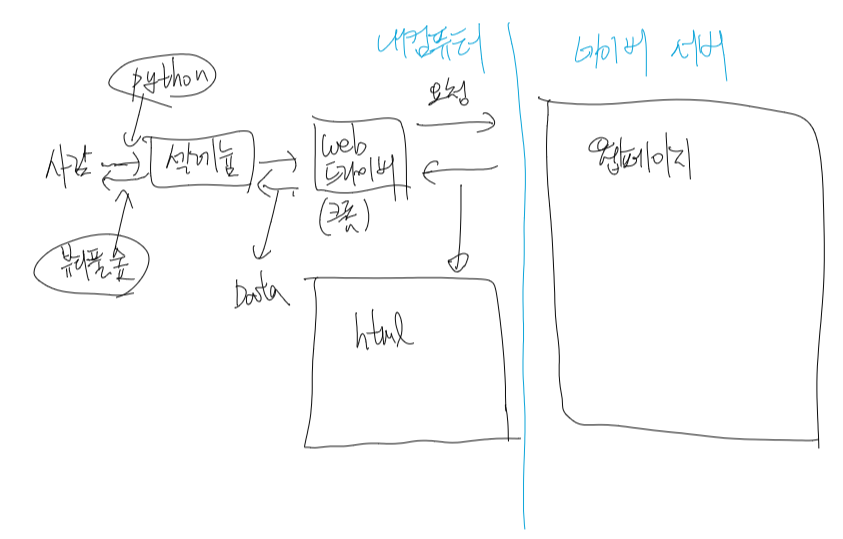
● 시간

time.sleep(2) # 2초 기다려라

● 텍스트 제거 및 입력

|  |
| --- |
| **# 한글자씩 입력창에 입력**  s\_date = driver.find\_element\_by\_xpath("""//\*[@id="blog\_input\_period\_begin"]""") # 날짜 입력할 부분 선택  s\_date.clear( ) # 날짜 입력 부분에 기존에 입력되어 있던 날짜를 제거  start\_date = “2019-04-05”  for c in start\_date:  s\_date.send\_keys(c)  time.sleep(0.3) |

**● 크롤링이란**



**1. 파이썬 환경 준비**

● 사용하는 도구

#파이썬(3.5 or 3.6) #셀레늄 #뷰티풀숲 #크롬드라이버

● 크롤링할 데이터

# 이미지 # 텍스트

● 설치

# 파이썬 설치 (버전 : 3.5.4) : www.python.org

# 파이썬 라이브러리 설치 (명령 프롬프트 : cmd 검색)

- pip install bs4

- pip install selenium

※ 셀레늄

\* selenium(셀레니움) 라이브러리는 본래 웹사이트의 테스트를 위한 목적으로 제작되었지만 자바스 크립트 같이 동적으로 구성된 페이지를 크롤링하거나 사이트 캡쳐 등을 할 때에도 사용됩니다. 셀레니움 설치는 selenium 웹사이트에서 파일을 다운로드 받아 직접 설치하거나 pip 같은 패키지 관리자로 간단하게 설치할 수 있으며 대부분 간편한 pip를 이용한 설치 방법을 사용합니다.

\* selenium 은 웹 브라우저를 제어해서 크롤링을 진행합니다. 그래서 selenium 을 사용할 때 웹 브라우저를 지정해 주어야 하는데 여기서는 구글사의 크롬을 이용하겠습니다. 크롬을 사용하려면 실습을 진행하는 컴퓨터에 크롬이 설치되어 있어야 하며 selenium 이 사용할 크롬 드라이버도 있어야 합니다. 크롬 드라이버는 아래 링크에서 받아 주세요

chromedriver 2.40 설치 : https://chromedriver.storage.googleapis.com/index.html?path=2.40/

<크롤링시 주의할 점]>

크롤링 시 특정 사이트에서는 비정상적 접근으로 인식하여 해당 아이피를 차단할 수도 있습니다. 과도한 크롤링은 해당 회사나 단체에 피해를 줄 수 있으니 충분한 검토 후 작업하시기 바랍니다. 이 글은 크롤링 하는 원리와 방법에 대해 설명하는게 목적입니다. 사용자의 과도한 크롤링 행위로 타 회사나 기관에 피해를 줄 경우 모든 책임은 해당 행위를 수행한 사람에게 있음을 미리 밝혀 둡니다.

