



■ 파이썬 웹크롤링 과 자동화에 대한 A to Z

(/book/4706)

1. 준비단계 - 스트레칭

1.1 마음가짐 - 무엇이든 할 수 있게 하는 힘

1.2 목 돌리기 - 필요한 파이썬 문법만 알고가자

2. 기초단계 - 유산소운동

2.1 마우스 자동화 - pyautogui 사용법 (1)

2.2 키보드 자동화 - pyautogui 사용법 (2)

2.3 메시지 박스 - pyautogui 사용법(3)

2.4 이미지로 좌표찾기 - pyautogui 사용법 (4)

2.5 사이트 정보 가져오기 - requests 사용법

2.6 사이트 정보 추출하기 - beautifulsoup 사용법 (1)

2.7 사이트 정보 추출하기 - beautifulsoup 사용법 (2)

2.8 사이트 자동화하기 - selenium 사용법(1)

3. 응용단계 - 웨이트 트레이닝

3.0 파이썬 엑셀 다루기 - openpyxl 사용법

3.1 웹스크래핑 예제(1) - 네이버금융 실시간 검색 순위 스크래핑 후 엑셀에 저장하기

3.2 웹스크래핑 예제(2) - 네이버금융 실시간 주가 크롤링하기

Published with WikiDocs (/)



■ 파이썬 웹크롤링 과 자동화에 대한 A to Z (/book/4706) / 2. 기초단계 - 유산소운동 (/85383)

/ 2.8 사이트 자동화하기 - selen ... (/91474)

2.8 사이트 자동화하기 - selenium 사용법(1)

BeautifulSoup 라이브러리만으로도 다양한 사이트의 정보를 추출할 수 있습니다. 그런데, BeautifulSoup에게는 한계가 있습니다.

그 한계란 무엇일까요?

바로, "자바스크립트로 동적으로 생성된 정보는 가져올 수 없다!"입니다.

자바스크립트가 발전을 하면서, Ajax(비동기 통신) 형태로 서버와 데이터를 주고 받아 화면에 뿌려주는 사이트가 많아 졌습니다. 이러한 형식으로 데이터를 주고 받으면 url 변경이나 새로고침 없이 데이터를 가져오게 됩니다.

아마도 스크래핑을 시도하다가 데이터를 가져와야 정상인데, 아무것도 가져오지 못하는 현상을 마주친적이 있을 거예요. 대부분의 경우가 자바스크립트로 HTML을 만들어서 그렇습니다.

Selenium 라이브러리를 사용하는 이유는 다음과 같습니다.

1. 자바스크립트가 동적으로 만든 데이터를 크롤링 하기 위해
2. 사이트의 다양한 HTML 요소에 클릭, 키보드 입력 등 이벤트를 주기 위해

Selenium을 잘 활용하면, 평소에 반복적으로 하고 있는 웹상의 업무를 자동화할 수도 있습니다.

1. 자동으로 로그인하기
2. 메일보내기 자동화
3. 블로그 이웃새글 자동좋아요 누르기
4. 인스타그램 자동으로 좋아요, 댓글 작성하기
5. 등등 정말 많은 다양한 일

이제 배울 이유가 충분해 졌나요? 그럼 바로 들어가 보도록하겠습니다.





1. selenium 모듈 설치

```
- pip install selenium
```

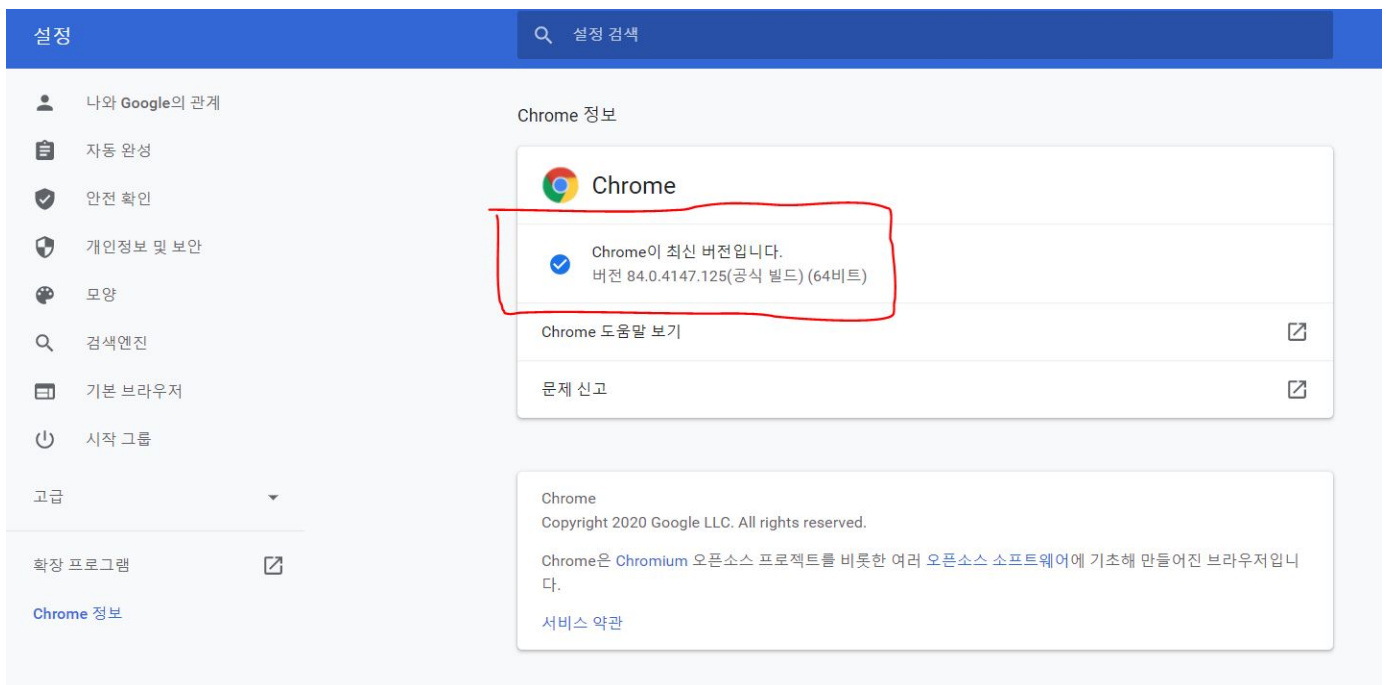
2. 크롬드라이버 설치

크롬 드라이버란 크롬 브라우저를 제어하기 위한 드라이버 입니다. 크롬 외에도 IE, 파이어폭스등 다양한 브라우저도 가능합니다. <https://chromedriver.chromium.org/downloads> 로 이동해서 자신의 크롬버전 및 운영체제와 동일한 드라이버를 다운로드 합니다.

Index of /84.0.4147.30/

Name	Last modified	Size	ETag
Parent Directory		-	
 chromedriver_linux64.zip	2020-05-28 21:05:07	5.06MB	befffb1bca07d8f4fd23213b292ef963b
 chromedriver_mac64.zip	2020-05-28 21:05:09	6.99MB	b2ff30e148ae11a78e0f13e93b29f271
 chromedriver_win32.zip	2020-05-28 21:05:11	4.63MB	3bf0e106a93382efd7a5bb3b55b182a6
 notes.txt	2020-05-28 21:05:15	0.00MB	a505de7f878e415f1b06a44935f109bf

- 크롬 버전은 다음과 같이 확인 할 수 있습니다.



설정

Chrome 정보

Chrome이 최신 버전입니다.
버전 84.0.4147.125(공식 빌드) (64비트)

Chrome 도움말 보기

문제 신고

Chrome
Copyright 2020 Google LLC. All rights reserved.
Chrome은 Chromium 오픈소스 프로젝트를 비롯한 여러 오픈소스 소프트웨어에 기초해 만들어진 브라우저입니다.
서비스 약관

- 프로젝트 폴더에 넣어준다.

```
> 2020-07-27
> 2020-08-04
> 2020-08-06
> 2020-08-17
> 과제
✓ test
  ├── chromedriver.exe
  └── selenium_test.py
```

3. selenium으로 크롬브라우저창 열기

- driver 경로를 파일경로와 같은 곳에 둘 경우

```
from selenium import webdriver
driver = webdriver.Chrome()
url = 'https://www.google.com'
driver.get(url)
```

- driver 경로를 파일경로와 다른 곳에 둘 경우

```
from selenium import webdriver
driver = webdriver.Chrome('driver의경로')
url = 'https://www.google.com'
driver.get(url)
```

4. 원하는 태그 찾기

- 요소검사를 진행해서 id나 class 또는 태그명을 확인한다.
 - driver.find_element_by_css_selector (단수)
 - driver.find_elements_by_css_selector (복수)

5. 키 입력하기

- send_keys('파이썬')

마지막 편집일시 : 2020년 8월 17일 8:36 오후

댓글 0 피드백

- 이전글 : 2.7 사이트 정보 추출하기 - beautifulsoup 사용법 (2)
- 다음글 : 3. 응용단계 - 웨이트 트레이닝

↑ TOP