

## Data Crawling 2





Selenium



# 정적 웹페이지 vs 동적 웹페이지

## 정적 웹페이지

- 웹 서버에 이미 저장된 기본 html 전송
- 데이터 자체의 변화 없는 한 고정 데이터 전달
- Ex) 홈페이지 회사 소개

## 동적 웹페이지

- 요청 정보 처리 후 제작된 html 페이지 전송
- 사용자가 요청하는 상황 등에 따라 다른 데이터 전달
- Ex) 뉴스 댓글, 상품 검색



# Selenium

정적 웹페이지

VS

동적 웹페이지

BeautifulSoup  
(페이지 내의 고정 html만 전달받음)

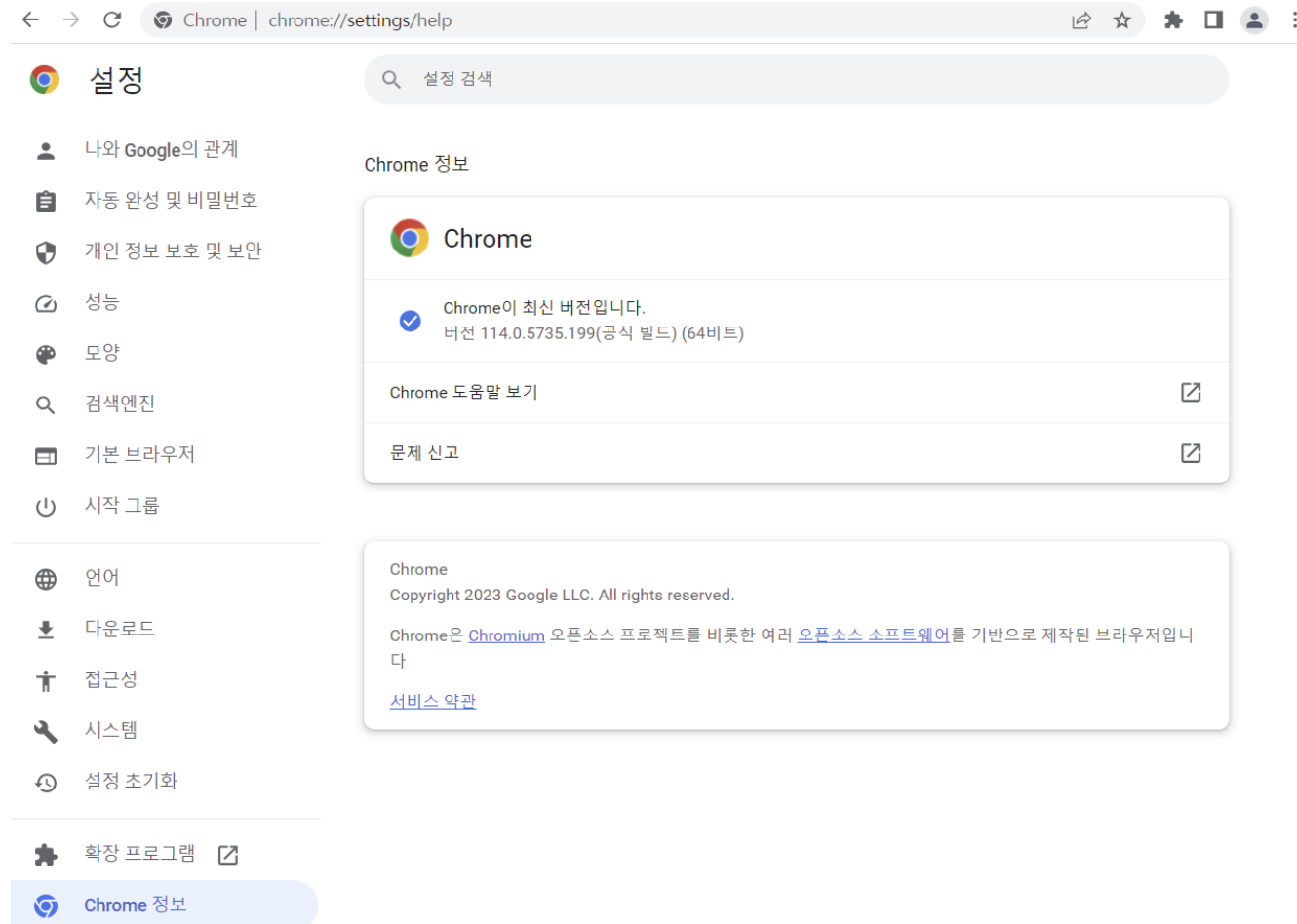
Selenium

(정적, 동적 웹페이지 모두 크롤링 가능)

\* 원래 테스트 엔지니어링을 돕는 프로그램이었기 때문!



# 웹 드라이버 설치: 크롬 브라우저 버전 확인



Chrome | chrome://settings/help

설정

나와 Google의 관계

자동 완성 및 비밀번호

개인 정보 보호 및 보안

성능

모양

검색엔진

기본 브라우저

시작 그룹

언어

다운로드

접근성

시스템

설정 초기화

확장 프로그램

Chrome 정보

Chrome 정보

Chrome

Chrome이 최신 버전입니다.  
버전 114.0.5735.199(공식 빌드) (64비트)

Chrome 도움말 보기

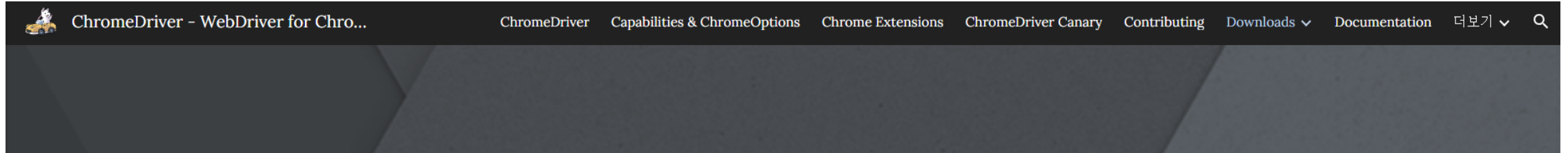
문제 신고

Chrome  
Copyright 2023 Google LLC. All rights reserved.  
Chrome은 [Chromium](#) 오픈소스 프로젝트를 비롯한 여러 [오픈소스 소프트웨어](#)를 기반으로 제작된 브라우저입니다.  
[서비스 약관](#)

chrome://settings/help



# 웹 드라이버 설치



## Current Releases

- If you are using Chrome version 115 or newer, please consult [the Chrome for Testing availability dashboard](#). This page provides convenient [JSON endpoints](#) for specific ChromeDriver version downloading.
- For older version of Chrome, please see below for the version of ChromeDriver that supports it.

For more information on selecting the right version of ChromeDriver, please see the [Version Selection](#) page.

### ChromeDriver 114.0.5735.90

Supports Chrome version 114

For more details, please see the [release notes](#).

### ChromeDriver 114.0.5735.16

Supports Chrome version 114

For more details, please see the [release notes](#).

### ChromeDriver 113.0.5672.63

Supports Chrome version 113

- Resolved issue 4205: Same object ids in Classic and BiDi [Pri-1]
- Resolved issue 4302: Don't assume that Mapper is in the first tab in ExecuteGetWindowHandles [Pri-1]
- Resolved issue 4356: Chrome 110 not utilizing pref value "download.default\_directory" [Pri-1]







For more details, please see the [release notes](#).



<https://sites.google.com/chromium.org/driver/downloads?authuser=0>

# 웹 드라이버 설치

## Index of /114.0.5735.90/

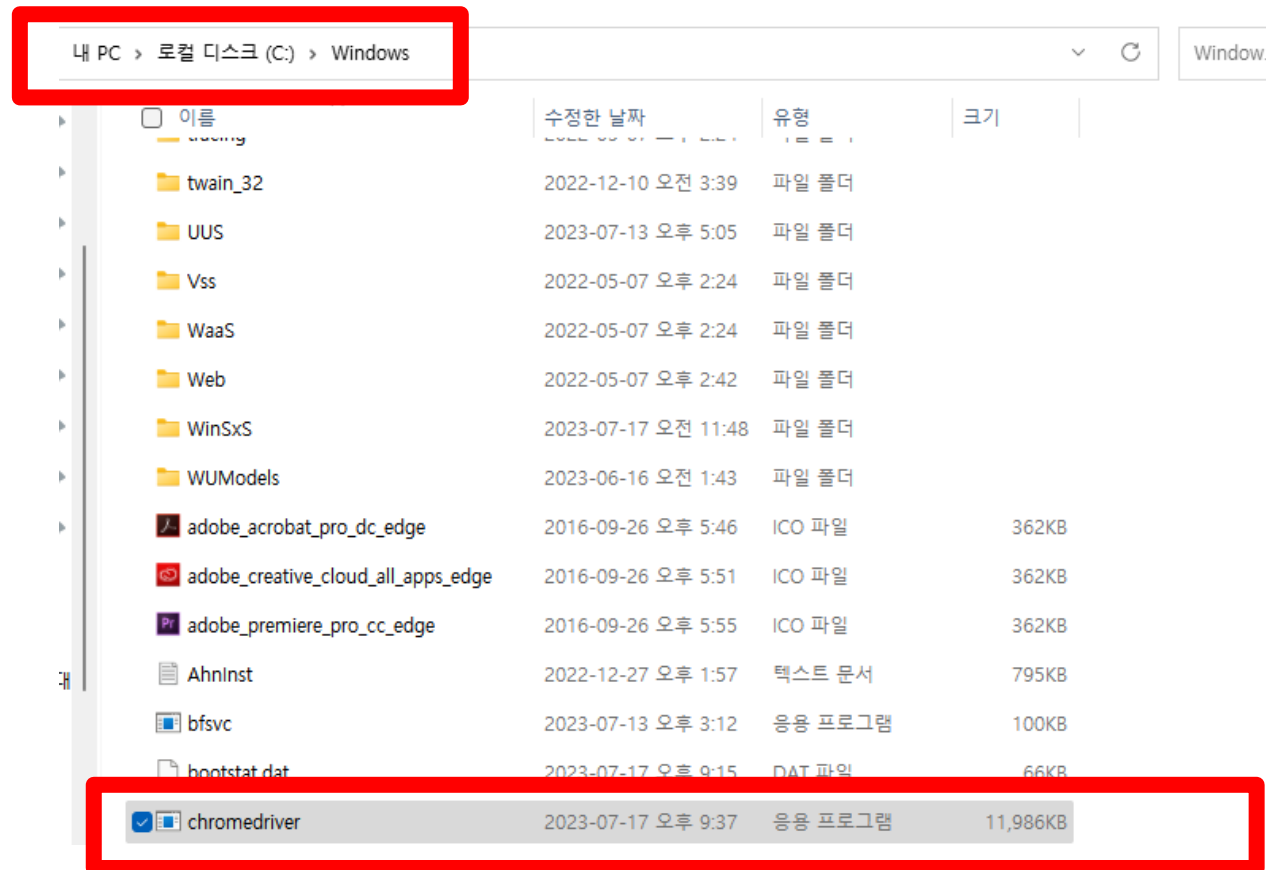
	<u>Name</u>	Last modified	Size	ETag
	<a href="#">Parent Directory</a>	-	-	-
	<a href="#">chromedriver linux64.zip</a>	2023-05-31 08:57:22	7.06MB	cd6613edf6628041684393706b62d3a6
	<a href="#">chromedriver mac64.zip</a>	2023-05-31 08:57:25	8.29MB	b44390afbdddadf8748a1d151483b2472
	<a href="#">chromedriver mac arm64.zip</a>	2023-05-31 08:57:29	7.40MB	0d515e46bea141705e49edaba1d49819
	<a href="#">chromedriver win32.zip</a>	2023-05-31 08:57:32	6.30MB	7d455bed57ef682d41108e13d45545ca
	<a href="#">notes.txt</a>	2023-05-31 08:57:38	0.00MB	1670f6dde7877ca84ecd4c56b9cc759c



# 웹 드라이버 설치: Windows의 경우

1. 압축파일 압축 해제

2. chromedriver.exe 파일을 로컬 디스크 (C:) 내의 Windows 폴더에 넣기





# 웹 드라이버 설치: MacOS의 경우

1. 압축파일 압축 해제
2. chromedriver.exe 파일을 크롤링 파일이 있는 폴더에 복사해서 옮기기
3. 다음과 같이 코드 입력

```
from webdriver_manager.chrome import ChromeDriverManager  
  
driver = webdriver.Chrome(ChromeDriverManager().install())
```



# 주의사항

GoogleColab에서는 selenium이 지원되지 않음  
Jupyter Notebook 환경 권장



# Input 태그



데이터를 입력받을 수 있는 태그

Ex) 검색창, 아이디 입력 창 등

1. input 요소를 찾아서 변수에 할당
2. 변수명.send\_keys(입력)을 통해 입력
3. 변수명.send\_keys(keys.RETURN)을 통해 입력한 것 엔터로 실행



# Selenium으로 크롤링하기

`driver.find_element(By문, 키워드)`

⇒ By에 해당하는 메소드로 키워드에 해당하는 것 반환 (한 개만)

`driver.find_elements(By문, 키워드)`

=> By에 해당하는 메소드로 키워드에 해당하는 것 반환 (모두, 여러 개)

Selenium 객체는 객체 내에서의 중복 크롤링이 가능



# Selenium By문 정리

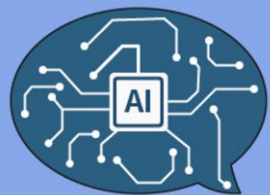
Find\_element(By.ID, 키워드)

⇒ ID 속성이 키워드에 해당하는 객체를 반환

Find\_element(By.CSS\_SELECTOR, 키워드)

⇒ 키워드에 해당하는 객체를 반환 (BeautifulSoup의 select 구문과 같음)





감사합니다

