

크롤링 차근차근 시작하기 (2/E) - 중급편

Selenium: ChromeOptions 필수옵션

여러분의 파이썬/장고 페이스메이커가 되겠습니다.

webdriver.Chrome 생성자 주요 인자

executable_path

Chrome Driver 위치 지정

상대경로/절대경로 모두 가능

디폴트로 환경변수 PATH상에서 chromedriver 찾을

```
webdriver.Chrome(  
    executable_path='chromedriver',  
    options=None,  
    생략 ...)
```

options

chrome_options(deprecated)에서 인자 이름 변경

webdriver.ChromeOptions() 객체 전달

Headless 모드

"창이 없는" 브라우저 구동

GUI가 없는 환경에서도 구동 가능

예전에는 PhantomJS 브라우저에서만 지원했지만, 2017년부터 Chrome/Firefox에서 공식 지원

Headless 모드에서는 User-Agent헤더에 Headless임을 알림

```
Mozilla/5.0 (Macintosh; Intel Mac OS X 10_14_3) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/72.0.3626.119 Safari/537.36  
Mozilla/5.0 (Macintosh; Intel Mac OS X 10_14_3) AppleWebKit/537.36 (KHTML, like Gecko) HeadlessChrome/72.0.3626.119 Safari/537.3
```

```
import os
from selenium import webdriver
```

```
options = webdriver.ChromeOptions()
```

```
options.add_argument('--lang=%s' % 'ko-KR') # 언어 설정 (브라우저에서는 대개 OS 언어 설정을 그대로 가져옵니다.)
```

```
options.add_argument('--disable-dev-shm-usage')
```

Large Pages를 렌더링할 때 Chrome이 Crash될 경우 => 메모리 부족이 원인일수도.
shared memory를 끄고 temporary path경로를 사용토록 한다. => Chrome 65부터

```
# 커스텀 User Data 디렉토리 지정 여부 (Profile/extensions, Cache, Cookies, Local Storage, Blob Storage 등)
```

```
user_data_path = os.path.join(os.environ['LOCALAPPDATA'], r'Google\Chrome\User Data') # Windows
```

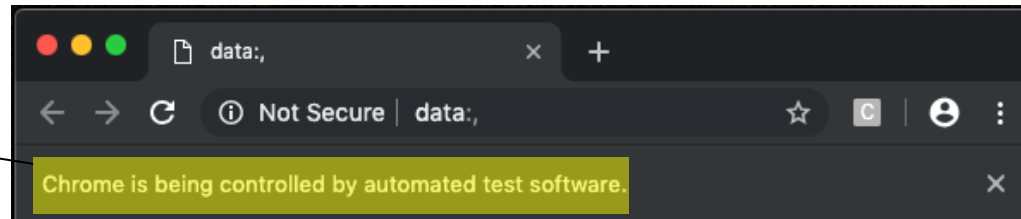
Mac OS X → os.path.join(os.environ['HOME'], 'Library/Application Support/Google/Chrome')

```
options.add_argument('--user-data-dir=%s' % user_data_path)
```

ChromeDriver에서는 디폴트로 매 세션마다 임시 Profile을 지정
--user-data-dir 옵션을 통해, 사용자 Profile 디렉토리 지정
확장도 지정 경로에 설치

```
options.add_argument('--headless')
```

```
options.add_argument('--disable-infobars')
```



```
fake_user_agent = ('Mozilla/5.0 (Macintosh; Intel Mac OS X 10_14_3) AppleWebKit/537.36 '
                  '(KHTML, like Gecko) Chrome/72.0.3626.119 Safari/537.36')
```

```
options.add_argument('--user-agent=%s' % fake_user_agent)
```

Headless모드일 경우, User-Agent에서 Headless를 제거코자 할 경우

창 크기 관련 옵션

```
options.add_argument("--window-size=1920,1080")
```

크롤링 대상 페이지의 전체 페이지가 다 그려지도록, 창 크기를 지정할 필요가 있습니다.

```
options.add_argument('--start-maximized')
```

```
options.add_argument('--start-fullscreen')
```

에러) unknown error: DevToolsActivePort file doesn't exist

--user-data-dir를 지정하면, 다음 에러가 발생할 수도 있습니다.

headless 모드 지정을 통해 회피가 되기도 합니다. → bug같아요. ☹️

~~크롬 기본 User Data 경로가 아닌 다른 경로를 지정하면, 아래와 같이 DevToolsActivePort 파일이 생성이 됩니다. (내용은 매번 새로이 지정됨)~~

```
~/anaconda3/lib/python3.7/site-packages/selenium/webdriver/remote/errorhandler.py in check_response(self, response)
    240         alert_text = value['alert'].get('text')
    241         raise exception_class(message, screen, stacktrace, alert_text)
--> 242         raise exception_class(message, screen, stacktrace)
    243
    244     def _value_or_default(self, obj, key, default):

WebDriverException: Message: unknown error: Chrome failed to start: exited abnormally
(unknown error: DevToolsActivePort file doesn't exist)
(The process started from chrome location /Applications/Google Chrome.app/Contents/MacOS/Google Chrome is no longer
running, so ChromeDriver is assuming that Chrome has crashed.)
(Driver info: chromedriver=73.0.3683.68 (47787ec04b6e38e22703e856e101e840b65afe72),platform=Mac OS X 10.14.3 x86_6
4)
```

```
$ cat DevToolsActivePort
61663
/devtools/browser/7b16ad8c-1e6b-431b-9cae-be70bf2da3ba
```

브라우저 확장을 사용할려면?

Chrome Extension Downloader → <https://chrome-extension-downloader.com/>

AdGuard AdBlocker : <https://chrome.google.com/webstore/detail/adguard-adblocker/bgnkhhnnamicmpeenaelnjfhikgbkllg?hl=ko>

광고, 팝업 등의 사이트 광고를 차단 → 페이지 로딩이 빨라지고, 대역폭을 절약

Block image : <https://chrome.google.com/webstore/detail/block-image/pehaalcefcjfccdpbckoablngfkfgfgj?hl=ko>

이미지와 동영상 로딩을 차단

Stealth Mode : <https://chrome.google.com/webstore/detail/stealth-mode/hfkipddhkjcbfmnclailpehabmekgphfh?hl=ko>

서드파티 쿠키를 차단하고 온라인 추적자가 사용자 정보를 수집할 수 없도록 합니다.

```
from selenium import webdriver
```

```
options = webdriver.ChromeOptions()
```

지정 User Data 디렉토리의 Default 프로파일의 extensions에 설치

```
options.add_extension("./chrome-extensions/AdGuard-AdBlocker_v3.0.11.crx")
```

```
options.add_extension("./chrome-extensions/Block-image_v1.1.crx")
```

```
options.add_extension("./chrome-extensions/Stealth-Mode_v1.0.9.crx")
```

```
with webdriver.Chrome("./drivers/chromedriver", options=options) as driver:
```

```
    driver.implicitly_wait(10)
```

```
    # ...
```

인생은 짧습니다.
파이썬/장고를 쓰세요.

여러분의 파이썬/장고 페이스메이커가 되겠습니다.

- Ask Company