

크롤링 차근차근 시작하기 (2/E) - 중급편

Internet Explorer Driver 설정하기

여러분의 파이썬/장고 페이스메이커가 되겠습니다.

Selenium 기본 설치 (1) (지난 슬라이드 복&붙)

python3

공식 배포판 : <https://www.python.org>

Anaconda 배포판 : <https://www.anaconda.com/distribution/> (추천)

주의) 설치 시에 환경변수 PATH에 꼭 체크해주세요.

라이브러리

```
pip3 install selenium beautifulsoup4 requests jupyter
```

Selenium 기본 설치 (2) (지난 슬라이드 복&붙)

사용할 Driver 다운로드 (원하는 경로의 drivers 폴더에 압축풀어서 모아두기)

Internet Explorer Driver Server : 32비트용 권장 (64비트 머신이라도 32비트 권장)

윈도우만 지원

Chrome Driver

Chrome 설치 필요 : <https://www.google.com/chrome/>

Firefox Gecko Driver

Firefox 설치 필요 : <https://www.mozilla.org/firefox/>

Edge

윈도우만 지원

정기적으로 Driver 버전을 체크하여, 최신버전으로 업데이트해주세요.
(요즘 브라우저는 자동 업데이트)

Internet Explorer Driver

윈7/8/8.1과 IE7/8/9/10/11에서 테스트됨.

윈10에서도 잘 동작합니다.

IE6은 2014년 4월 이후로 지원 중단

맥/리눅스에서 사용하려면, VirtualBox/VMWare/Parallels를 통해 Windows 구동 가능 → 별도 에피소드에서 소개

드라이버 다운로드

IEDriverServer.exe

32비트 드라이버 : 32비트 IE 구동 (추천)

64비트 드라이버 : 64비트 IE 구동

<https://selenium-release.storage.googleapis.com/index.html>

다운받아서, 환경변수 PATH 상의 경로에 둡니다.

NoSuchWindowException 예외 발생

Message: Unable to get browser

대응 1) 32비트 Driver 사용

대응 2) 브라우저 옵션

Security Tab

모든 Region에 대해, "Enable Protected Mode" 옵션 해제 → 다음 페이지에서 소개

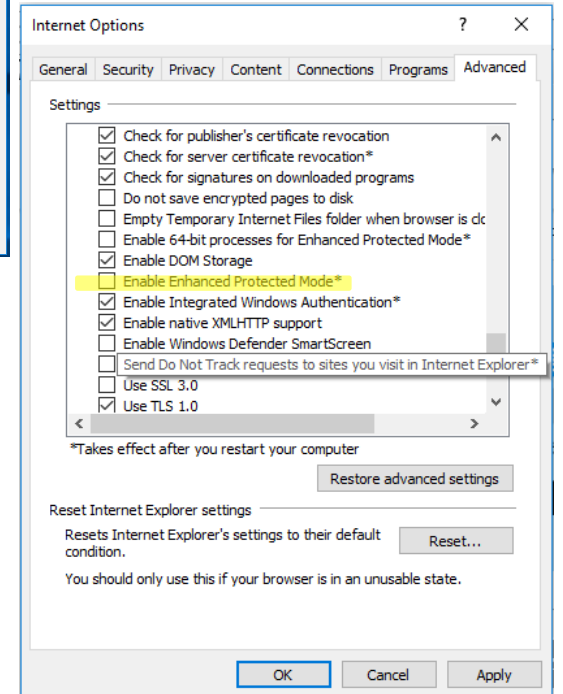
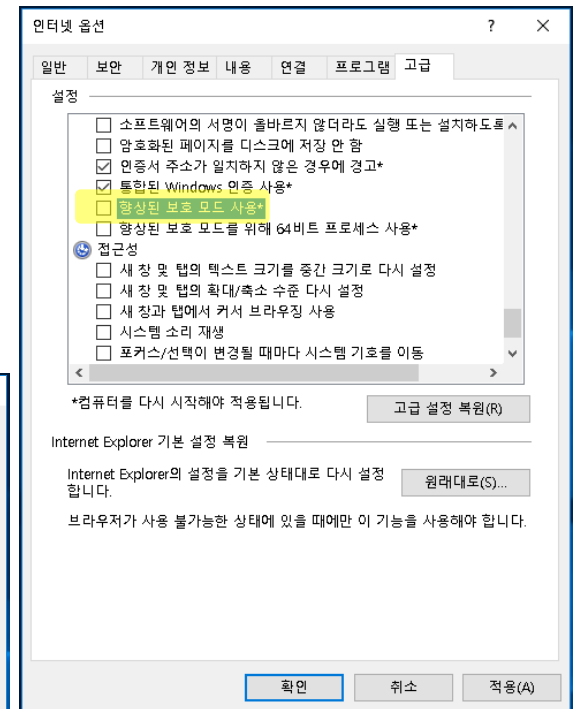
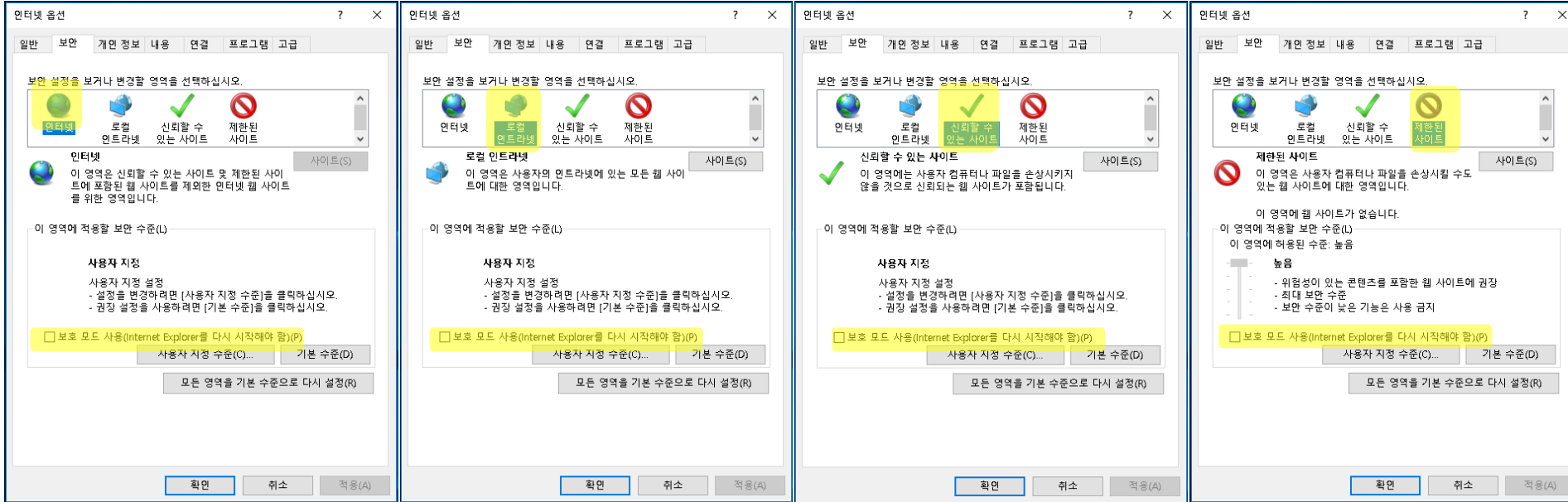
```
In [2]: driver = webdriver.Ie("./drivers/IEDriverServer-3.141.59.exe")
```

```
In [*]: driver.get("http://naver.com")
```

"보호 모드 사용"을 모드를 모두 끄기

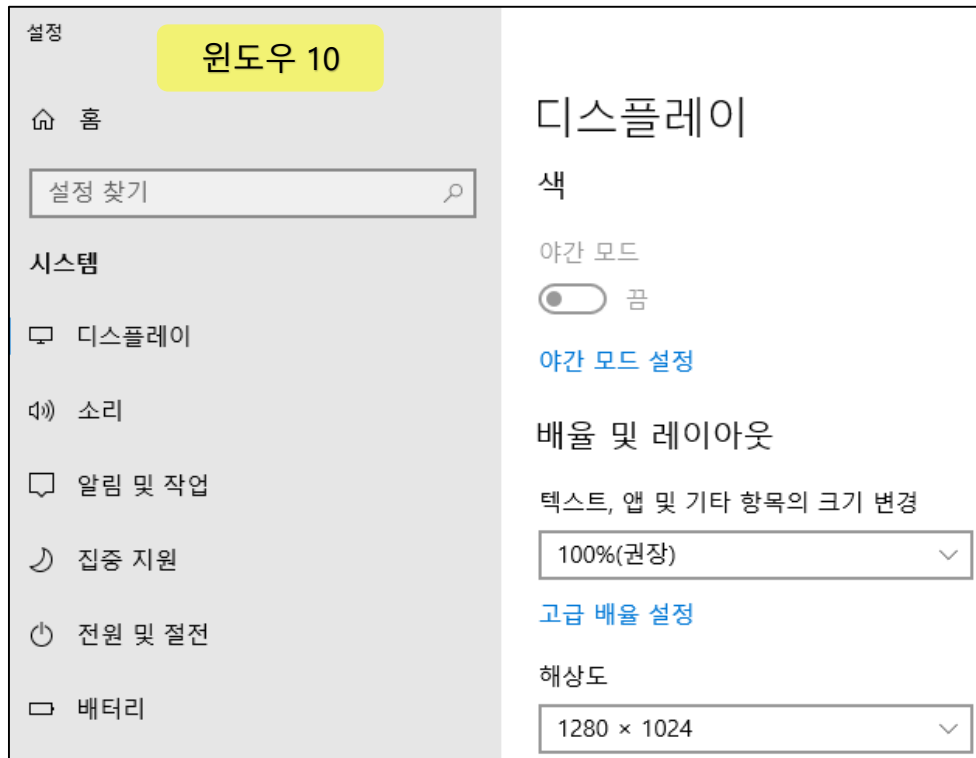
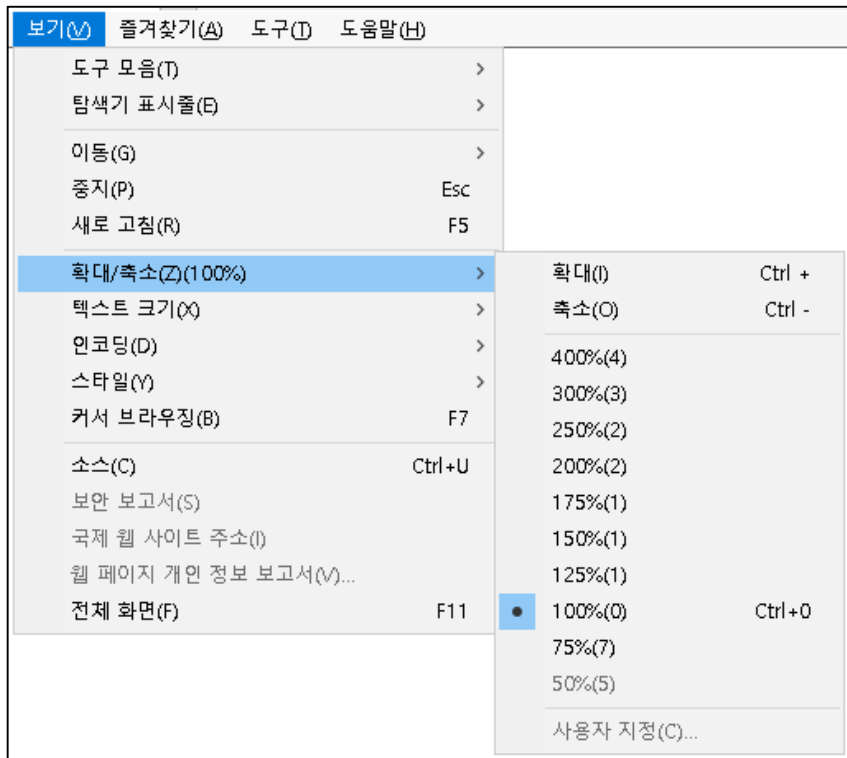
Internet Explorer 브라우저 : 도구 → 인터넷 옵션

윈도우 Vista/7의 IE7 이상에서 반드시.



Zoom Level은 필히 100%

마우스 이벤트가 정확한 위치에서 발생토록 하기 위함.



IE11에서 추가 레지스트리 설정

생성된 IE 인스턴스로의 Connection을 유지하기 위해

FEATURE_BFCACHE 위치

32비트 윈도우

HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Internet Explorer\Main\FeatureControl\FEATURE_BFCACHE

64비트 윈도우

HKEY_LOCAL_MACHINE\SOFTWARE\Wow6432Node\Microsoft\Internet Explorer\Main\FeatureControl\FEATURE_BFCACHE

FEATURE_BFCACHE 내에 생성할 값

타입 : DWORD

이름 : iexplore.exe

값 : 0

Demo Code

```
from selenium import webdriver

with webdriver.Ie("./drivers/IEDriverServer-3.141.59.exe") as driver:
    driver.implicitly_wait(10)
    driver.get("http://naver.com")
    element_list = driver.find_elements_by_css_selector(".PM_CL_realtimeKeyword_rolling_base .ah_k")
    keyword_list = [ele.get_attribute("innerHTML") for ele in element_list]

print(keyword_list)
```

인생은 짧습니다.
파이썬/장고를 쓰세요.

여러분의 파이썬/장고 페이스메이커가 되겠습니다.

- Ask Company