

크롤링 차근차근 시작하기 (2/E) - 중급편

Selenium Remote Driver

여러분의 파이썬/장고 페이스메이커가 되겠습니다.

Selenium Remote WebDriver

네트워크로 연결된, 다른 원격 서버(들)에서 실행시킬 수 있는 WebDriver

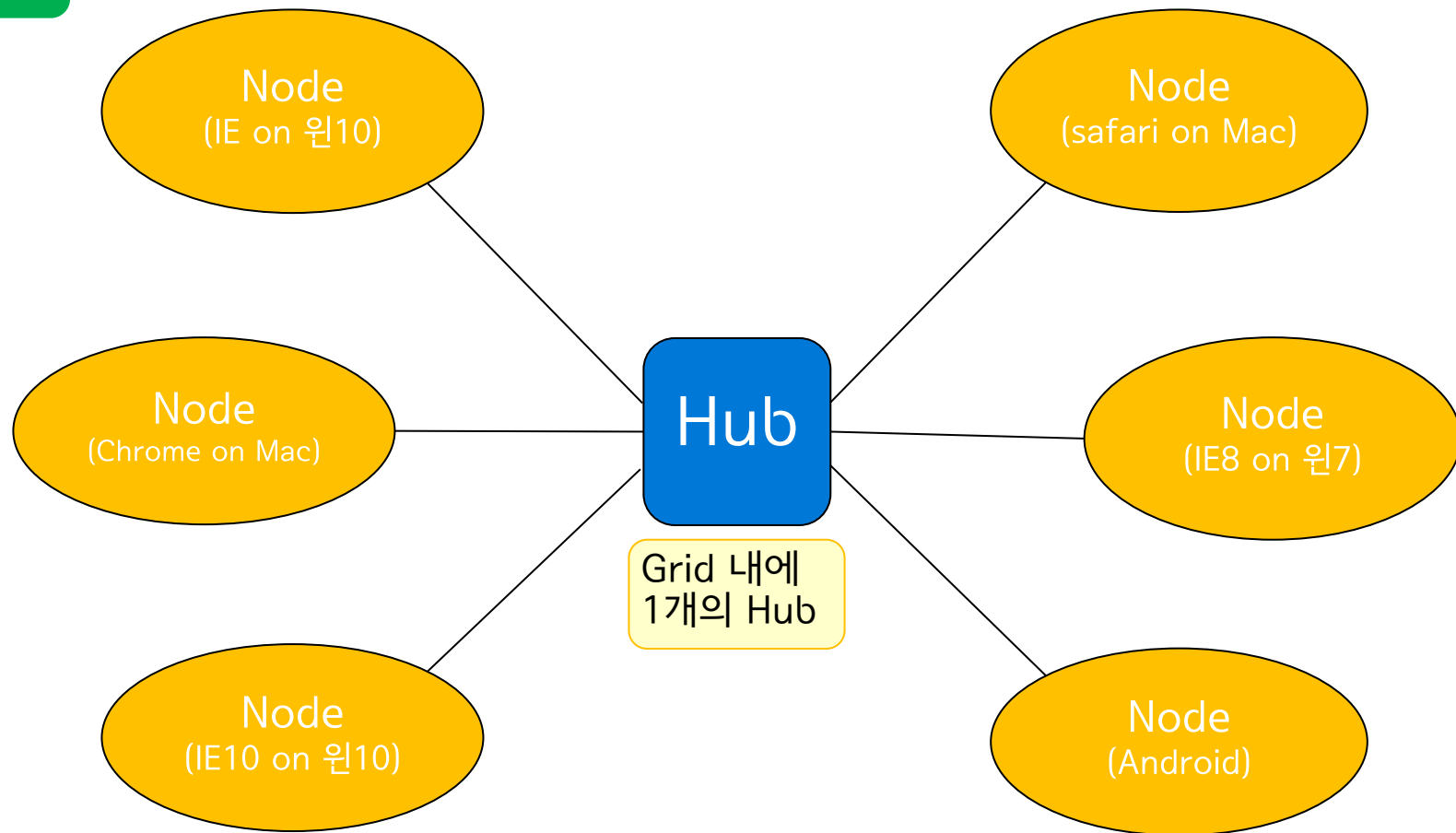
- 서로 다른 운영체제/브라우저를 동시에 운영 가능

Hub/Node에서 selenium-server-standalone-버전.jar 를 다른 옵션으로 구동

- 각 세션의 로그는 메모리에 기록
- Selenium#stop() 혹은 Selenium#quit() 호출 시에 정리 → 세션 종료를 잊을 경우, 메모리 누수로 이어집니다.

<https://seleniumhq.github.io/docs/remote.html>

Grid



Running IEDriverServer.exe Remotely

IEDriverServer.exe는 로컬 머신 내에서의 접속 만을 허용.

IEDriverServer.exe의 소스코드를 변경하지 않고서는 불가.

- 변경된 버전을 Hub/Node가 동일하게 사용해야합니다.

➔ Selenium Remote WebDriver를 이용하면, OK.

<https://github.com/SeleniumHQ/selenium/wiki/InternetExplorerDriver#running-iedriverserverexe-remotely>

허브/노드에서의 설치

해당 머신 측에

JRE 설치가 필요

<https://www.oracle.com/technetwork/java/javase/downloads/jre8-downloads-2133155.html>

Selenium Standalone Server 필요

<https://www.seleniumhq.org/download/>

Hub에서 서버 구동

허브

```
java -jar selenium-server-standalone-버전.jar \  
-role hub \  
-port 4444
```

hub, node, standalone 중 택일

Endpoints

/grid/console : Grid 웹 콘솔

/grid/register : Node에서 허브에 등록하는 Endpoint

/wd/hub : WebDriver Executor, Endpoint

VirtualBox에서 지원하는 네트워크

NAT (디폴트) : 각 VM마다 별도 가설 네트워크가 생성. 인터넷 가능.

NAT network : NAT + 여러 VM들이 한 네트워크

Bridge Adapter : VM이 호스트와 같은 물리적인 네트워크를 사용.

Internal Network : NAT - 인터넷 연결 불가.

Host-only Network : Internal Network + 호스트와의 연결만 허용.

<https://www.virtualbox.org/manual/ch06.html>

Nodes에서 서버 구동

Tip) 윈도우 명령프롬프트에서 여러 줄 명령 끝에는 역슬래시가 아닌 ^를 쓰셔야 합니다.

노트

필히 -Dwebdriver.chrome.driver 옵션을 -jar 옵션보다 앞에 지정

```
java -Dwebdriver.chrome.driver="chromedriver.exe" \  
-Dwebdriver.ie.driver="IEDriverServer.exe" \  
-Dwebdriver.gecko.driver="geckodriver.exe" \  
-jar selenium-server-standalone-버전.jar \  
-role node \  
-hub http://허브_아이피:허브_포트번호/grid/register \  
-port 5555
```

Exception in thread "main" com.beust.jcommander.ParameterException:
Was passed main parameter '-
Dwebdriver.chrome.driver=drivers\chromedriver-win32-73.0.3683.68.exe'
but no main parameter was defined in your arg class

System Property 설정

Hub 등록에 실패했을 때

```
21:42:48.967 INFO [GridLauncherV3.parse] - Selenium server version: 3.141.59, revision: e82be7d358  
21:42:49.093 INFO [GridLauncherV3.lambda$buildLaunchers$7] - Launching a Selenium Grid node on port 40033  
2019-04-22 21:42:49.197:INFO::main: Logging initialized @443ms to org.seleniumhq.jetty9.util.log.StdErrLog  
21:42:49.454 INFO [WebDriverServlet.<init>] - Initialising WebDriverServlet  
21:42:49.538 ERROR [BaseServer.start] - Port 40033 is busy, please choose a free port and specify it using -port option  
Exception in thread "main" java.lang.RuntimeException: java.net.BindException: Can't assign requested address  
    at org.openqa.selenium.grid.server.BaseServer.start(BaseServer.java:221)  
    at org.openqa.selenium.remote.server.SeleniumServer.boot(SeleniumServer.java:131)  
    at org.openqa.grid.internal.utils.SelfRegisteringRemote.startRemoteServer(SelfRegisteringRemote.java:119)  
    at org.openqa.grid.selenium.GridLauncherV3.lambda$buildLaunchers$7(GridLauncherV3.java:284)  
    at org.openqa.grid.selenium.GridLauncherV3.lambda$launch$0(GridLauncherV3.java:86)  
    at java.util.Optional.map(Optional.java:215)  
    at org.openqa.grid.selenium.GridLauncherV3.launch(GridLauncherV3.java:86)  
    at org.openqa.grid.selenium.GridLauncherV3.main(GridLauncherV3.java:70)  
Caused by: java.net.BindException: Can't assign requested address  
    at sun.nio.ch.Net.bind0(Native Method)  
    at sun.nio.ch.Net.bind(Net.java:433)  
    at sun.nio.ch.Net.bind(Net.java:425)  
    at sun.nio.ch.ServerSocketChannelImpl.bind(ServerSocketChannelImpl.java:223)
```


Demo

```
from selenium import webdriver
from selenium.webdriver.common.desired_capabilities import DesiredCapabilities

capabilities = DesiredCapabilities.CHROME.copy()
capabilities['browserName'] = 'internet explorer'
# capabilities → {'browserName': 'internet explorer', 'platform': 'ANY', 'version': ''}

kwargs = dict(
    # command_executor="http://localhost:4444/wd/hub",
    desired_capabilities=capabilities
)

with webdriver.Remote(**kwargs) as driver:
    driver.get("http://naver.com")
    print(driver.title)
```

주의) DesiredCapabilities.* 사전을 변경하지 마세요.

DesiredCapabilities.* 내 값들은 상수가 아니라, 단순히 dict 값입니다.
그렇기에 아래와 같이 변경하지마시고, 필요하다면 .copy() 해서 쓰세요.

DesiredCapabilities

브라우저 선택

- browserName : android, chrome, firefox, htmlunit, internet explorer, iPhone, ipad, opera, safari 중 택일
- version : 브라우저 버전을 지정. 빈 문자열로 지정하면 자동 탐지.
- platform : 운영체제를 지정. WINDOWS, XP, VISTA, MAC, LINUX, UNIT, ANDROID, ANY 중 택일

<https://github.com/SeleniumHQ/selenium/wiki/DesiredCapabilities>

참고) docker-selenium

Docker 컨테이너를 활용한 Hub/Nodes

```
$ docker network create grid
```

Hub

```
$ docker run -d -p 4444:4444 --net grid \  
  --name selenium-hub \  
  selenium/hub:3.141.59-mercury
```

Nodes

```
$ docker run -d --net grid -e HUB_HOST=selenium-hub \  
  -v /dev/shm:/dev/shm \  
  selenium/node-chrome:3.141.59-mercury
```

```
$ docker run -d --net grid -e HUB_HOST=selenium-hub \  
  -v /dev/shm:/dev/shm \  
  selenium/node-firefox:3.141.59-mercury
```

<https://github.com/SeleniumHQ/docker-selenium#using-docker-networking>

인생은 짧습니다.
파이썬/장고를 쓰세요.

여러분의 파이썬/장고 페이스메이커가 되겠습니다.

- Ask Company