Python 调用浏览器

好记性不如烂笔头，今天把selenium调用浏览器操作总结一下。

    首先，得先明确，为什么要采用selenium来模拟登陆网页。最主要的原因我认为还是在于通过这种模拟登录方式获取的页面html代码，可以把js里的内容也获取到，而通过urllib方式模拟登录的方式虽然也可以伪装成浏览器的形式获取页面html代码，但是这里面的js，css代码是没有的，也就是没有动态的内容，达不到全面抓取数据的目的；当然除了selenium这种方式外，还有其他的途径也能获取到js等动态代码，这里先不做介绍了，以后再分别细说。

    说了selenium的优点，就需要再强调一下它的不足，selenium调用浏览器时，跟我们手动打开浏览器效果一样，你可以想想自己打开浏览器时的镜像，从开启到加载完毕，要耗费好几秒时间，对于一般的脚本来说，好几秒可是很长的时间了，但是它就是这么傲娇，加载不完就别想玩后面的；如果你只是想偶尔打开一次抓一次，几秒倒也没什么，但如果你要循环执行的话，这个方法的效率就是低的不要不要的了。

    总之，还是要开发人员自己权衡利弊，决定是否使用。

    好了，接下来我们就来介绍python环境下，如何利用selenium体统的webdriver来浏览三大常用浏览器：IE，Chrome，Firefox。

    1 利用Chrome浏览器，安装的chrome浏览器是没有没有提供现成的driver的，因此需要我们自己下载ChromeDriver.exe，这里可以有[下载地址](http://www.xdowns.com/soft/38/63/2015/Soft_136971.html" \t "_blank)，下载完后把它放到chrome浏览器所在的安装目录下就可以了，保证让他跟chrome.exe呆一起就可以了。相关脚本如下

from selenium import webdriver

from selenium.webdriver.common.keys import Keys

import os

chromedriver = "D:\Program Files\ChromeGAE\Chrome\chromedriver.exe"

os.environ["webdriver.chrome.driver"] = chromedriver

browser = webdriver.Chrome(chromedriver)

url = "http://www.baidu.com"

browser.get(url)

这样基本就可以直接用了。

    2 利用IE浏览器，IE浏览器和chrome比较相似，也需要我们自己去找到Driver，名字是IEDriverServer.exe，这个随便一百度也能找到，就不放连接了，下载后跟chrome那里一样，也是放到ieexplore.exe的同级目录下，这个一般是在C:\Program Files\Internet Explorer下，毕竟是系统自带的嘛。脚本如下

from selenium import webdriver

from selenium.webdriver.common.keys import Keys

import oschromedriver = "C:\Program Files\Internet Explorer\IEDriverServer.exe"

os.environ["webdriver.ie.driver"] = IEdriver

browser = webdriver.Ie(IEdriver)

url = "http://www.baidu.com"

browser.get(url)

3 利用Firefox浏览器，firefox这个相比起前面两个就方便多了，果然是亲开发的良心浏览器，自带driver光环，拿来就可以用。

from selenium import webdriver

from selenium.webdriver.common.keys import Keys

browser = webdriver.Firefox()

url = "http://www.baidu.com"

browser.get(url)

是不是方便多了，但是！但是！但是！这里不得不说的是，firefox也是有大坑的，firefox会经常更新，一单更新了，你的selenium没跟到位，那就费了，银次你需要关闭firefox的自动更新，并且将firefox保持在一个较低的版本（45以前的应该还是可以的），我当初遇到的问题就是firefox更新后，脚本可以调用浏览器打开，但是却打不开url，各种找原因，看到其他网友就说是版本太高的原因，方法就是再找到老版本的firefox安装。

    好了，就说这些了，如有不对之处还望多批评指正。