

无头浏览器

我们已经基本了解了selenium的基本使用了. 但是呢, 不知各位有没有发现, 每次打开浏览器的时间都比较长. 这就比较耗时了. 我们写的是爬虫程序. 目的是数据. 并不是想看网页. 那能不能让浏览器在后台跑呢? 答案是可以的.

咱直接上案例吧. 拿出最开始我们看到的那个网页. 抓取电影票房. 并且用正常的有浏览器窗口的方式来抓取. 然后再改成后台运行不就好了么

```
1 from selenium.webdriver import Chrome
2 from selenium.webdriver.support.select import
   Select
3 import time
4
5
6 web = Chrome()
7 web.get('https://www.endata.com.cn/BoxOffice/B0/Y
   ear/index.html')
8
9
```

```
10 # 切换select
11 sel = Select(web.find_element_by_xpath('//*
    [@id="OptionDate"]'))
12 for i in range(len(sel.options)):
13     sel.select_by_index(i) # 按照索引位置切换
14     time.sleep(1)
15     table = web.find_element_by_xpath('//*
    [@id="TableList"]/table')
16
    print("=====
    ==")
17     print(table.text)
```

接下来, 我们对程序进行修改. 让selenium在后台安安静静的执行.

```
1 from selenium.webdriver import Chrome
2 from selenium.webdriver.chrome.options import
    Options
3 from selenium.webdriver.support.select import
    Select
4 import time
5
6 # 准备无头浏览器配置信息
7 opt = Options()
```

```
8 opt.add_argument('--headless')
9 opt.add_argument('--disable-gpu')
10
11 web = Chrome(options=opt) # 将无头信息进行配置
12 web.get('https://www.endata.com.cn/BoxOffice/B0/Y
    ear/index.html')
13
14
15 # 切换select
16 sel = Select(web.find_element_by_xpath('//*
    [@id="OptionDate"]'))
17 for i in range(len(sel.options)):
18     sel.select_by_index(i) # 按照索引位置切换
19     time.sleep(1)
20     table = web.find_element_by_xpath('//*
    [@id="TableList"]/table')
21
22     print("=====
    ==")
23     print(table.text)
24
25 # 打印浏览器源代码
26 time.sleep(3) # 等待js加载完数据
27 print(web.page_source)
```

