@[toc]

一、selenium介绍

Selenium是一系列基于Web的自动化工具，提供一套测试函数，用于支持Web自动化测试。函数非常灵活，能够完成界面元素定位、窗口跳转、结果比较。具有如下特点：

一、多浏览器支持

```
	可以对多浏览器进行测试，如IE、Firefox、Safari、Chrome、Android手机浏览器等。
```

二、支持多种语言

```
	如Java、C#、Python、Ruby、PHP等。
```

三、支持多种操作系统

```
	如Windows、Linux、IOS、Android等。
```

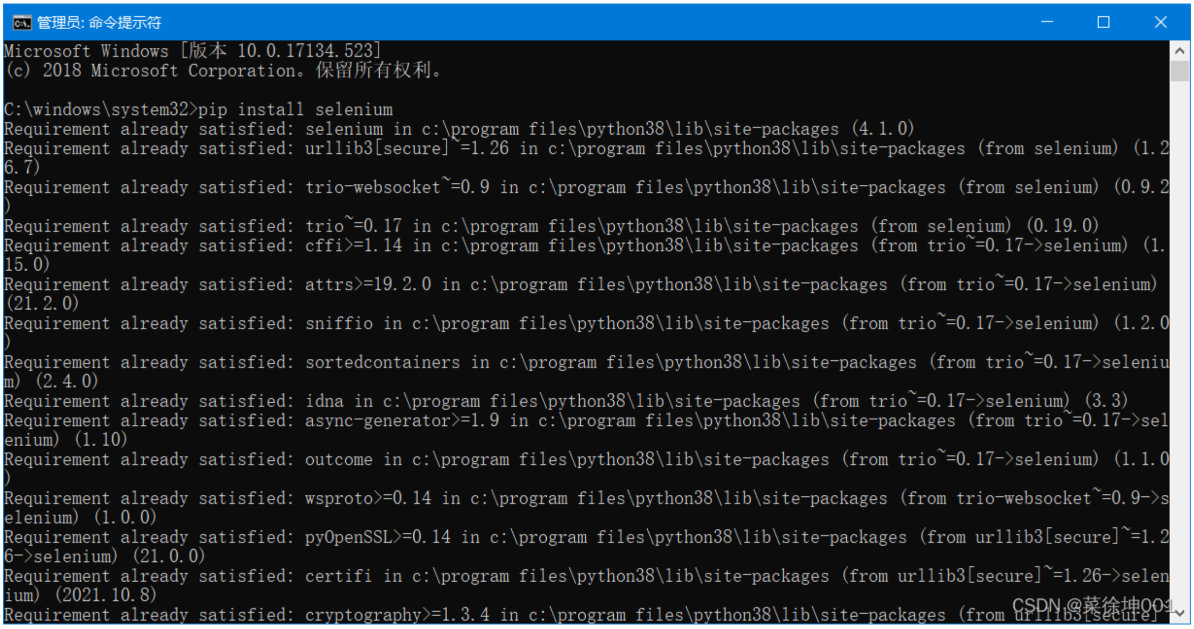
四、开源免费

```
	官网：http://www.seleniumhg.org/
```

二、准备工作

1、下载selenium

可以直接用 pip install selenium下载

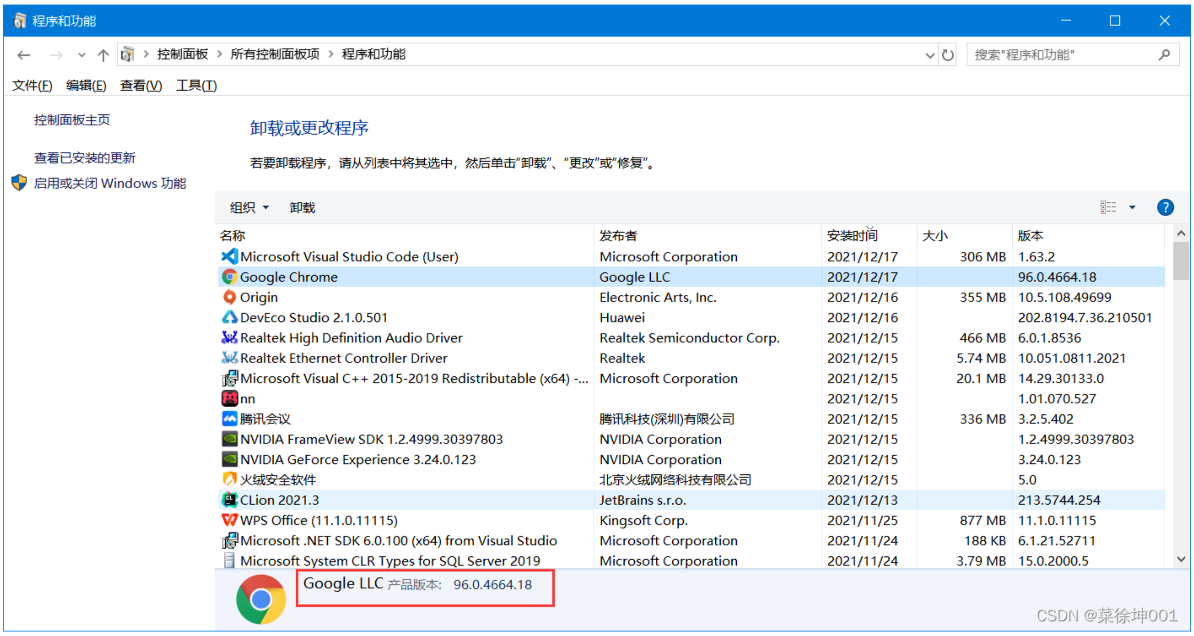


```
Microsoft Windows [版本 10.0.17134.523]
(c) 2018 Microsoft Corporation。保留所有权利。

C:\windows\system32>pip install selenium
Requirement already satisfied: selenium in c:\program files\python38\lib\site-packages (4.1.0)
Requirement already satisfied: urllib3[secure]~=1.26 in c:\program files\python38\lib\site-packages (from selenium) (1.2
6.7)
Requirement already satisfied: trio-websocket~=0.9 in c:\program files\python38\lib\site-packages (from selenium) (0.9.2
)
Requirement already satisfied: trio~=0.17 in c:\program files\python38\lib\site-packages (from selenium) (0.19.0)
Requirement already satisfied: cffi>=1.14 in c:\program files\python38\lib\site-packages (from trio~=0.17->selenium) (1.
15.0)
Requirement already satisfied: attrs>=19.2.0 in c:\program files\python38\lib\site-packages (from trio~=0.17->selenium)
(21.2.0)
Requirement already satisfied: sniffio in c:\program files\python38\lib\site-packages (from trio~=0.17->selenium) (1.2.0
)
Requirement already satisfied: sortedcontainers in c:\program files\python38\lib\site-packages (from trio~=0.17->seleniu
m) (2.4.0)
Requirement already satisfied: idna in c:\program files\python38\lib\site-packages (from trio~=0.17->selenium) (3.3)
Requirement already satisfied: async-generator>=1.9 in c:\program files\python38\lib\site-packages (from trio~=0.17->sel
enium) (1.10)
Requirement already satisfied: outcome in c:\program files\python38\lib\site-packages (from trio~=0.17->selenium) (1.1.0
)
Requirement already satisfied: wsproto>=0.14 in c:\program files\python38\lib\site-packages (from trio-websocket~=0.9->s
elenium) (1.0.0)
Requirement already satisfied: pyOpenSSL>=0.14 in c:\program files\python38\lib\site-packages (from urllib3[secure]~=1.2
6->selenium) (21.0.0)
Requirement already satisfied: certifi in c:\program files\python38\lib\site-packages (from urllib3[secure]~=1.26->selen
ium) (2021.10.8)
Requirement already satisfied: cryptography>=1.3.4 in c:\program files\python38\lib\site-packages (from urllib3[secure]
```

2、下载chrome浏览器驱动

查看chrome的版本



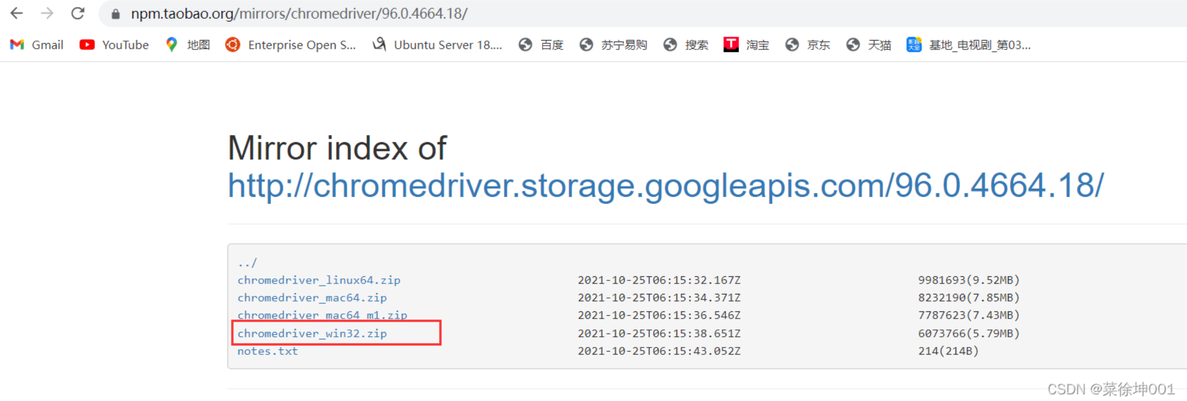
卸载或更改程序

若要卸载程序，请从列表中将其选中，然后单击"卸载"、"更改"或"修复"。

名称	发布者	安装时间	大小	版本
Microsoft Visual Studio Code (User)	Microsoft Corporation	2021/12/17	306 MB	1.63.2
Google Chrome	Google LLC	2021/12/17		96.0.4664.18
Origin	Electronic Arts, Inc.	2021/12/16	355 MB	10.5.108.49699
DevEco Studio 2.1.0.501	Huawei	2021/12/16		202.8194.7.36.210501
Realtek High Definition Audio Driver	Realtek Semiconductor Corp.	2021/12/15	466 MB	6.0.1.8536
Realtek Ethernet Controller Driver	Realtek	2021/12/15	5.74 MB	10.051.0811.2021
Microsoft Visual C++ 2015-2019 Redistributable (x64) -...	Microsoft Corporation	2021/12/15	20.1 MB	14.29.30133.0
nn		2021/12/15		1.01.070.527
腾讯会议	腾讯科技(深圳)有限公司	2021/12/15	336 MB	3.2.5.402
NVIDIA FrameView SDK 1.2.4999.30397803	NVIDIA Corporation	2021/12/15		1.2.4999.30397803
NVIDIA GeForce Experience 3.24.0.123	NVIDIA Corporation	2021/12/15		3.24.0.123
火绒安全软件	北京火绒网络科技有限公司	2021/12/15		5.0
CLion 2021.3	JetBrains s.r.o.	2021/12/13		213.5744.254
WPS Office (11.1.0.11115)	Kingsoft Corp.	2021/11/25	877 MB	11.1.0.11115
Microsoft .NET SDK 6.0.100 (x64) from Visual Studio	Microsoft Corporation	2021/11/24	188 KB	6.1.21.52711
Microsoft System CLR Types for SQL Server 2019	Microsoft Corporation	2021/11/24	3.79 MB	15.0.2000.5

Google LLC 产品版本: 96.0.4664.18

在https://npm.taobao.org/mirrors/chromedriver/下载对应的驱动



Mirror index of http://chromedriver.storage.googleapis.com/96.0.4664.18/

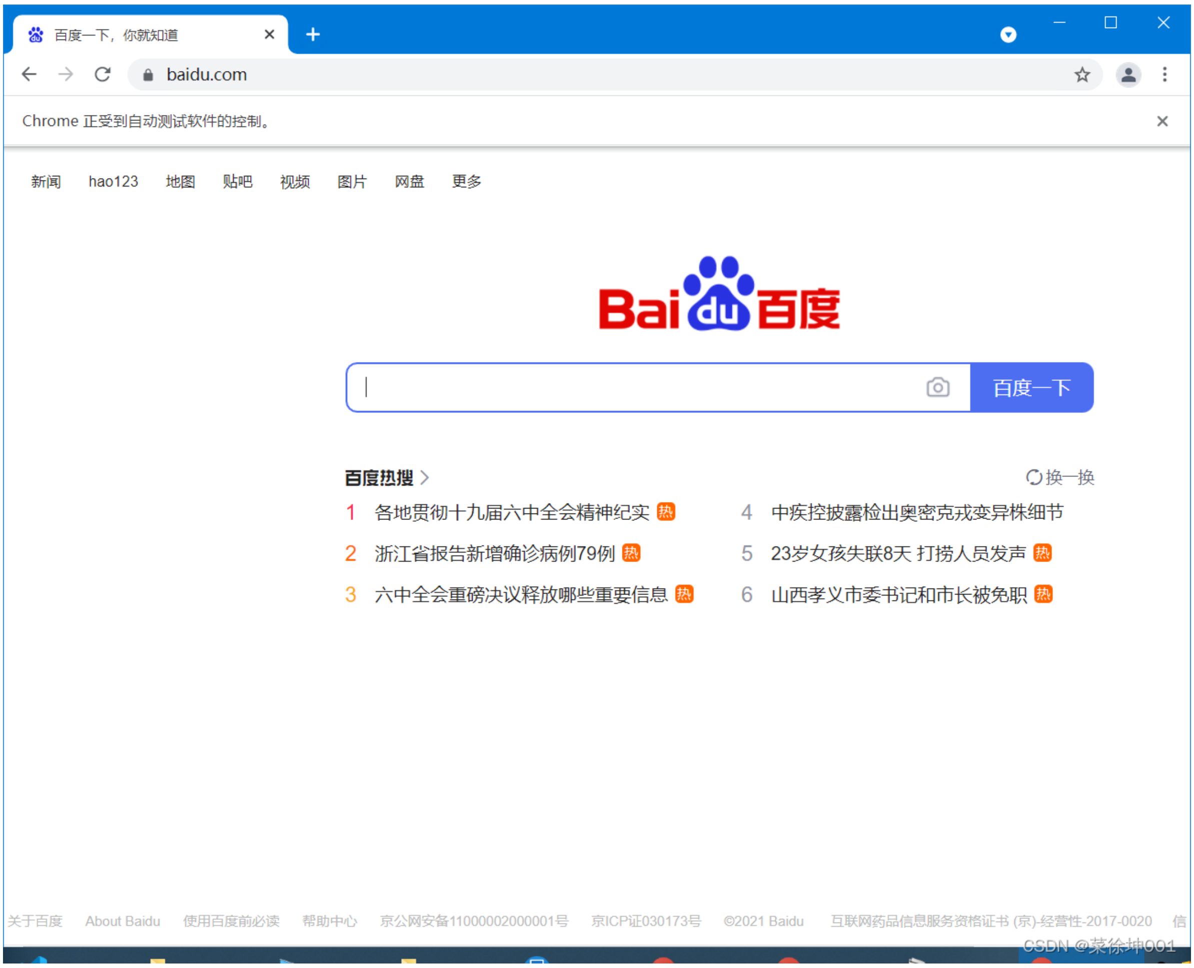
```
../
chromedriver_linux64.zip      2021-10-25T06:15:32.167Z      9981693(9.52MB)
chromedriver_mac64.zip        2021-10-25T06:15:34.371Z      8232190(7.85MB)
chromedriver_mac64_m1.zip     2021-10-25T06:15:36.546Z      7787623(7.43MB)
chromedriver_win32.zip        2021-10-25T06:15:38.651Z      6073766(5.79MB)
notes.txt                     2021-10-25T06:15:43.052Z      214(214B)
```

三、自动化测试

- 测试一下是否安装成功

```
from selenium import webdriver
driver=webdriver.Chrome('D:\\py\\chromedriver_win32\\chromedriver.exe')
#进入网页
driver.get("https://www.baidu.com/")
```

可以看到成功运行谷歌浏览器


百度一下，你就知道
baidu.com
新闻
hao123
地图
贴吧
视频
图片
网盘
更多
Bai du 百度
百度一下
百度热搜 >
换一换
关于百度 About Baidu 使用百度前必读 帮助中心 京公网安备11000002000001号 京ICP证030173号 ©2021 Baidu 互联网药品信息服务资格证书 (京)-经营性-2017-0020 信
CSDN @菜徐坤001

Chrome 正受到自动测试软件的控制。

1 各地贯彻十九届六中全会精神纪实 热
2 浙江省报告新增确诊病例79例 热
3 六中全会重磅决议释放哪些重要信息 热
4 中疾控披露检出奥密克戎变异株细节
5 23岁女孩失联8天 打捞人员发声 热
6 山西孝义市委书记和市长被免职 热

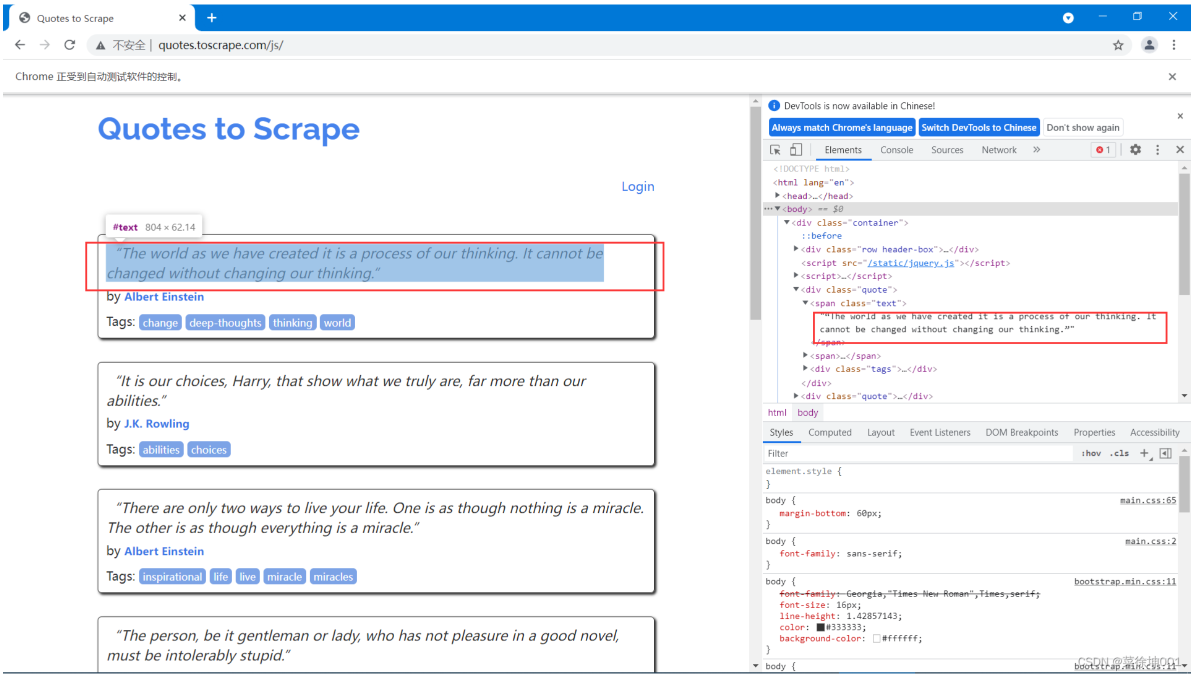
四、爬取名言

```
from bs4 import BeautifulSoup as bs
from selenium import webdriver
import csv
from selenium.webdriver.chrome.options import Options
driver=webdriver.Chrome('D:\\py\\chromedriver_win32\\chromedriver.exe')
driver.get('http://quotes.toscrape.com/js/')
```

- 定义csv表头和文件路径，以及存放内容的列表

```
#定义csv表头
quote_head=['名言','作者']
#csv文件的路径和名字
quote_path='..\\source\\mingyan_csv.csv'
#存放内容的列表
quote_content=[]
```

- 进入开发者模式，定位名言的位置，观察名言的位置



Quotes to Scrape

“The world as we have created it is a process of our thinking. It cannot be changed without changing our thinking.”
by Albert Einstein
Tags: change deep-thoughts thinking world

“It is our choices, Harry, that show what we truly are, far more than our abilities.”
by J.K. Rowling
Tags: abilities choices

“There are only two ways to live your life. One is as though nothing is a miracle. The other is as though everything is a miracle.”
by Albert Einstein
Tags: inspirational life live miracle miracles

“The person, be it gentleman or lady, who has not pleasure in a good novel, must be intolerably stupid.”

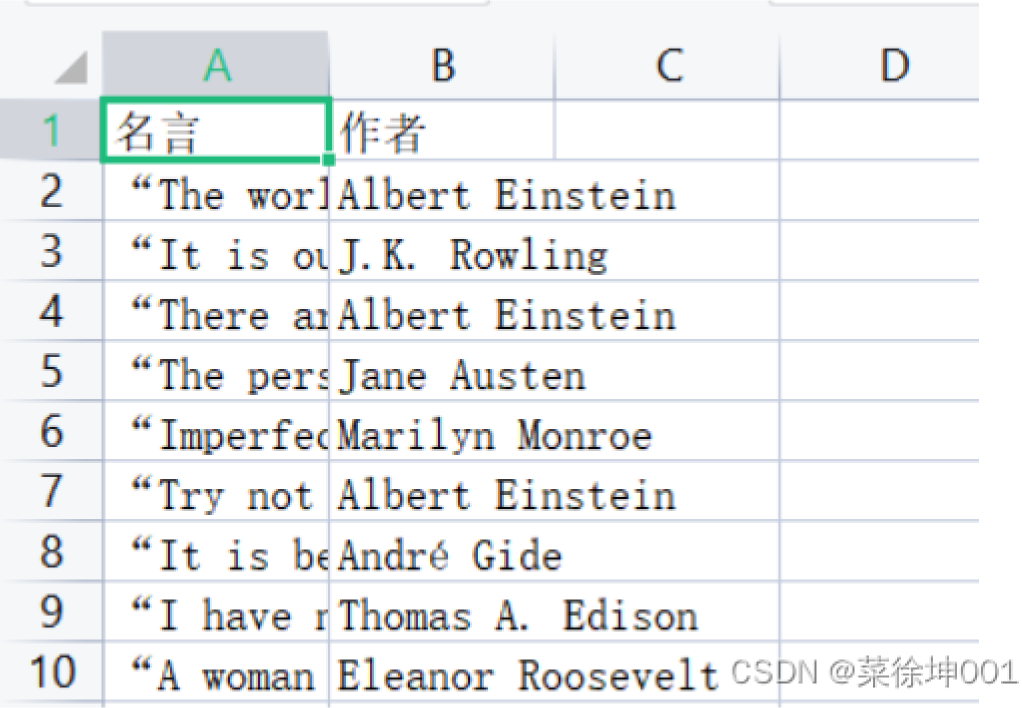
```
<!DOCTYPE html>
<html lang="en">
 ▸<head>…</head>
··· ▾<body> == $0
   ▾<div class="container">
      ::before
     ▸<div class="row header-box">…</div>
      <script src="/static/jquery.js"></script>
     ▸<script>…</script>
     ▾<div class="quote">
       ▾<span class="text">
          "“The world as we have created it is a process of our thinking. It
          cannot be changed without changing our thinking.”"
        </span>
       ▸<span>…</span>
       ▸<div class="tags">…</div>
      </div>
     ▸<div class="quote">…</div>
```

```
element.style {
}
body {                                              main.css:65
  margin-bottom: 60px;
}
body {                                              main.css:2
  font-family: sans-serif;
}
body {                                              bootstrap.min.css:11
  ~~font-family: Georgia,"Times New Roman",Times,serif;~~
  font-size: 16px;
  line-height: 1.42857143;
  color: ■#333333;
  background-color: □#ffffff;
}
body {                                              bootstrap.min.css:11
```

- 爬取信息并存入

```
def write_csv(csv_head,csv_content,csv_path):
    with open(csv_path, 'w', newline='',encoding='utf-8') as file:
        fileWriter =csv.writer(file)
        fileWriter.writerow(csv_head)
        fileWriter.writerows(csv_content)
        print('爬取信息成功')
        quote=driver.find_elements_by_class_name("quote")
#将要收集的信息放在quote_content里
for i in tqdm(range(len(quote))):
    quote_text=quote[i].find_element_by_class_name("text")
    quote_author=quote[i].find_element_by_class_name("author")
    temp=[]
    temp.append(quote_text.text)
    temp.append(quote_author.text)
    quote_content.append(temp)
write_csv(quote_head,quote_content,quote_path)
```

- 爬取成功



	A	B	C	D
1	名言	作者		
2	“The worl	Albert Einstein		
3	“It is ou	J.K. Rowling		
4	“There ar	Albert Einstein		
5	“The pers	Jane Austen		
6	“Imperfec	Marilyn Monroe		
7	“Try not	Albert Einstein		
8	“It is be	André Gide		
9	“I have r	Thomas A. Edison		
10	“A woman	Eleanor Roosevelt		

五、爬取淘宝商品信息

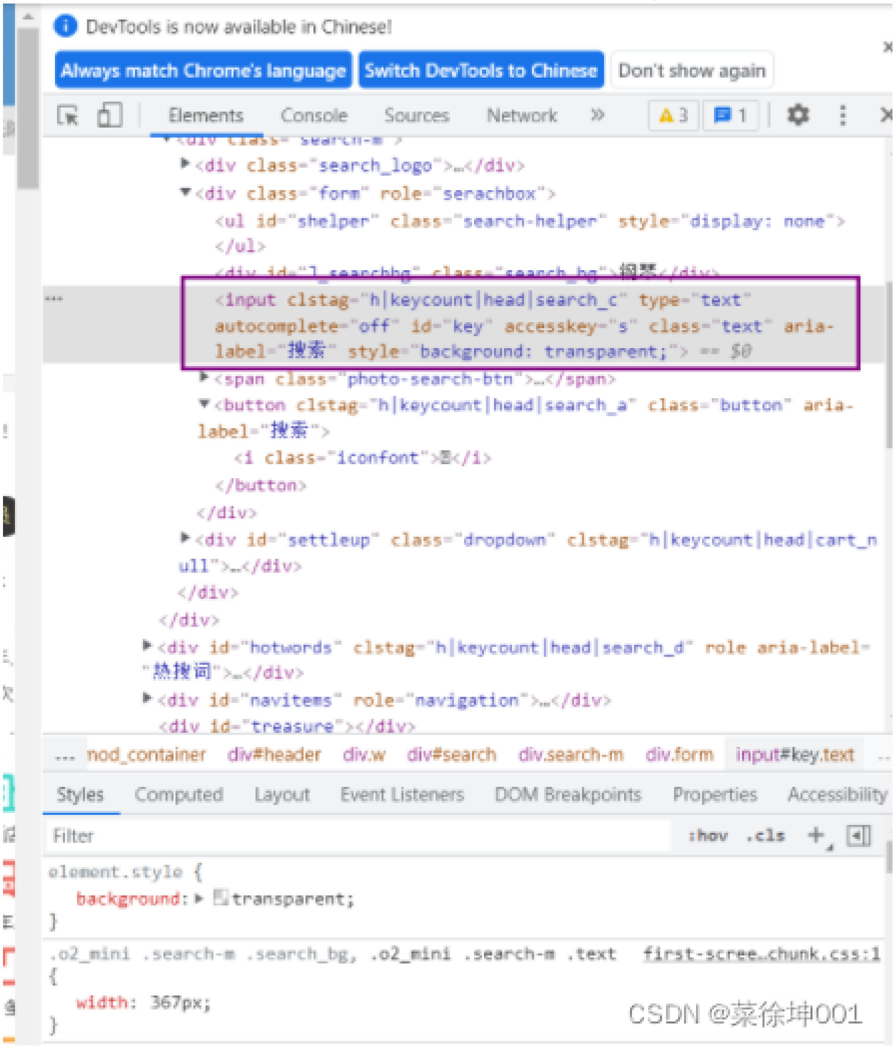
- 导入包并进入淘宝首页

```
from selenium import webdriver
import time
import csv
driver=webdriver.Chrome('D:\\py\\chromedriver_win32\\chromedriver.exe')
driver.get("https://www.taobao.com/")
```

- 定义存放爬取信息的路径和存放内容的列表以及爬取数量

```
#定义存放图书信息的列表
goods_info_list=[]
#爬取200本
goods_num=200
#定义表头
goods_head=['价格','名字','链接']
#csv文件的路径和名字
goods_path='..\\source\\goods.csv'
```

- 进入开发者模式，定位搜索框，可以看到搜索框的id为key，获取按钮输入信息



```
▼<div class="search-m">
  ▶<div class="search_logo">…</div>
  ▼<div class="form" role="serachbox">
     <ul id="shelper" class="search-helper" style="display: none">
     </ul>
     <div id="l_searchbg" class="search_bg">钢琴</div>
     <input clstag="h|keycount|head|search_c" type="text"
     autocomplete="off" id="key" accesskey="s" class="text" aria-
     label="搜索" style="background: transparent;"> == $0
    ▶<span class="photo-search-btn">…</span>
    ▼<button clstag="h|keycount|head|search_a" class="button" aria-
    label="搜索">
       <i class="iconfont"></i>
     </button>
   </div>
  ▶<div id="settleup" class="dropdown" clstag="h|keycount|head|cart_n
  ull">…</div>
  </div>
</div>
▶<div id="hotwords" clstag="h|keycount|head|search_d" role aria-label=
"热搜词">…</div>
▶<div id="navitems" role="navigation">…</div>
 <div id="treasure"></div>
```

```
element.style {
  background: ▶ transparent;
}
.o2_mini .search-m .search_bg, .o2_mini .search-m .text    first-scree…chunk.css:1
{
  width: 367px;
}
```

- 获取输入框并输入信息

```
#向输入框里输入薯片
p_input = driver.find_element_by_id("key")
p_input.send_keys('薯片')
```

- 获取商品价格、名称、链接的函数

```
#获取商品价格、名称、链接
def get_prince_and_name(goods):
    #直接用css定位元素
    #获取价格
    goods_price=goods.find_element_by_css_selector('div.p-price')
    #获取元素
    goods_name=goods.find_element_by_css_selector('div.p-name')
    #获取链接
    goods_herf=goods.find_element_by_css_selector('div.p-
img>a').get_property('href')
    return goods_price,goods_name,goods_herf
```

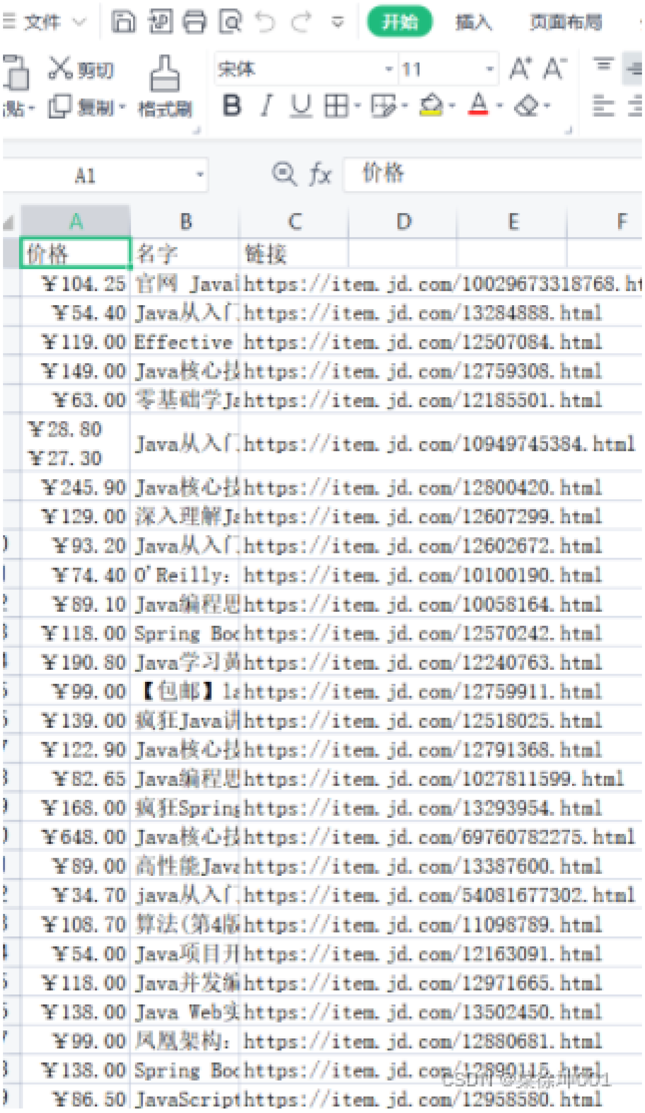
- 定义滑动到页面底部函数，滑动到页面底部会刷出下一页

```
def  drop_down(web_driver):
    #将滚动条调整至页面底部
    web_driver.execute_script('window.scrollTo(0, document.body.scrollHeight)')
```

- 定义爬取一页的函数

```
#获取爬取一页
def crawl_a_page(web_driver,goods_num):
    #获取图书列表
    drop_down(web_driver)
    goods_list=web_driver.find_elements_by_css_selector('div#J_goodsList>ul>li')
    #获取一个图书的价格、名字、链接
    for i in tqdm(range(len(goods_list))):
        goods_num-=1
        goods_price,goods_name,goods_herf=get_prince_and_name(goods_list[i])
        goods=[]
        goods.append(goods_price.text)
        goods.append(goods_name.text)
        goods.append(goods_herf)
        goods_info_list.append(goods)
        if goods_num==0:
            break
    return goods_num
```

- 运行结果



A	B	C	D	E	F
价格	名字	链接			
￥104.25	官网 Javai	https://item.jd.com/10029673318768.ht			
￥54.40	Java从入门	https://item.jd.com/13284888.html			
￥119.00	Effective	https://item.jd.com/12507084.html			
￥149.00	Java核心技	https://item.jd.com/12759308.html			
￥63.00	零基础学Ja	https://item.jd.com/12185501.html			
￥28.80 ￥27.30	Java从入门	https://item.jd.com/10949745384.html			
￥245.90	Java核心技	https://item.jd.com/12800420.html			
￥129.00	深入理解Ja	https://item.jd.com/12607299.html			
￥93.20	Java从入门	https://item.jd.com/12602672.html			
￥74.40	O'Reilly:	https://item.jd.com/10100190.html			
￥89.10	Java编程思	https://item.jd.com/10058164.html			
￥118.00	Spring Boo	https://item.jd.com/12570242.html			
￥190.80	Java学习黄	https://item.jd.com/12240763.html			
￥99.00	【包邮】1z	https://item.jd.com/12759911.html			
￥139.00	疯狂Java讲	https://item.jd.com/12518025.html			
￥122.90	Java核心技	https://item.jd.com/12791368.html			
￥82.65	Java编程思	https://item.jd.com/1027811599.html			
￥168.00	疯狂Spring	https://item.jd.com/13293954.html			
￥648.00	Java核心技	https://item.jd.com/69760782275.html			
￥89.00	高性能Java	https://item.jd.com/13387600.html			
￥34.70	java从入门	https://item.jd.com/54081677302.html			
￥108.70	算法(第4版	https://item.jd.com/11098789.html			
￥54.00	Java项目开	https://item.jd.com/12163091.html			
￥118.00	Java并发编	https://item.jd.com/12971665.html			
￥138.00	Java Web实	https://item.jd.com/13502450.html			
￥99.00	凤凰架构:	https://item.jd.com/12880681.html			
￥138.00	Spring Boo	https://item.jd.com/12890115.html			
￥86.50	JavaScript	https://item.jd.com/12958580.html			

六、总结

Selenium是一系列基于Web的自动化工具，提供一套测试函数，用于支持Web自动化测试。函数非常灵活，能够完成界面元素定位、窗口跳转、结果比较。

七、参考链接

https://blog.csdn.net/wanglian2017/article/details/72843984?ops_request_misc=%257B%2522request%255Fid%2522%253A%2522163974182016780269820611%2522%252C%2522scm%2522%253A%252220140713.130102334.pc%255Fall.%2522%257D&request_id=163974182016780269820611&biz_id=0&utm_medium=distribute.pc_search_result.none-task-blog-2~all~first_rank_ecpm_v1~rank_v31_ecpm-1-72843984.pc_search_result_cache&utm_term=selenium%E4%BB%8B%E7%BB%8D&spm=1018.2226.3001.4187 https://blog.csdn.net/weixin_44019406/article/details/88681312