爬虫主体使用Python+selenium写成

代码的引用库

from selenium.webdriver.common.by import By  
from selenium import webdriver  
from selenium.webdriver.common.keys import Keys  
import time

我将爬虫A分成两个函数来写，一个用来搜索，一个用来获取商品的review。当然，在这之前我要先把浏览器的驱动导入

driver = webdriver.Edge()

如果你没有把驱动加载系统的环境变量中，请在括号内把你浏览器的驱动地址写入

第一部分，搜索函数：

def search(keywords):  
 driver.get('https://www.amazon.cn/')  
 inputtag = driver.find\_element(By.ID,'twotabsearchtextbox')  
 inputtag.send\_keys(keywords)  
 inputtag.send\_keys(Keys.ENTER)

首先定义search这个函数

第一行drive，get（）即打开亚马逊官网

第二行drive.find\_element（BY.ID，）即用ID去定位元素，这里定位的是亚马逊主页的搜索框。

第三行和第四行send. Keys即输入关键词和回车。

第二部分，获取商品的review

def review():  
 result = driver.find\_element(By.XPATH,'//\*[@id="search"]/div[1]/div[1]/div/span[1]//span/span[4]').text  
a = int(result)  
a = a-1  
for i in range(a):  
 h2s = driver.find\_elements(By.XPATH, '//\*[@id="search"]//ancestor::h2')  
 time.sleep(2)  
 for h2 in h2s:  
 print(h2.find\_element(By.CSS\_SELECTOR, "span").text)  
 print(h2.find\_element(By.XPATH, '//\*[@id="search"]//div/a/span[1]').text)  
 forward = driver.find\_element(By.XPATH, '(//\*[@id="search"]//div/div/span/a)[last()]')  
 forward.click()  
 time.sleep(2)

h2s = driver.find\_elements(By.XPATH, '//\*[@id="search"]//ancestor::h2')  
这句话是获取到search下h2元素的所有父元素，方便后面使用

print(h2.find\_element(By.CSS\_SELECTOR, "span").text)  
 print(h2.find\_element(By.XPATH, '//\*[@id="search"]//div/a/span[1]').text)  
这个是输出商品的名称和价格

forward = forward = driver.find\_element(By.XPATH, '(//\*[@id="search"]//div/div/span/a)[last()]')

time.sleep(2)

向后翻页，

完整代码如下

from selenium.webdriver.common.by import By  
from selenium import webdriver  
from selenium.webdriver.common.keys import Keys  
import time  
  
driver = webdriver.Edge()  
def search(keywords):  
 driver.get('https://www.amazon.cn/')  
 inputtag = driver.find\_element(By.ID,'twotabsearchtextbox')  
 inputtag.send\_keys(keywords)  
 inputtag.send\_keys(Keys.ENTER)  
  
def review():  
 result = driver.find\_element(By.XPATH,'//\*[@id="search"]/div[1]/div[1]/div/span[1]//span/span[4]').text  
a = int(result)  
a = a-1  
for i in range(a):  
 h2s = driver.find\_elements(By.XPATH, '//\*[@id="search"]//ancestor::h2')  
 time.sleep(2)  
 for h2 in h2s:  
 print(h2.find\_element(By.CSS\_SELECTOR, "span").text)  
 print(h2.find\_element(By.XPATH, '//\*[@id="search"]//div/a/span[1]').text)  
 forward = driver.find\_element(By.XPATH, '(//\*[@id="search"]//div/div/span/a)[last()]')  
 forward.click()  
 time.sleep(2)

if \_\_name\_\_ == '\_\_main\_\_':  
 search('Beer glasses')  
 review()