本节知识点概述

爬取淘宝商品数据

全能的Selenium

Selenium的环境搭建

Selenium的简单使用

实战环节

使用Selenium爬取淘宝

思路讲解+代码分享

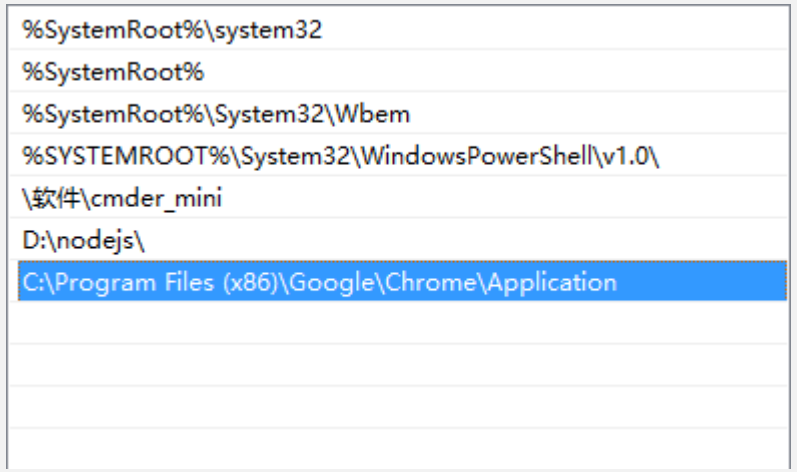
说在最后

经验分享

回顾课程

Selenium的环境搭建

1. 安装：在终端输入 pip install selenium
2. 下载：下载chromedriver，解压后放在...\Google\Chrome\Application\
3. 环境变量：将改目录添加至环境变量



Selenium的简单使用

1. 导入Selenium包
2. 选择Chrome浏览器
3. 打开百度首页
4. 找到搜索框，输入关键词
5. 打印源码

```
# -*- coding:utf-8 -*-
#!/usr/bin/env python

from selenium import webdriver

driver = webdriver.Chrome()
driver.get("http://www.baidu.com")
```

```
# -*- coding:utf-8 -*-
#!/usr/bin/env python

from selenium import webdriver
from selenium.webdriver.common.keys import Keys

driver = webdriver.Chrome()
driver.get("http://www.baidu.com")

elem = driver.find_element_by_xpath('//*[@id="kw"]')
elem.send_keys("Python selenium", Keys.ENTER)
print(driver.page_source)
```

使用Selenium爬取淘宝

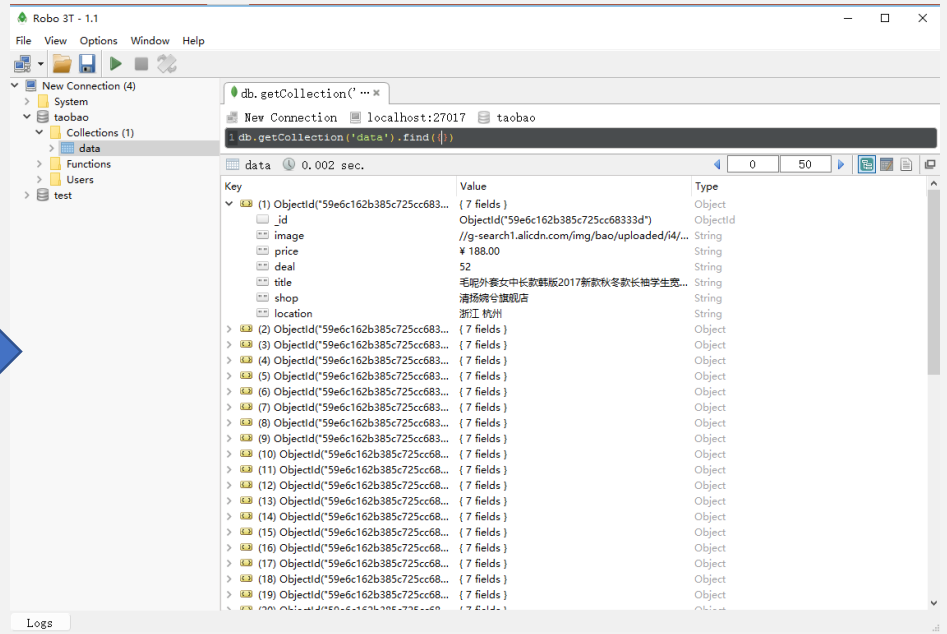
```python
# -*- coding:utf-8 -*-
#!/usr/bin/env python

from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.common.exceptions import TimeoutException
from pyquery import PyQuery as pq
from pymongo import MongoClient
import re

browser = webdriver.Chrome()
wait = WebDriverWait(browser, 10)

client = MongoClient()
db=client.taobao
data = db.data

def search(kd):
    try:
        browser.get('https://www.taobao.com/')
        input = wait.until(EC.presence_of_element_located((By.CSS_SELECTOR, "#q")))
        submit = wait.until(EC.element_to_be_clickable((By.CSS_SELECTOR,'#J_TSearchForm > div.search-button > button')))
        input.send_keys(kd)
        submit.click()
        total = wait.until(EC.presence_of_element_located((By.CSS_SELECTOR,'#mainsrp-pager > div > div > div > div.total')))
        get_products()
        return total.text
    except TimeoutException:
        return search()
```

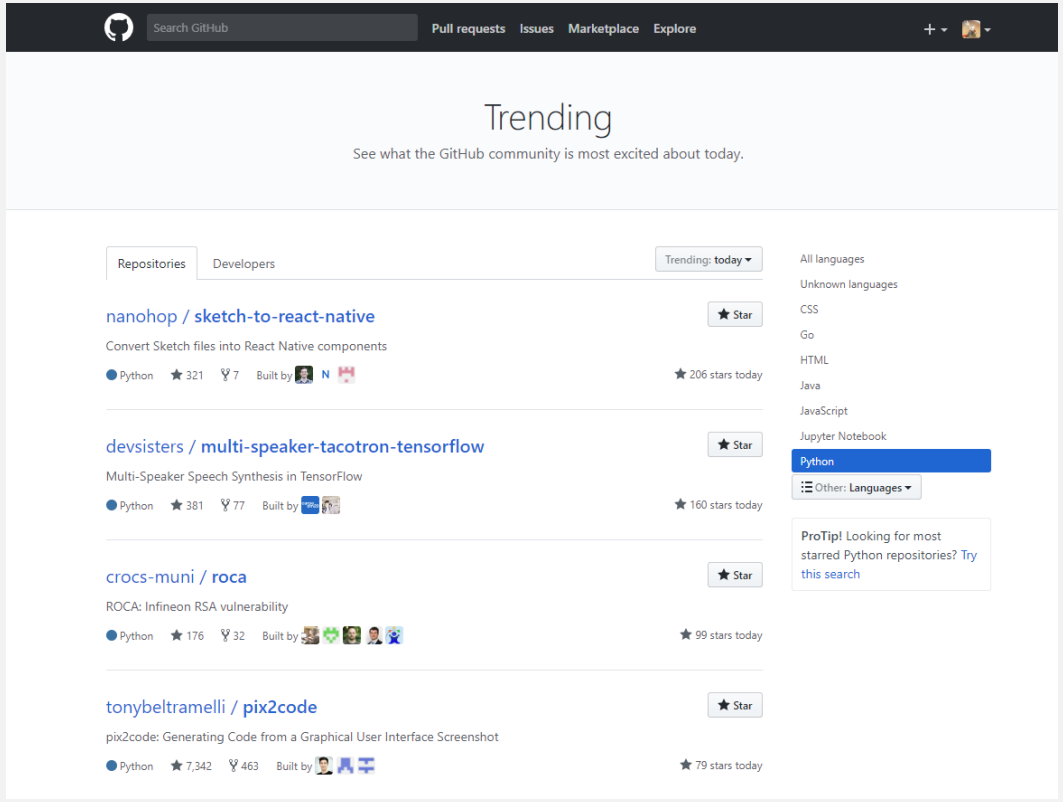


说在最后

常用的学习方法

经常浏览的网站

一年的学习经历



说在最后

- 第一课_什么是爬虫
- 第二课_初识Python爬虫
- 第三课_使用Requests爬取豆瓣短评
- 第四课_使用xpath解析豆瓣短评
- 第五课_使用pandas保存豆瓣短评数据
- 第六课_爬取知乎用户数据
- 第七课_爬取拉勾职位信息
- 第八课_爬取淘宝商品价格

完成作业

- 了解Selenium是什么
- 用Selenium去爬取更多网站
- 回顾入门课程，迎接进阶部分

爬虫工程师

更多数据科学课程，上DC学院： class.pkbigdata.com

关注 DataCastle

关注造数

造数