everyday 用户手册

1. 使用步骤

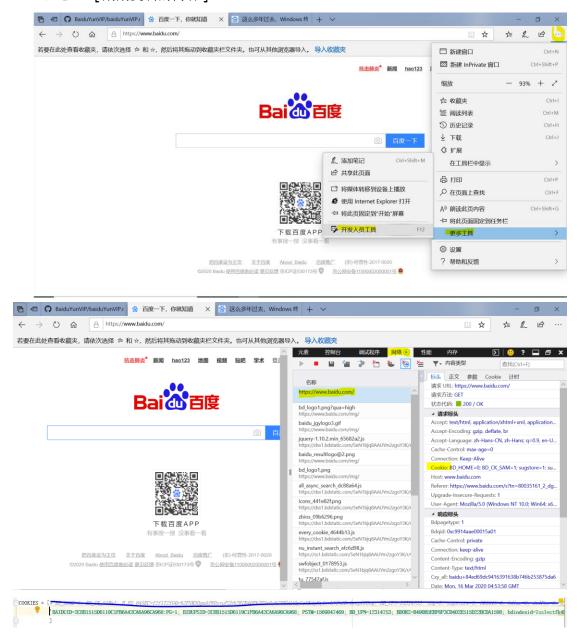
1.1 获得 cookies 值

打开浏览器,登录百度账号。百度账号网址:

https://github.com/lpg-it/BaiduYunVIP/blob/master/baiduYunVIP.md

登陆成功,点击浏览器的【...】键,【更多工具】->【开发人员工具】,点击 network/网络,键盘输入【ctrl+R】,出现【www.baidu.com】,点击,复制 cookies 里的内容,到 config.py 中的 cookie 粘贴。

COOKIES = ['刚刚复制的内容']



1.2 更改关键词

打开 demo.py,将肺炎更改为关键词。支持多关键词,但是!! 最好不要使用,被反爬了就得不偿失了

```
Jif __name__ == "__main__":
    keywords = ['肺炎']
    result = []
    for keys, value in CITY_CODE.items():
        baidu_index = BaiduIndex(keywords, '2019-12-01', '2020-03-15', value)
1
        for index in baidu_index.get_index():
J
            # print(index)
ì
            ##index.update({"省份":keys})
            index. update({"city":keys})
            result.append(index)
            #print(result)
            pd. DataFrame (result). to_csv('肺炎.csv', encoding='gbk')
ì
ì
        print("{0} is successfully saved".format(keys))
```

1.3 运行脚本

打开命令行窗口,输入【cd ‹demo.py 所在的路径›】,实例:

```
Microsoft Windows [版本 10.0.18362.657]
(c) 2019 Microsoft Corporation。保留所有权利。

C:\Users\l>cd C:\Users\l\Desktop\everyday\spider-BaiduIndex-master\new_spider_without_selenium\
```

(注意, ↔不要输入,不要到 demo.py,只需到达 demo.py 的上一级)

【回车】,输入【python demo.py】,坐等运行结束

2 注意事项

- 2.1 运行过程会打印访问页面,这是为了出错好调试,不用管.
- 2.2 每次更改关键词,建议更改 ip, 更改 ip 只需要断开网络,重连。可以百度搜索 【ip 地
- 址】看是否改变。一定要更改 cookies,如何更改见上。这是为了反反爬
- 2.3 禁止在一台电脑上运行多个进程,会遭受猛烈的反爬攻击。但是可以在多台电脑上运行,多台电脑运行一定要选择不同的 cookies。
- **2.4** 更改代码建议使用正规的代码编辑器,由于 python 是缩进敏感的语言,使用 notepad 修改可能会破坏缩进结构。(代码里所有的缩进都是 TAB 键)
- 2.5 有问题随时增加