

i春秋python_用python爬取i春秋的课程名

原创

[weixin_39648469](#) 于 2020-12-23 06:01:02 发布 35 收藏

文章标签: [i春秋python](#)

版权声明: 本文为博主原创文章, 遵循 [CC 4.0 BY-SA](#) 版权协议, 转载请附上原文出处链接和本声明。

本文链接: https://blog.csdn.net/weixin_39648469/article/details/111894662

版权

看课中内容是用get请求进行爬取课程, 自己实践的时候发现已经被改成post请求了, 下面开始

我用的火狐, 然后就是F12, 点击网络, 可能会有很多包, 但不影响, 点击删除就行, 然后点击第二页, 其他页面也行, 可以看出总共是有17页的

然后会出现很多包, 找到下面这个, 点击查看

点击后会看到消息头请求网址, 这个就是我们要用的url

复制到代码中

```
i_url = 'https://www.ichunqiu.com/courses/ajaxCourses'
```

接下来查看参数

然后发现“pageIndex”, 很明显是指的第几页, 可以试一下第四页数值是4, , , 。这是post要用的表单数据, 搞进代码, 因为pageIndex值需要改变, 我将它设为空值

```
post_data =  
{'courseTag': '', 'courseDifficulty': '', 'isExp': '', 'producerId': '', 'orderField': '', 'orderDirection': '', 'pageIndex': '', 'tagType': ''}
```

由于i春秋设置了防爬虫处理, 因此我们需要添加headers, headers就不多说了(不想暴露自己的配置, 我不会告诉你我还用的xp), 用火狐很简单, 点击编辑重发, 复制就行。接下来就是查看响应

很明显这是json数据, 先import json再说, 可以找到课是在course中的result下, 从0开始, 我们可以用“len(raw_data[course]['result'])”查看该页面有几个课程, 用raw_data[course]['result'][0]['courseName']来获取第一个课程名字, 下面是代码

```
#-----#Author:Glasses#-----
```

```
import requests  
import json  
def getData():
```

```
    i_url = 'https://www.ichunqiu.com/courses/ajaxCourses'  
    headers = {
```

```
        }  
    #headers 自己填
```

```
    for p_index in range(1, 18):
```

```
post_data=
{'courseTag':"','courseDiffcuty':"','lsExp':"','producerId':"','orderField':"','orderDirection':"','pageIndex':"','tagType':"}

post_data['pageIndex'] =p_index

r= requests.post(i_url,headers=headers,data=post_data,timeout=10)

raw_data=json.loads(r.text)for i in range(len(raw_data['course']['result'])):print raw_data['course']['result'][i]
['courseName']

getData()
```

来i春秋一起学习吧，

进阶吧！小白，只要摔不死我就继续摔！