

python爬取i春秋文章标题

原创

努力的学渣# 于 2020-01-01 10:46:27 发布 1035 收藏

分类专栏: [python](#) 文章标签: [爬虫](#) [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_41598660/article/details/103791281

版权



[python](#) 专栏收录该内容

5 篇文章 0 订阅

订阅专栏

最近自学python爬虫入门, 好奇想尝试一下

环境: python3.8

目标: i春秋的视频文章标题

代码如下

```
import requests
import re

def getTitle():
    url = 'https://bbs.ichunqiu.com/portal.php'
    headers = {
        'Host': 'bbs.ichunqiu.com',
        'Connection': 'close',
        'Cache-Control': 'max-age=0',
        'Upgrade-Insecure-Requests': '1',
        'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/55.0.2883.87 Safari/537.36',
        'Accept': 'text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8',
        'Accept-Language': 'zh-CN,zh;q=0.8',
    }
    req = requests.get(url=url,headers=headers)
    html = req.content
    titlere = r'target="blank" class="ui_colorG" style="color: #555555;">(.*?)</a></h3>'
    title = re.findall(titere,html.decode('utf-8'))
    num=0
    for titles in title:
        num=num+1
        print (num,titles)
if __name__ == '__main__':
    getTitle()
```

结果显示

```
E:\visual studio code > paching.py > getTitle
1 import requests
2 import re
3
4 def getTitle():
5     url = 'https://bbs.ichunqiu.com/portal.php'
6     headers = {}
7     headers['Host'] = 'bbs.ichunqiu.com',
8     headers['Connection'] = 'close',
9     headers['Cache-Control'] = 'max-age=0',
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL 2: Python

46 【超详细入门教程】通过xss获取管理员明文账号密码
47 500万人用山寨12123查违章？高仿APP不只是坑钱
48 llvm Pass简单编写
49 Citrix严重漏洞可能会使成千上万的公司遭受攻击
50 智联SRC双旦活动 2倍积分奖励
51 Netcat-NC神器使用总结
52 一种有趣的帐户接管手段
53 利用burp collaborator模块获得服务器端的真实ip（利用pingback
54 在Chrome中使用WebRTC ICE服务进行端口扫描
55 Xctf 是联赛还是比赛？
56 Kali for 树莓派安装Mysql的一个坑
57 一立陶宛男子因盗窃Google, Facebook的1.2亿美元被判入狱五年
58 某校园网任意用户密码重置
59 lfi phpinfo
60 41款APP侵权:10亿用户中招 QQ小米搜狐人人全在内
61 360专家划重点: 5G时代如何打好安全牌!
62 Windows 10资源管理器搜索问题一箩筐 微软正加紧修复
63 怎么通过浏览器IntentScheme触发app本地dos漏洞?
64 各种window权限维持姿势合集
65 WAES: Web Auto Enum & Scanner https://blog.csdn.net/weixin_41598660