

spider

名称	链接	说明
httpbin.org	链接	靶机
Playwright	链接	动态渲染 JS
GerapyPlaywright	链接	Playwright in Scrapy
Gerapy	链接	Distributed Crawler Management Framework Based on Scrapy, Scrapyd, Scrapyd-Client, Scrapyd-API, Django and Vue.js etd.
scrapy cookies	链接	将cookies存文件
py import	链接	import 疑难杂症
反爬	外部链接	web 端
acrawler.js反爬加密	掘金每日签到用到了	链接
mitmporxy安装及使用教程	链接	安装和使用
splash-docs	链接	官网
splash-scrapy	link	scrapy-splash;但有个engine=chromium无法填充的bug issue , 会导致有些场景渲染失败

在 console 注入 jquery

```
var
importJs=document.createElement('script');importJs.setAttribute("type","text
/javascript");importJs.setAttribute("src",
'https://ajax.microsoft.com/ajax/jquery/jquery-1.7.min.js');document.getElem
entsByTagName("head")[0].appendChild(importJs);
```

提取数组中的 href

```
$('section a').each(function(index){
    console.log($(this).attr("href"));
})
```

From:

<https://rd.irust.top/> - 学习笔记

Permanent link:

<https://rd.irust.top/doku.php?id=python:spider&rev=1664438160>

Last update: 2022/09/29 15:56

