

## 豆瓣电影--如何正确获取网页

笔记本: reptile\_draft

创建时间: 2021/6/27 19:40

更新时间: 2021/7/11 14:44

作者: 134exetj717

URL: about:blank

### ▼ GET

Scheme: https

Host: movie.douban.com

Filename: /j/chart/top\_list

type: 11

interval\_id: 100:90

action:

start: 0

limit: 20

地址: 154.8.131.172:443

所有	HTML	CSS	JS	XHR	字体	图像	媒体	WS	其他
状态	方法	域名	文件	发起者	类型	传输	大小		
200	GET	movi...	top_list?type=11&interval_id=1	jquery.js...	json	1.13 KB	94...		
200	GET	movi...	top_list_count?type=11&in	jquery.js...	json	553 字节	56...		
200	GET	movi...	top_list?type=11&interval_	jquery.js...	json	6.34 KB	17...		

3 个请求 已传输 18.68 KB / 8.01 KB 完成: 3.53 秒 DOMContentLoaded: 1.68 秒

消息头 Cookie 请求 响应 耗时 栈跟踪 安全性

过滤消息头 拦截 重发

GET https://movie.douban.com/j/chart/top\_list?type=11&interval\_id=100:90&action=8&start=0&limit=20

状态 200 OK ⓘ

版本 HTTP/1.1

传输 6.34 KB (大小 17.71 KB)

Referrer 政策 unsafe-url

响应头 (539 字节) 原始

Cache-Control: must-revalidate, no-cache, private

Connection: keep-alive

Content-Encoding: br

Content-Type: application/json; charset=utf-8

```
# -*- coding = utf-8 -*-
# @Time : 2021/6/23 20:41
# @Author : 希杰
# @file : pj_translate.py
```

```

# @software : PyCharm
import json

import requests
if __name__ == '__main__':
    post_url = 'https://fanyi.baidu.com/sug'
    #进行UA伪装
    headers = {
        'User-Agent': 'Mozilla / 5.0(Windows NT 10.0; Win64; x64; rv: 89.0)
Gecko / 20100101 Firefox / 89.0'
    }
    #post请求参数处理（同get请求一致）
    word = input('enter a word:\n')
    data = {
        'kw': word
    }
    #请求发送
    response = requests.post(url=post_url,data=data,headers=headers)
    #获取响应数据：json()方法返回的是obj（如果确认响应数据是json类型，才可以使用
    json()）对象
    dic_obj = response.json()
    print(dic_obj)
    #持久化存储
    fileName = word + '.json'
    fp = open('./' + fileName,'w',encoding='utf-8')
    json.dump(dic_obj,fp=fp,ensure_ascii=False)    #将你的对象内容存到文件中，确保
    #asciill码是无效的，因为中文不能用asciill码翻译
    print('over!!!')

```