

化妆品生产许可信息管理系统服务平台 (XHR请求的练习)

笔记本: reptile_draft

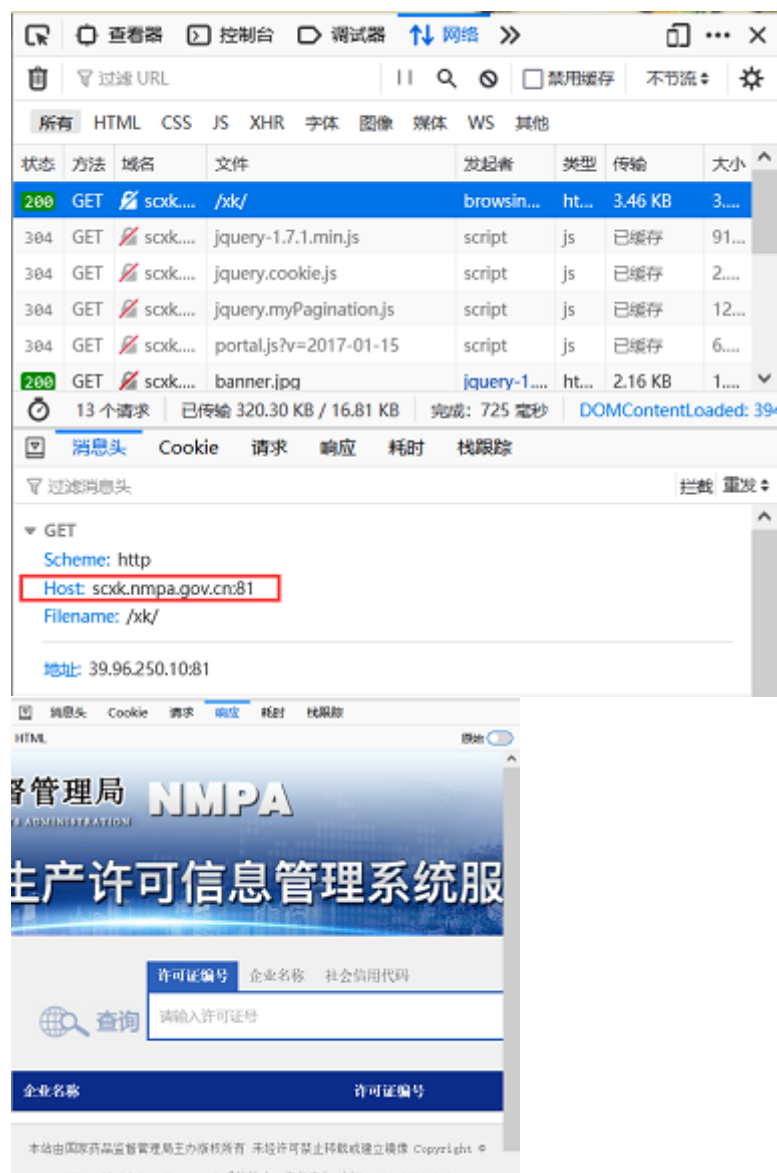
创建时间: 2021/6/27 20:04

更新时间: 2021/7/3 15:45

作者: 134exetj717

URL: about:blank

- 动态加载数据
 - 由阿贾克斯加载得到



上图可见, 明明地址是一样的, 但是我们观看响应的时候, 并没有相关企业的信息, 而是只有模板, 因此判断数据来源并不是url静态加载, 而很可能是动态加载, 拭目以待:



由此图可以发现，XHR（阿贾克斯请求）有这么一栏，里面的响应恰好就是我们的企业数据，因此判定数据来源就是动态加载。

- 解决方法：
 - 通过观察我们发现，有些网站虽然没有给资源的url，但是我们可以通过她自身提供的ID（唯一索引），与总网站的url域名相结合，就能够得到该资源的url。
- 当然，我们可能会遇到一个更为糟糕的情况，就是得到的资源网站也是动态加载的，那么此时如何判断是否动态加载呢？
 - 我们可以通过抓包工具，即检查的请求头得到的相应数据是否为空。
- 如果确认是动态加载的，那么我们就不能直接得到该资源的信息，此时还需要用到XHR来获取对应的相应信息。
- 同时，教大家一个好方法来获取json的规范格式化，比如headers不会自动换行这样，我们只需要搜索 `json解析` 就可以在线翻译，规范化，最后再复制即可。

```
# -*- coding = utf-8 -*-
# @Time : 2021/7/2 16:12
# @Author : 希杰
# @file : medicine.py
# @software : PyCharm

import requests
import json
if __name__ == "__main__":
    url = 'http://scxk.nmpa.gov.cn:81/xk/itownet/portalAction.do?method=getXkzsList'
    headers = {
        'User-Agent': 'Mozilla / 5.0(Windows NT 10.0; Win64; x64; rv: 89.0) Gecko / 20100101 Firefox / 89.0'
    }
    for page in range(1,6):
        page = str(page)
        data = {
            "on": "true",
            "page": page,
            "pageSize": "15",
            "productName": "",
            "conditionType": "1",
            "applyname": "",
```

```
        "applysn": ""
    }
    json_ids = requests.post(url=url,headers=headers,data=data).json()
    id_list = []
    for item in json_ids['list']:
        id_list.append(item['ID'])
    #获取企业详情数据
    post_url = 'http://scxk.nmpa.gov.cn:81/xk/itownet/portalAction.do?method=getXkzsById'
    all_data_list = []
    for item in id_list:
        data = {
            'id': item
        }
        detail_json = requests.post(url=post_url,data=data,headers=headers).json()
        # print(detail_json)
        all_data_list.append(detail_json)
    fp = open('./allData.json','w',encoding='utf-8')
    for item in all_data_list:
        json.dump(all_data_list,fp,ensure_ascii=False)
    print('over!!!')
```