

带数据击剑爬虫任务报告

爬虫任务：内江网络问政平台

http://www.njwlwz.gov.cn/MailBox_All.aspx?Mode=List&Open=1

任务团队：

带数据击剑（鲁艮，宋书波，王晓峰）

数据概况：

共 977 页，每页（除最后一页外）20 行数据，从 2013-7-3-2021-11-22 共 19536 条数据

任务完成情况：

可以正常爬取头部标题数据和内容数据

可以正常更新数据，并添加到历史数据中

由于使用 selenium 库，用时略长

说明介绍：

1、安装了 chromedriver 到项目的环境目录下，使用的 google 的 chrome 浏览器

2、最开始使用 requests 库，发现内江网页的数据是动态渲染的，使用 requests 库无法将网页源代码内的内容数据正常爬取下来，因此使用 selenium 库，自动化测试，模拟人的操作，selenium+webdriver 组合的方法有效解决页面动态加载元素难以获取的问题，将网页源代码的内容正常爬取下来。

3、跳转到爬取具体内容网页时，发现内容是在内嵌网页中的，网页源代码中没有具体的内容，通过 ID 定位到 iframe，找到嵌套网页，然后将具体的问政内容爬取下来。

4、header_crawler.py 是将头部标题内的数据爬取下来，并且保存到 header_crawler_data.xlsx 文件中。

头部标题内容是：

投诉	2021-11-22	关于绕城高速公路破坏房屋寻求领导给予帮助解决的问题	范家萍	市委书记	东兴区委	已办结
求助	2021-11-22	关于绕城高速公路破坏房屋寻求帮助与解决	范家萍	市长	东兴区政府	已办结
求助	2021-11-22	关于绕城高速公路破坏房屋寻求帮助解决的问题	范家萍	市委书记	东兴区委	已办结
建议	2021-11-22	希望得到支持	网友	资中县委	资中县委	已办结
咨询	2021-11-22	物业咨询	孔勇	市长	市中区政府	已办结

classification：类别，time：时间，title：标题，send_name：发送人，

receive_name：接收人，answer_name：回复人，state：状态

5：body_crawler.py 是将具体的问政内容以及回复爬取下来，并且将内容保存到 body_crawler_data.xlsx。

具体内容是：

【建议】希望得到支持

区县信箱

网友：2021-11-22 12:30:49

状态：已办结

前段时间中国铁路成都局进行了调整列车运行图。资中北站跨局直通列车及管内车次大幅减少资中北站停靠。特别是长途车次减幅十分明显。资中北站通达城市大幅减少。资中人民出行十分不利。因车次的减少。现在成都至资中和重庆至资中最大间隔3~4个小时才有一趟车！反观隔壁隆昌车次大幅增多且增加长途车次很多，出行越来越便利。资中县132万人口比隆昌多了近一半人多，需求应该更大才是，多了几十万人口，客流量不言而喻是能够得到支持的。反而现在资中北站比隆昌北站还少很多。还有1个多月就春运了，中国铁路成都局必将再次调整列车运行图。希望县委、县政府能够提前介入争取资中人民出行最大利益。从财政资金多挤出一点，为资中人民出行多着想一点。把资中交通越搞越好。希望能够得到支持谢谢。多争取北上南下的长途车次，加大资中北站往返成都、重庆的车次。资中人民感谢你们，谢谢。

电话:17390331168 邮编: 地址:

☒公开 ☒允许选登

回复

中共资中县委：2021-11-24 16:43:28

“网友”您好！

来信收悉。您反映的问题经县发改局核实，现回复如下：

关于您反映车次减少的问题，主要是由于疫情影响停运了一部分车次，现疫情得到控制，车次将逐渐恢复。另我县也多次与成都铁路局协商，争取增加车次停靠，多争取北上南下的长途车次，方便我县132万人民出行，目前中国铁路成都局还未正式公布调整列车运行图。

如您有新的意见和建议可向县发改局反映和咨询，联系人：李坪，联系电话：0832—5605751。

谢谢来信！

ask_name: 问政人姓名, ask_time: 问政时间, ask_body: 问政的具体内容,

answer_name: 回复人姓名（组织）, answer_time: 回复时间,

answer_body: 回复具体内容

6、up_data.py 是更新程序，将更新的数据（包括头部标题和具体的内容）放到 new_data.xlsx,然后将新的数据同步追加到 body_crawler_data.xlsx 文件和 header_crawler_data.xlsx 文件。已更新了 11 月 23 日的四条数据。