爬虫任务二

完成人：廖嘉欣

江北区：

<http://www.cqjbjyzx.gov.cn/lbv3/n_newslist_zz_item.aspx?ILWHBNjF4clKo8UY2fiQHA>==

大渡口区：

<http://www.ddkggzy.com/lbv3/n_newslist_zz_item.aspx?ILWHBNjF4clKo8UY2fiQHA>==

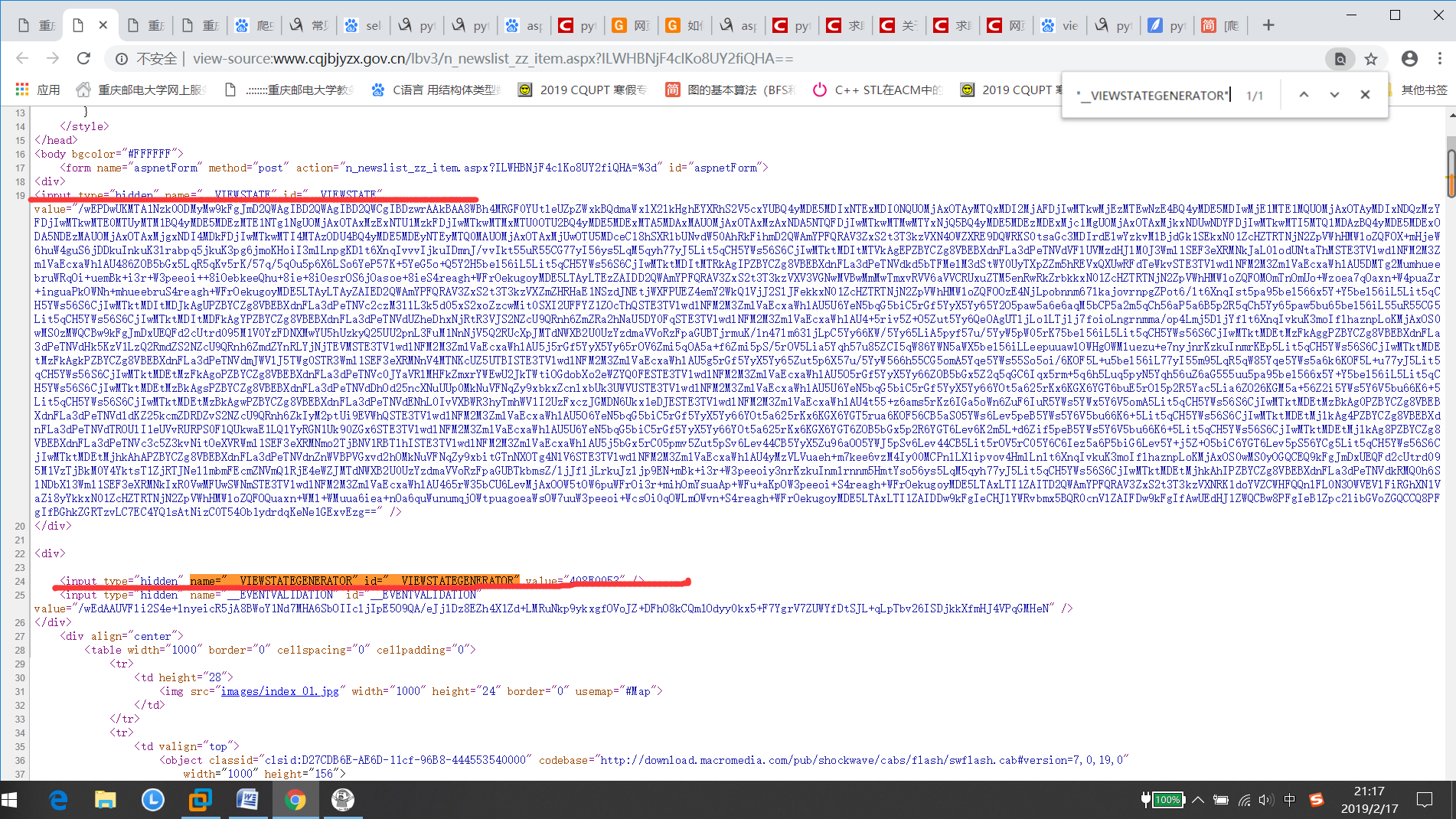
渝中区：

<http://www.cqyzbid.com/cqyzwz/jyxx/003001/003001004/MoreInfo.aspx?CategoryNum=003001004>

江北区：

解决翻页问题

发现跟任务一动态页面不一样（即不是ajax），考虑aspx

发现区别，即aspx不是抓包，（没有找到，），而是post里面做文章

然后百度：

**ViewState和EVENTVALIDATION的原理**

**参考博客：**[**https://blog.csdn.net/younghaiqing/article/details/71601860**](https://blog.csdn.net/younghaiqing/article/details/71601860)

**本次作业依然采取scrapy 框架**

只是post**ViewState和EVENTVALIDATION是没有效果的**

**Post参数**

**伪代码：**

**def1:**

**提取title,time,以及链接，把每一页的链接保存到本地或者全局变量列表里面，爬完了所有父网页，才进入子网页**

**传参链接到下一函数（改变传递的URL即response,用循环遍历全局变量列表，yield写在循环里）**

**def2:**

进入子网页提取

这次表格的每一行来处理