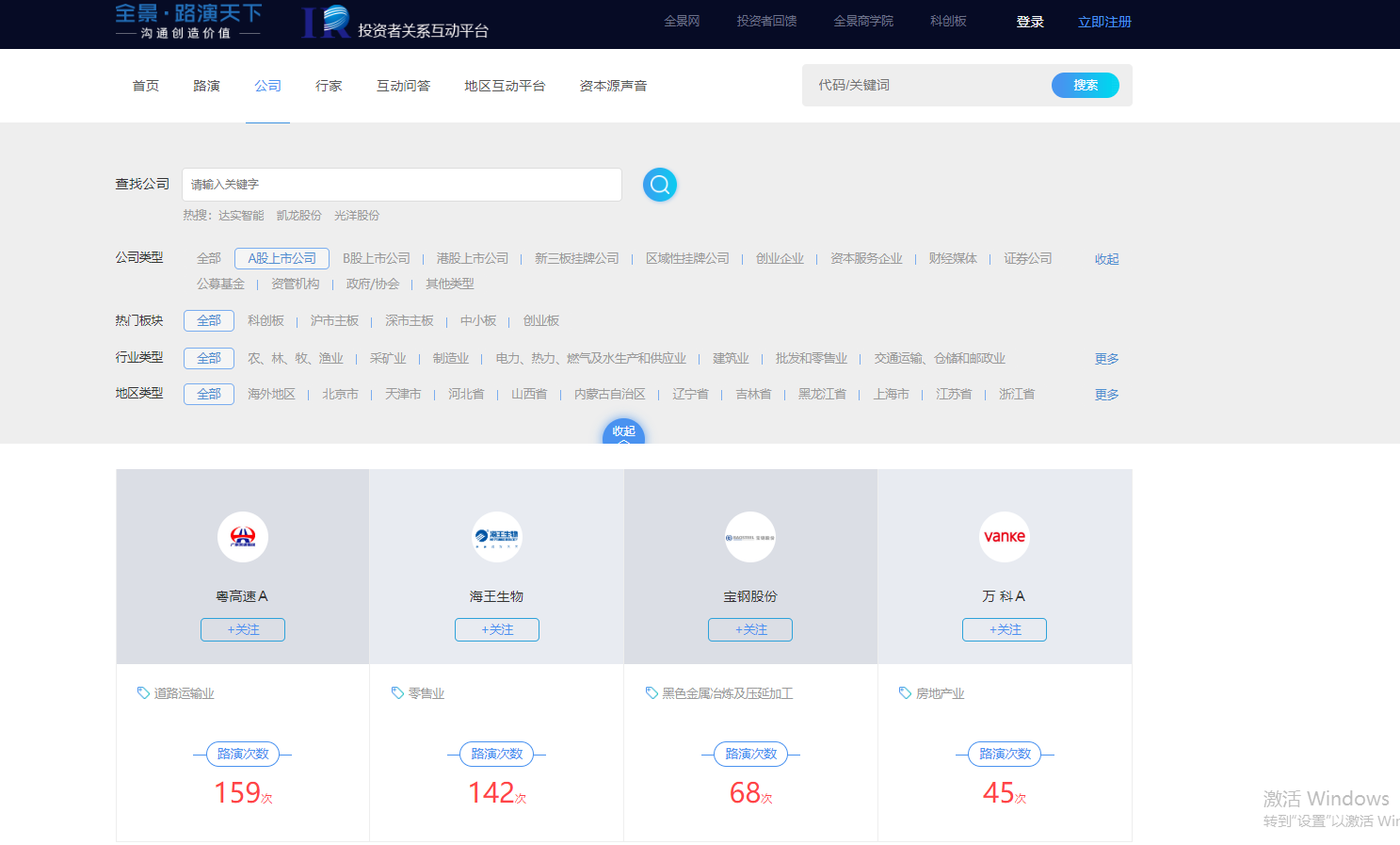
http://ir.p5w.net/company/

根据这个抓取问答股票代码，然后写入问答股票代码excel中

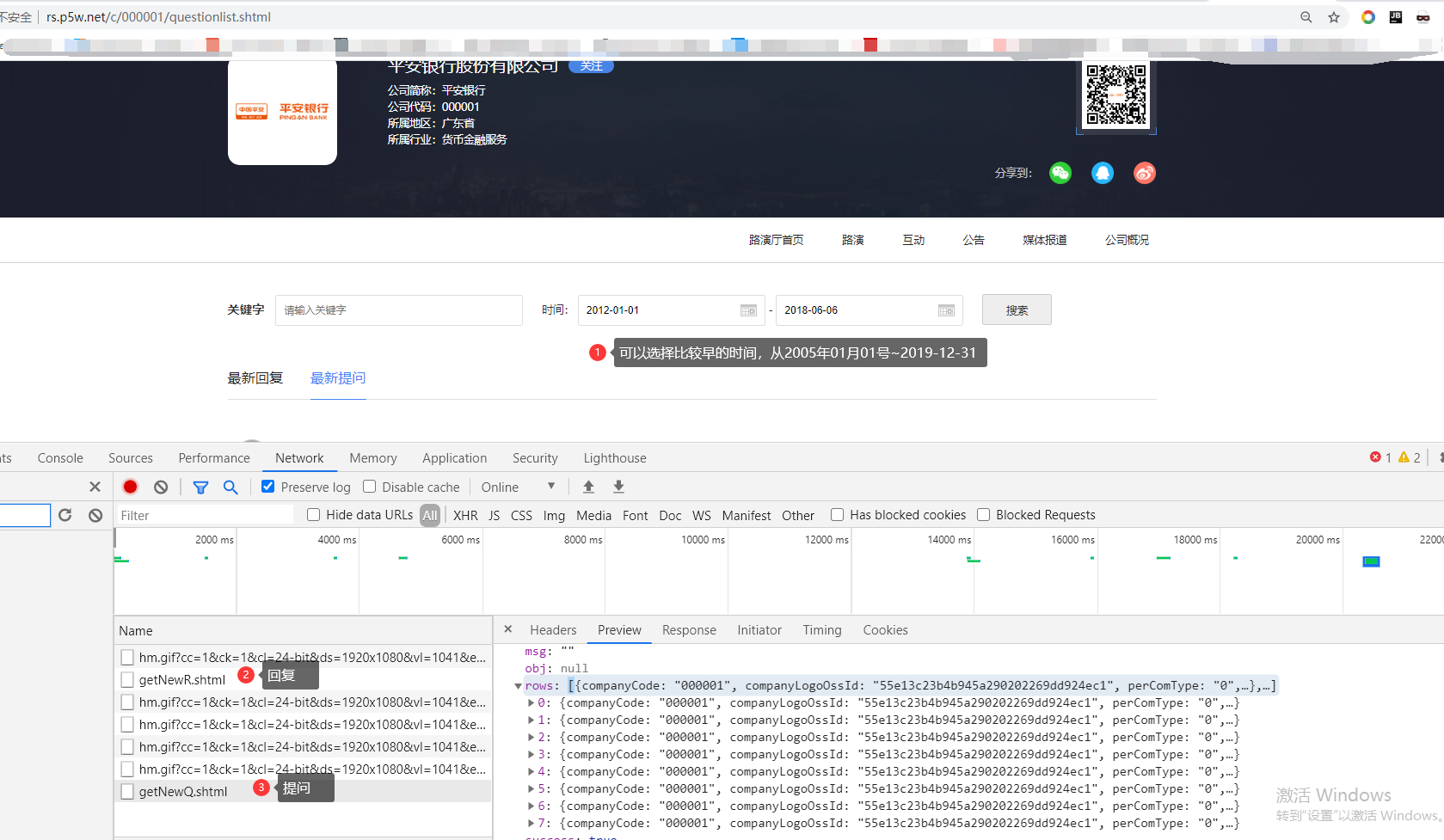
http://rs.p5w.net/question/ 投资者互动平台



抓取的时间为 从05年抓到2019年12月31号

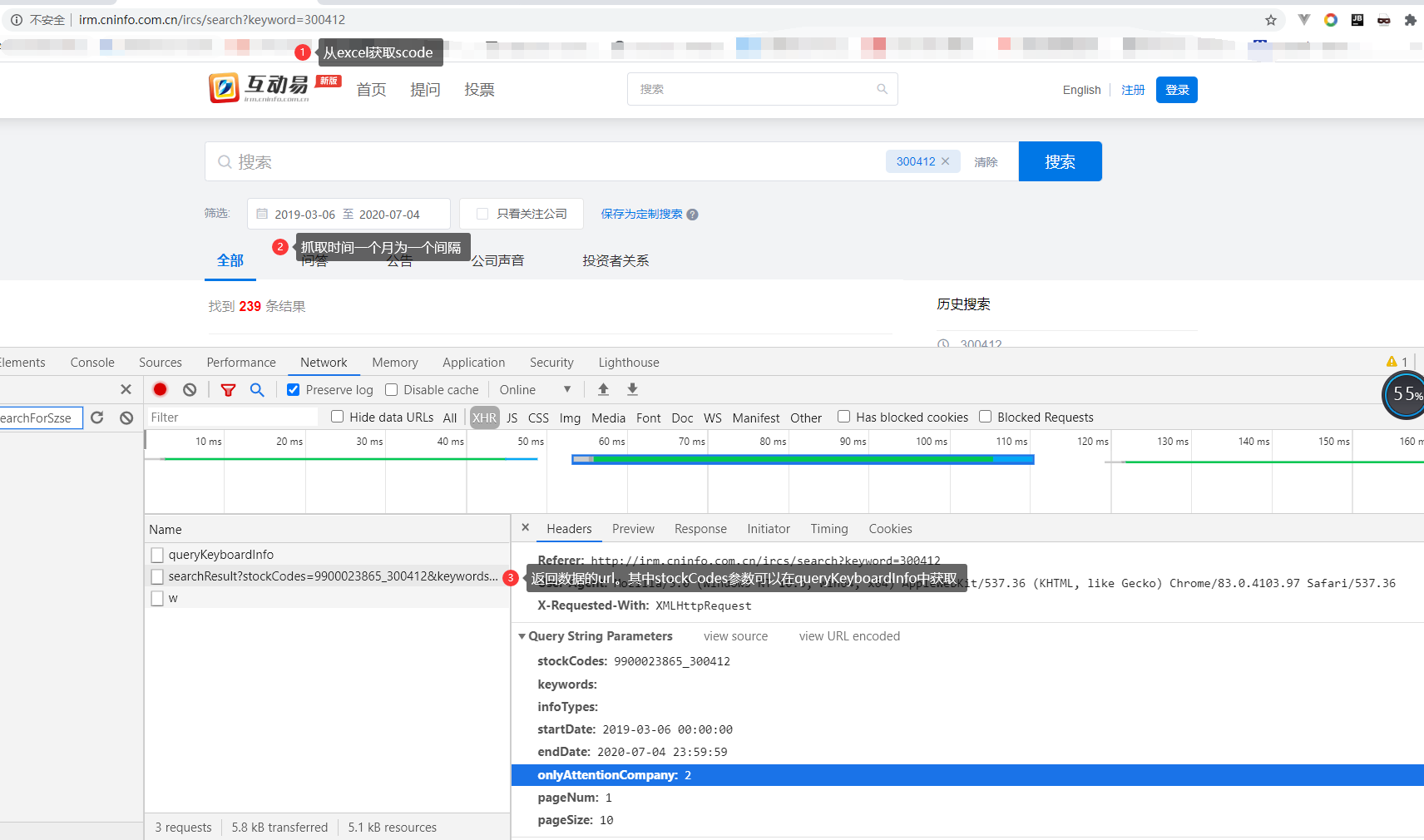
通过一番寻找，以000001为例：

<http://rs.p5w.net/c/000001/questionlist.shtml>



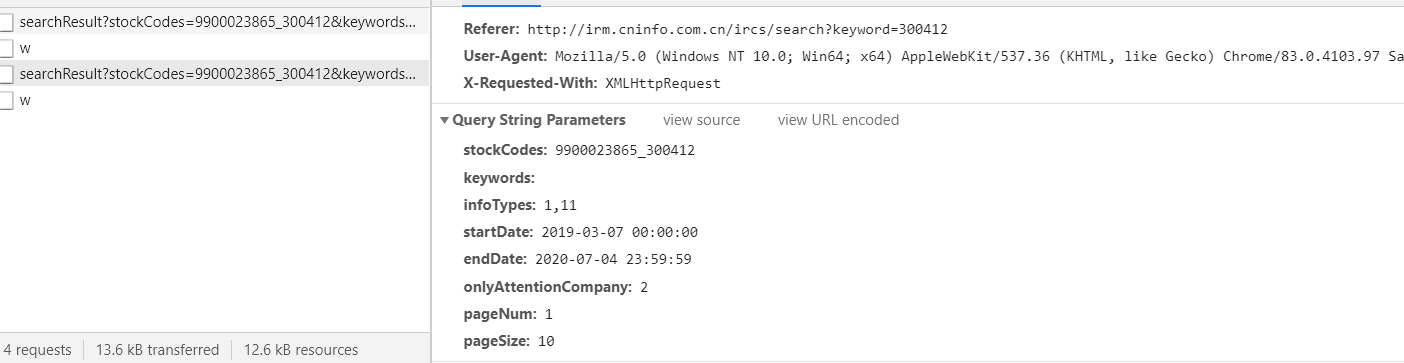
http://irm.cninfo.com.cn/szse/index.html 互动易（深交所）

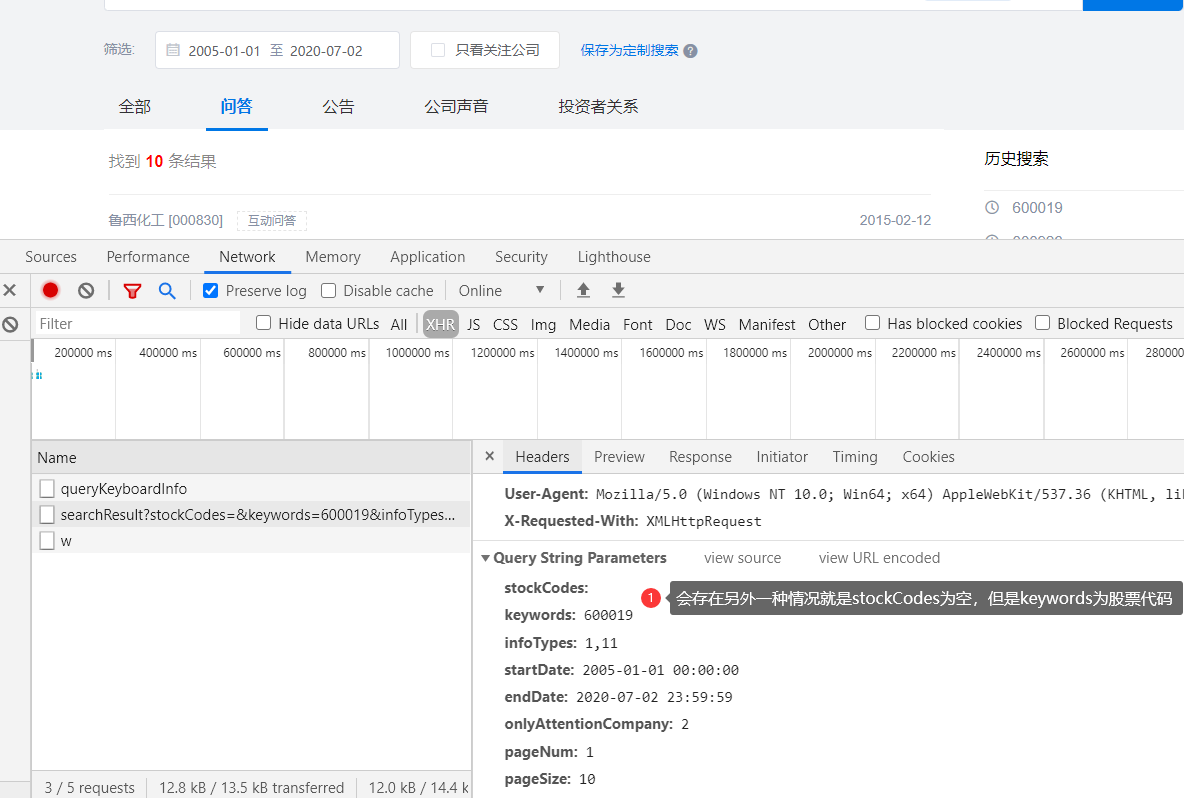
抓取的时间为 从05年抓到2019年12月31号



首先先抓取stockCodes参数，它是http://irm.cninfo.com.cn/ircs/index/queryKeyboardInfo

secid+stockCode的组合





将secid+stockCode的组合的情况写入到stockCodesexcel表格中

http://sns.sseinfo.com/qa.do e互动（上交所）

抓取的时间为 从05年抓到2019年12月31号，通过代码测试，数据从13年开始有数据

首先我们先获取所有上市公司的名称<http://sns.sseinfo.com/list/company.do>

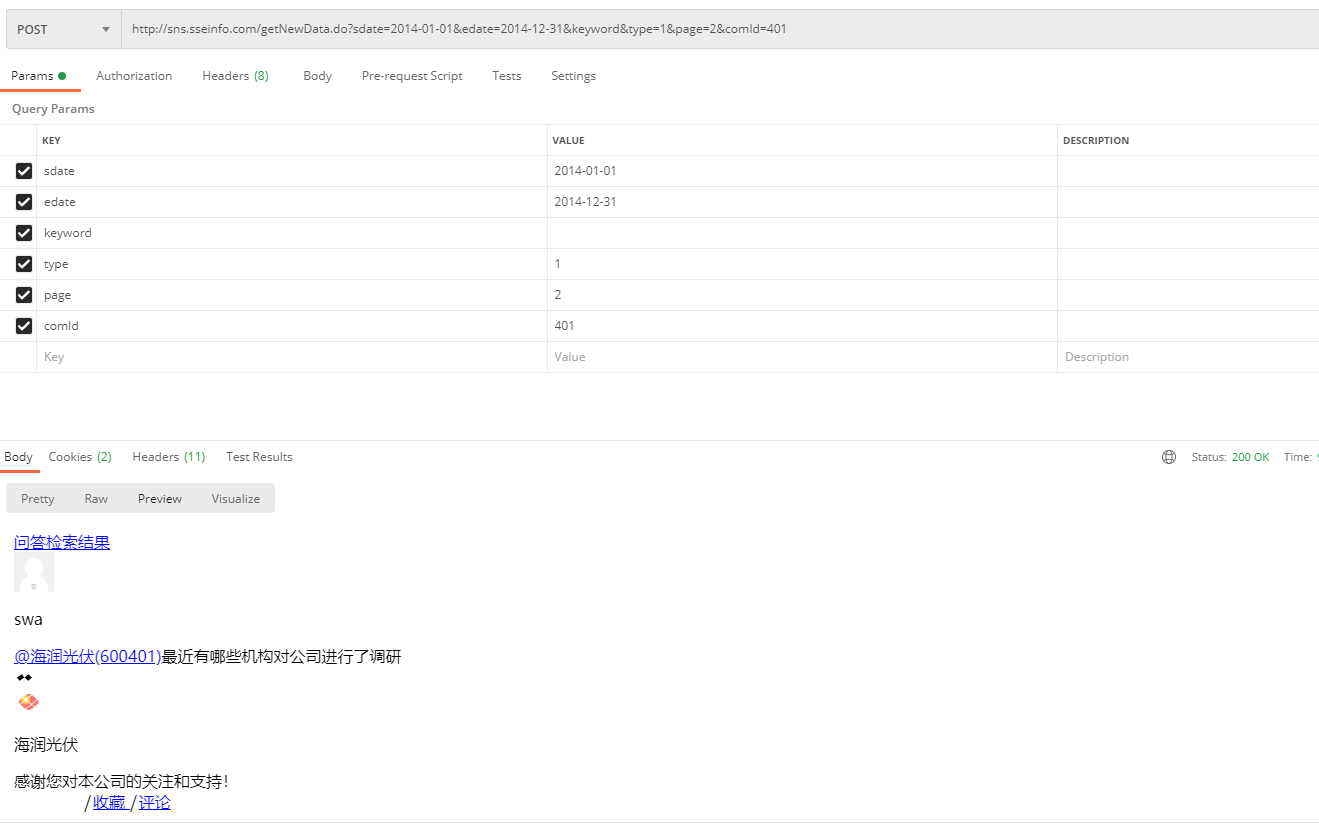
遇到的网站自身的问题：

第一：

正在获取第 -20- 家公司问答信息！共 -1688- 家公司！当前公司代码+uid：600401+401



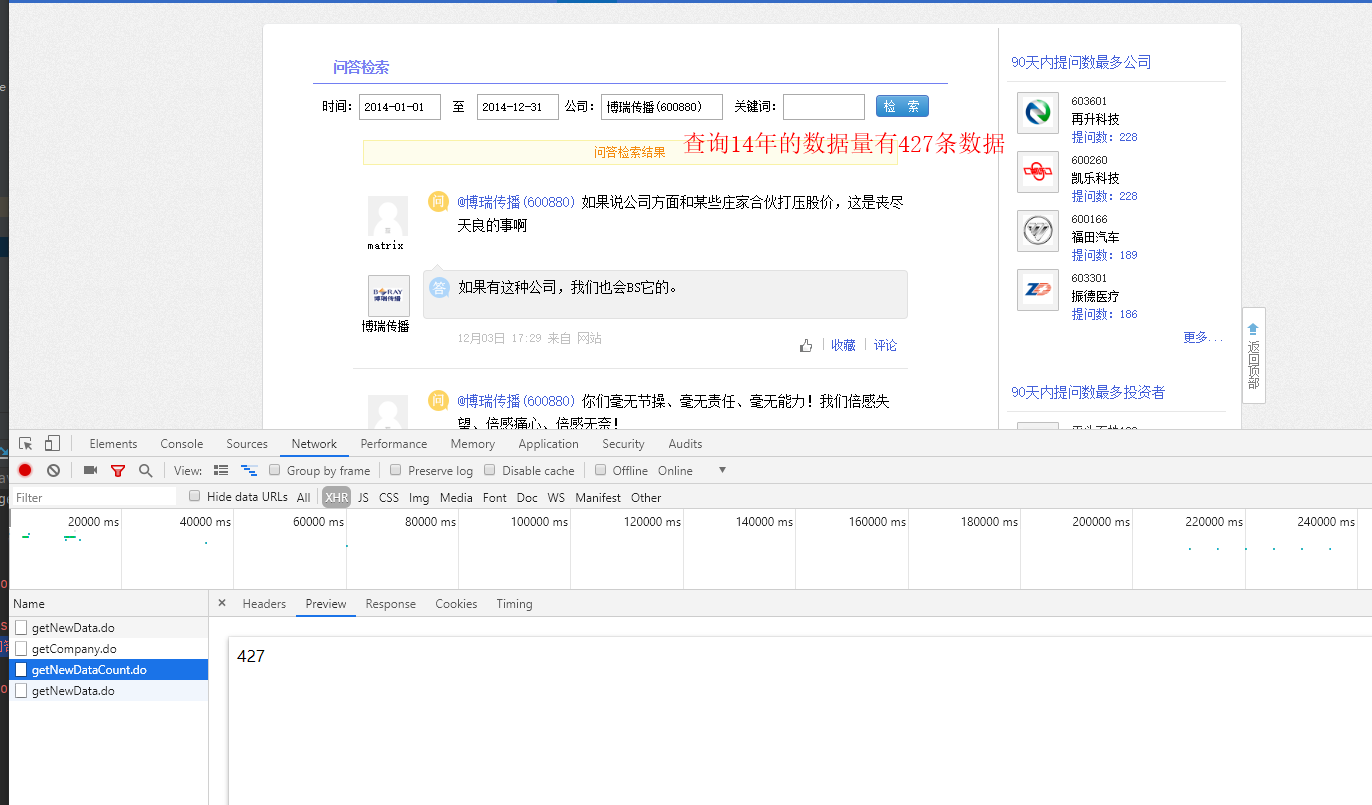
但是用postman能请求到数据



<http://sns.sseinfo.com/company.do?uid=401> 显示数据有1600多条数据，但是，但是跑出了2165。该公司已退市.用postman请求了，还真有217页数据

第二：

正在获取第 -21- 家公司问答信息！共 -1688- 家公司！当前公司代码+uid：600880+880





翻到最后一页发现一共有3页没有数据，导致这一年下载的数据只有397条数据

这种情况不止600800存在这种情况

还有600166股票代码等等



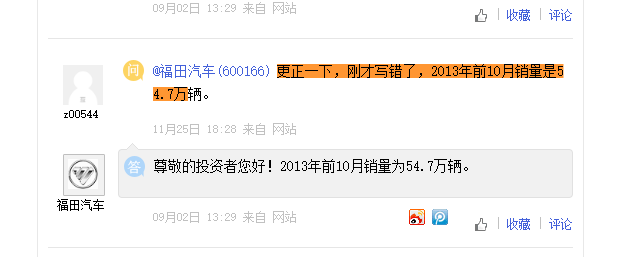


新的情况：



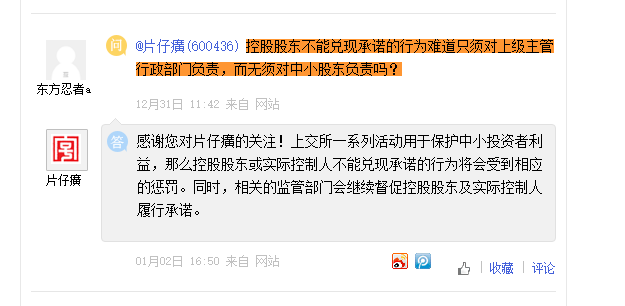
提问时间是11月份，回复时间是9月份

是将此数据删除，还是将提问时间减去一年？

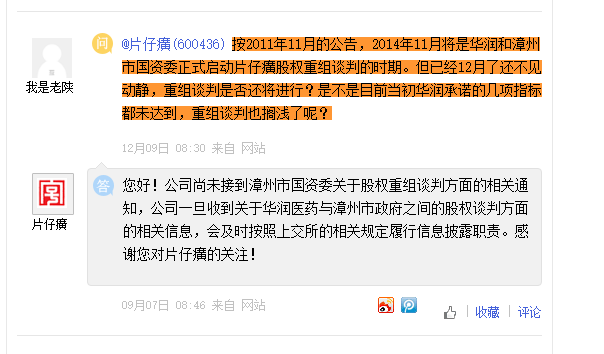


再根据上面的图，会发现一个问题，如果提问时间大于回复时间，就将提问时间的年份减去一年

600436也存在这个

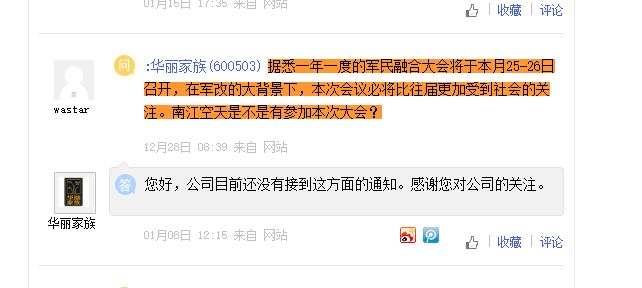






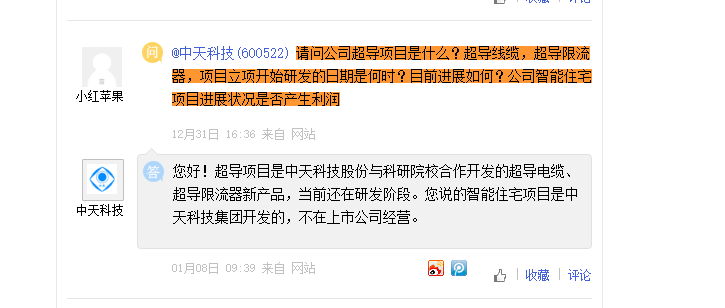


600503





600522



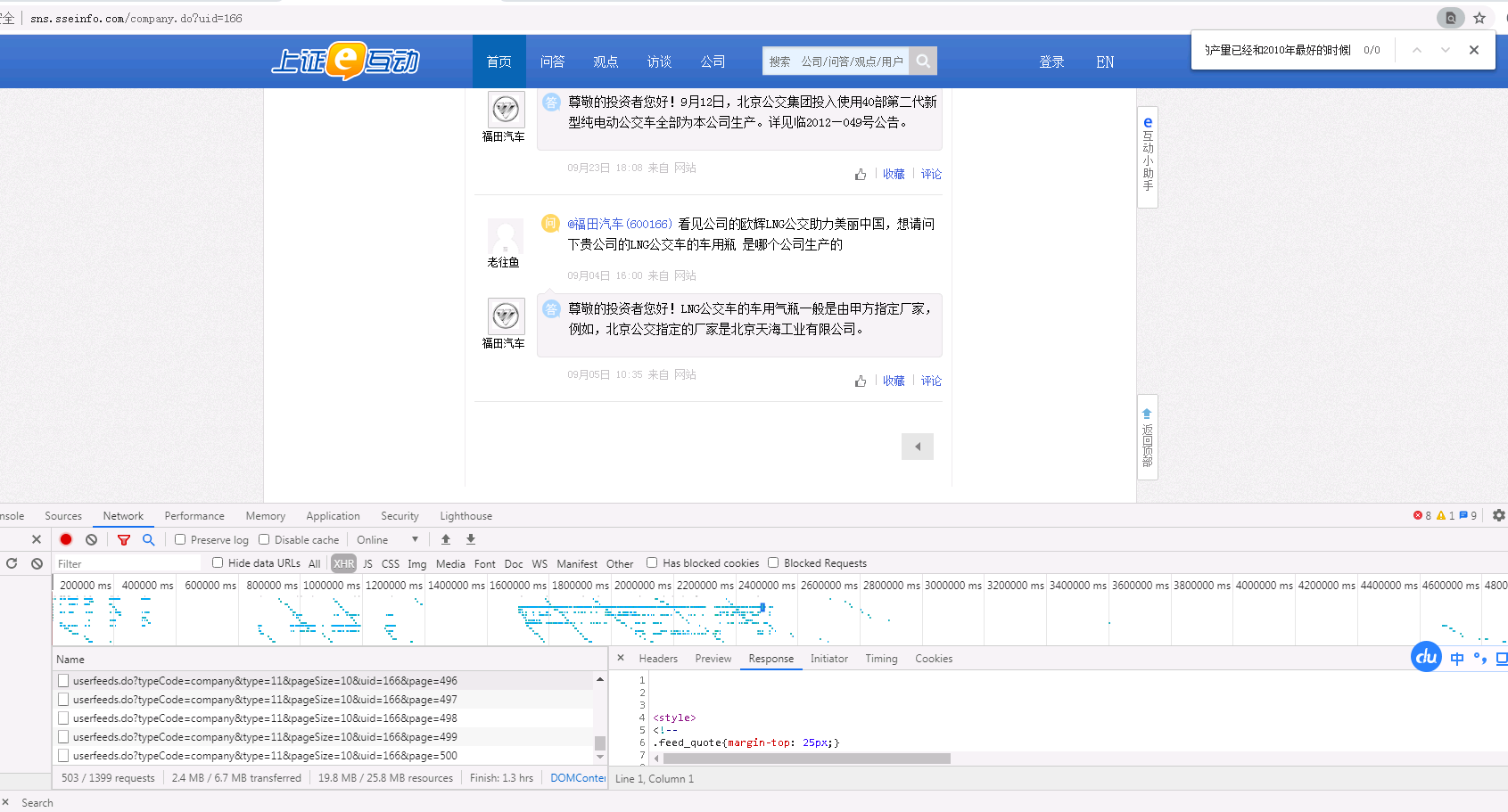


新问题：以600166为例



根据提问的问题去下面的链接查找

<http://sns.sseinfo.com/company.do?=uid=166>



未发现所提的问题，所以提问时间就写成2014年

新情况

