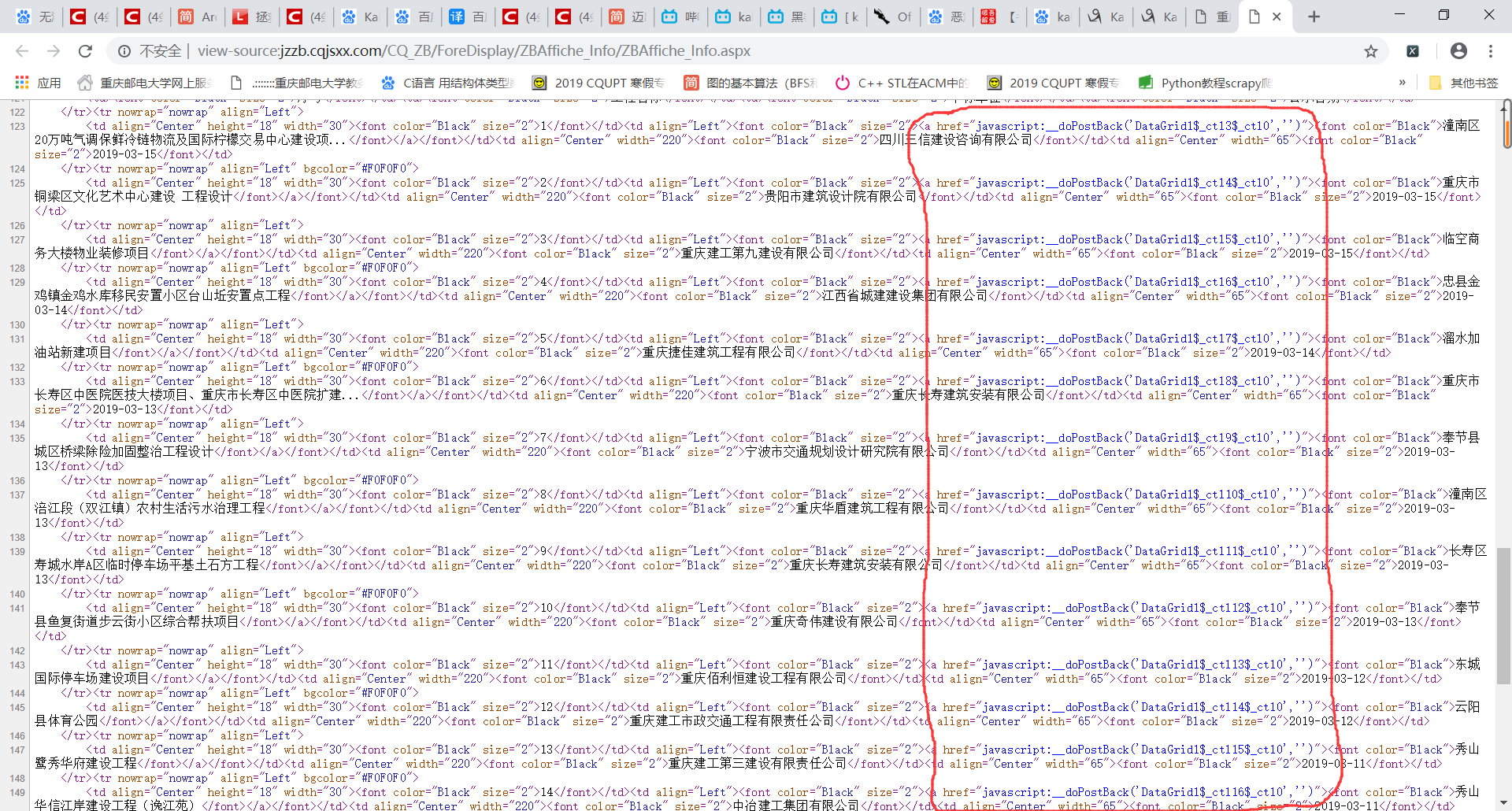
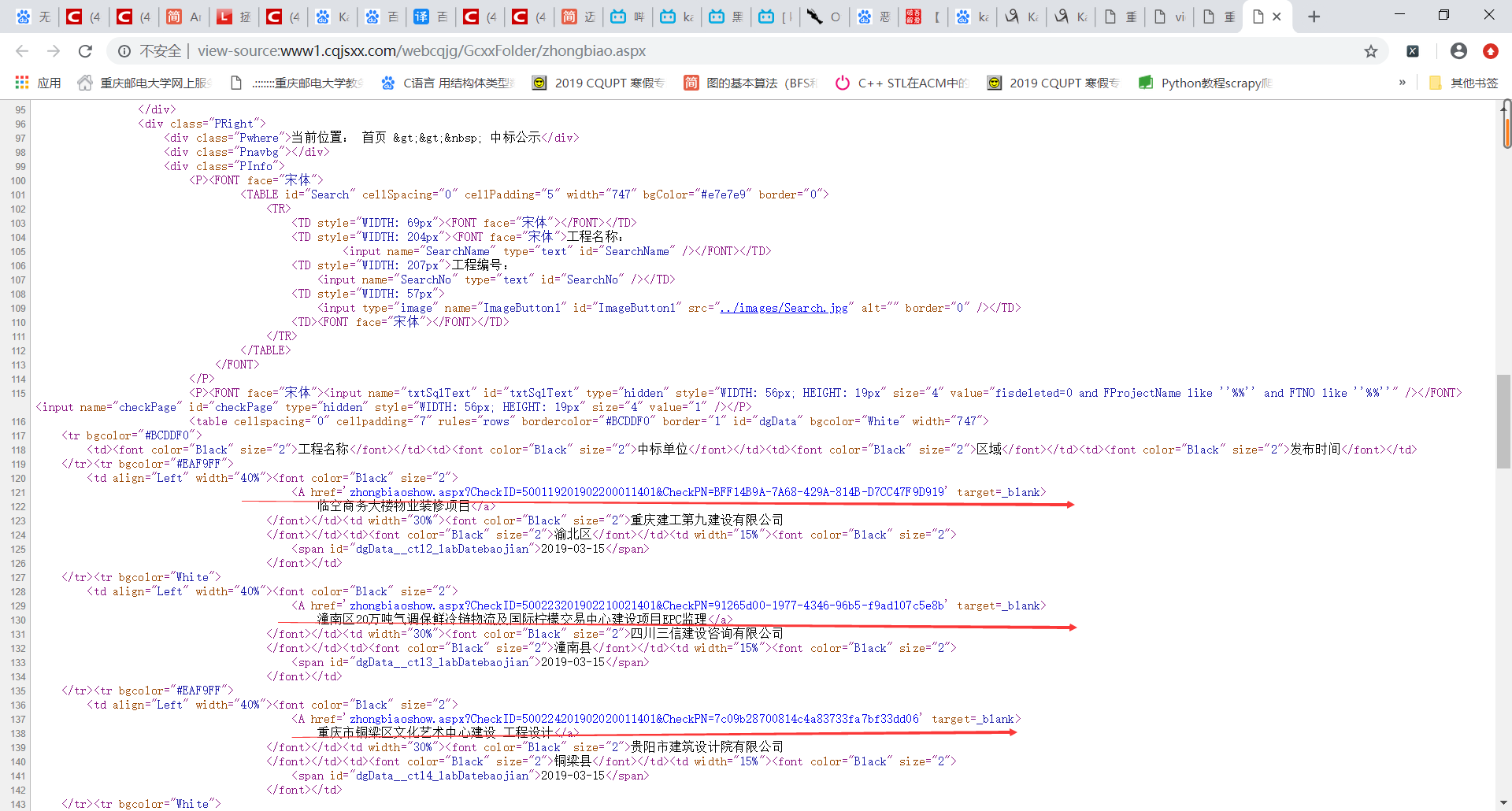
实验室爬虫说明文档三

本次爬取的是老地址，尝试以前的方法的post没有成功，所以这次直接使用selenium，简单粗暴，不过那个网站实在是太。。。不稳定了（有点pie），所以爬到470页时网就崩了emmm





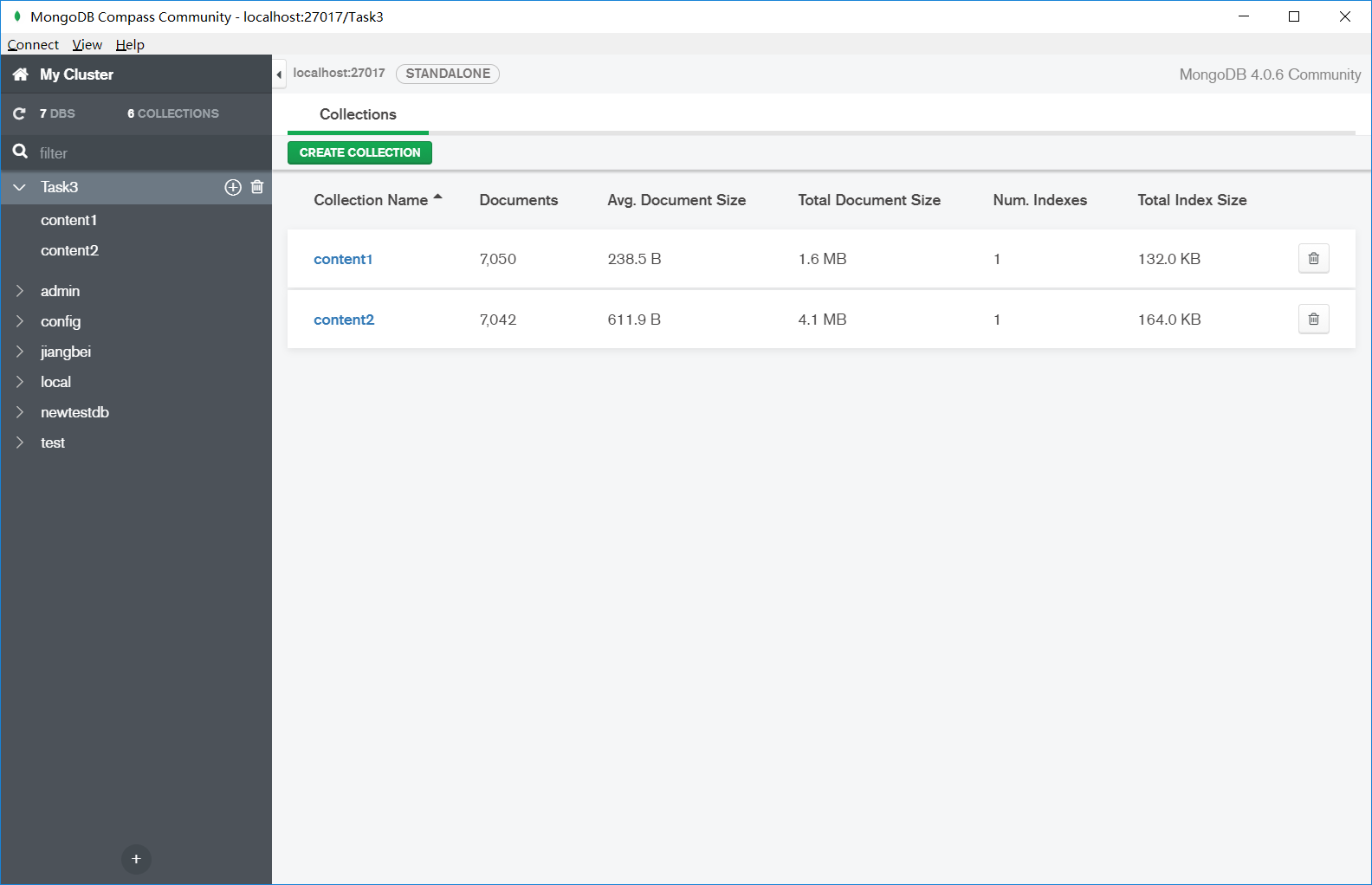
每个链接不是网址，而是要post，跟后面发的新网址不一样，那个是直接给出了网页链接

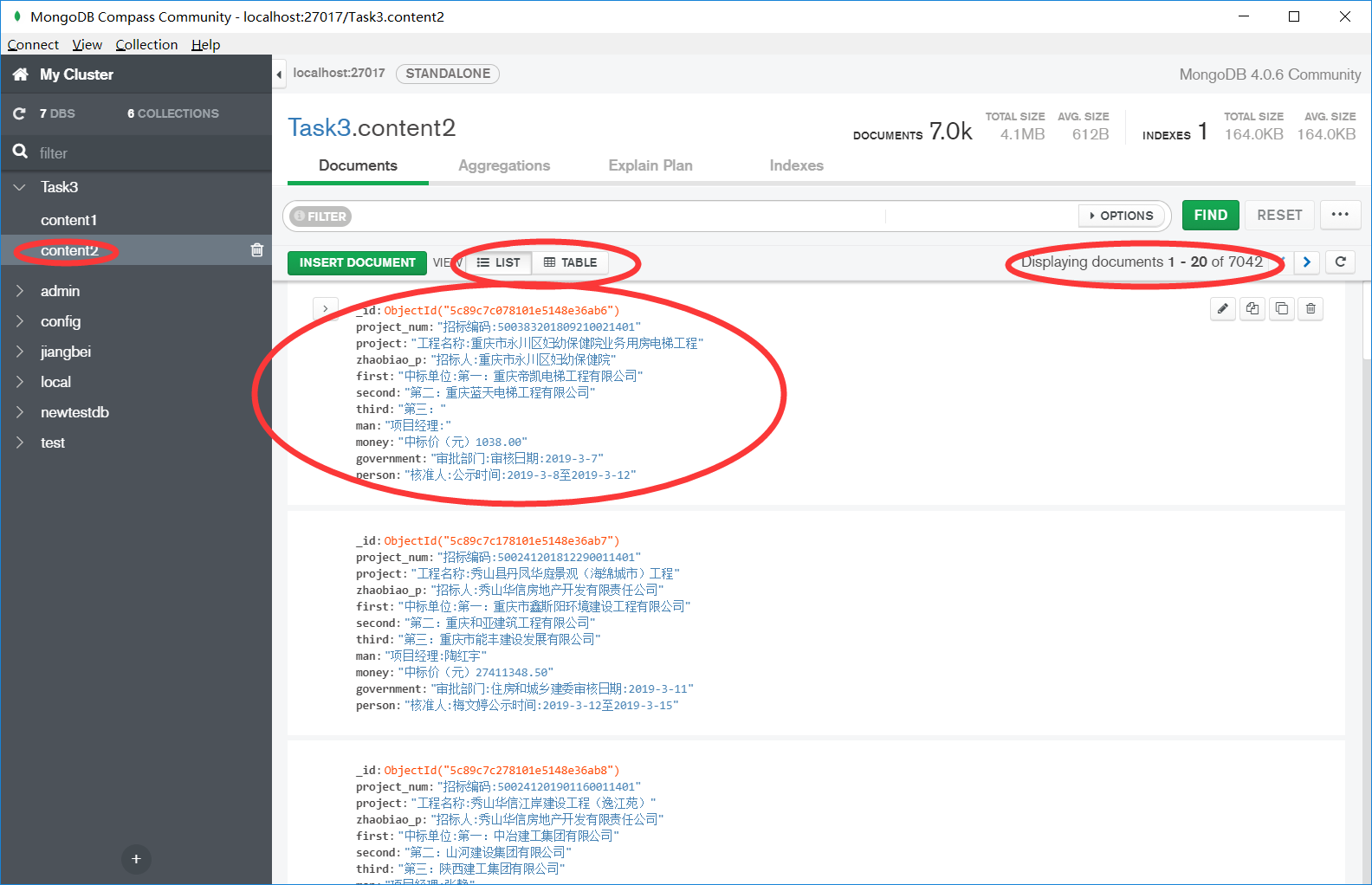


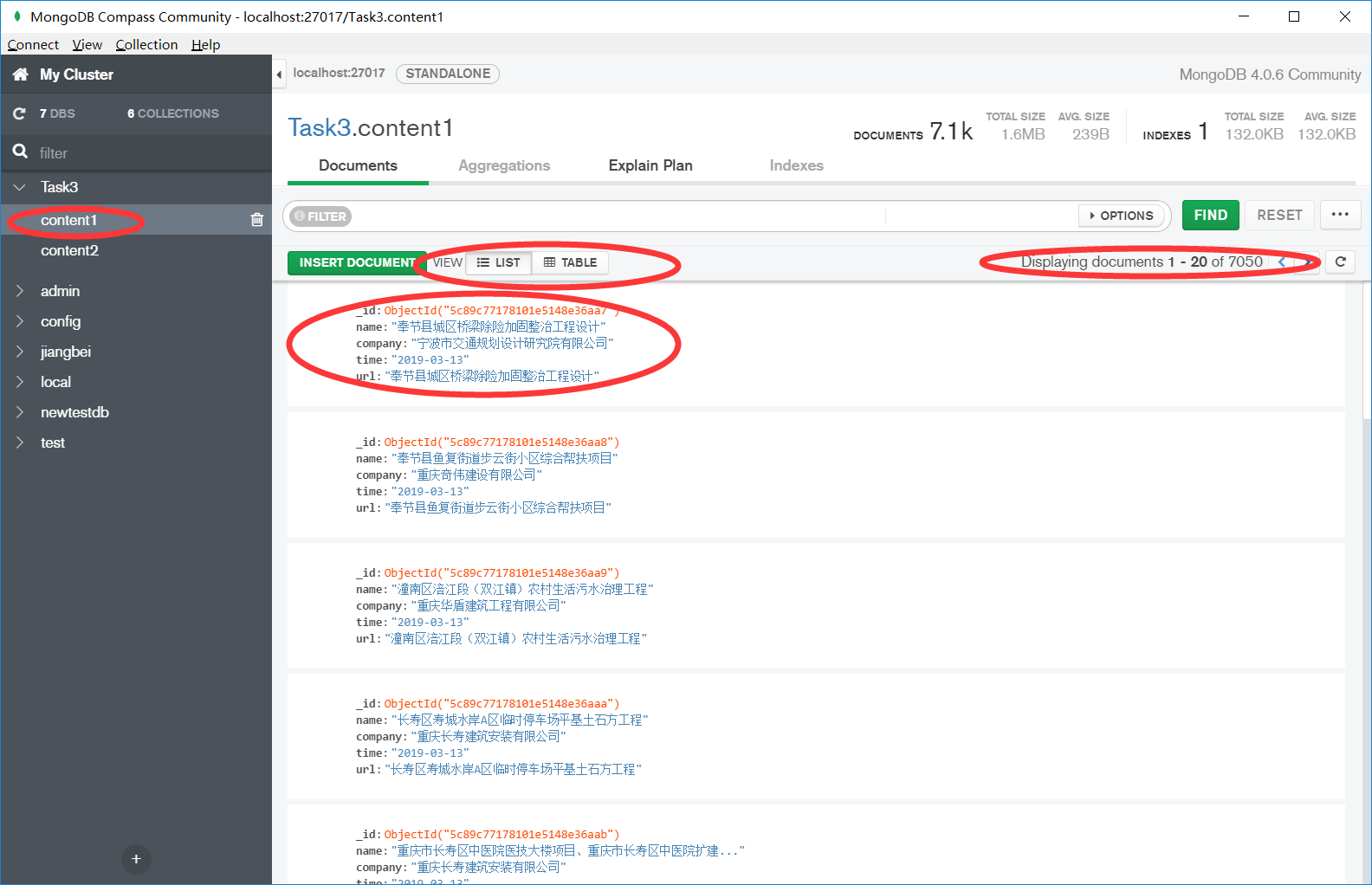
然后为了尝试新的，选择了原来的网址，之后23333

迫不得已才用selenium，但是真心慢啊，并再次吐槽这个网址后台是真的。。。。。

数据保存模式：mongodb数据库：







尝试从断点处重新爬取数据，发现直接selenium输入470页后回车（虽然网页刷新以为跳转，但还是在第一页）/点击确认（没有反应），所以决定硬核翻页到470页后再进行爬取（详见task32）

亲测有效！！！！！这样数据就全了（600页之前含600页），因为600页之后，后台就崩了2333333