

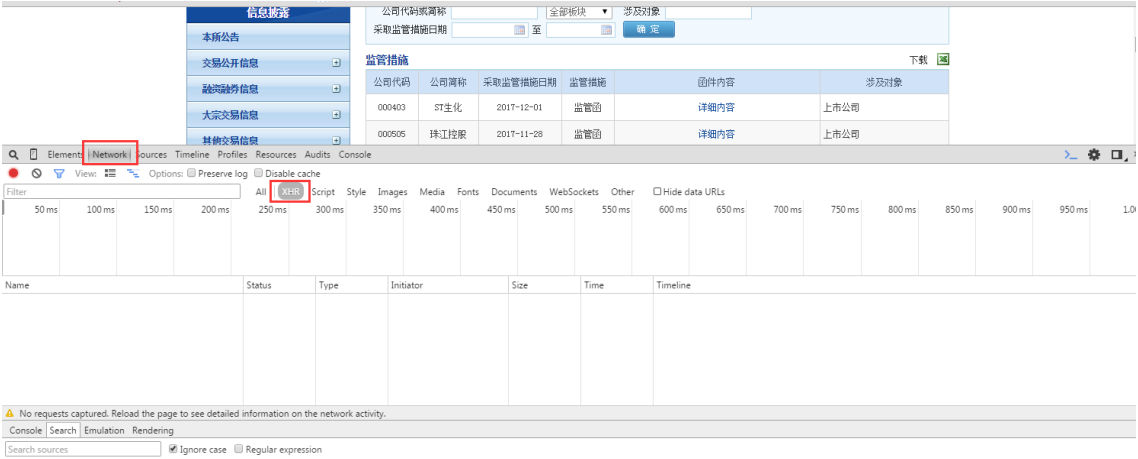
Python 爬虫，遇到网页需要翻页的情形，解决办法：

<http://www.szse.cn/main/disclosure/jgxxgk/jgcs/>

方法：

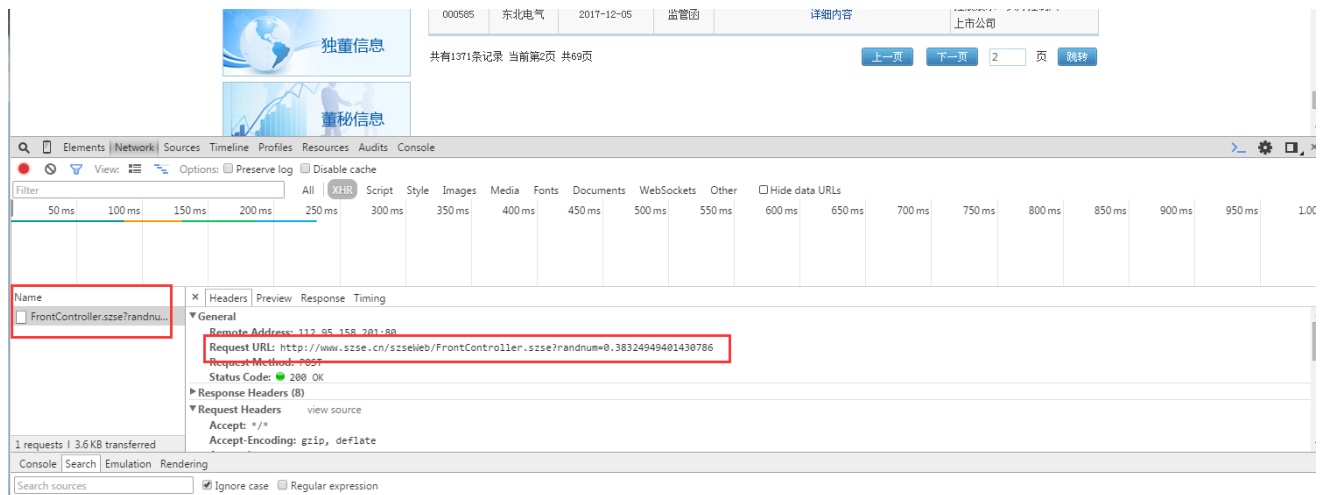
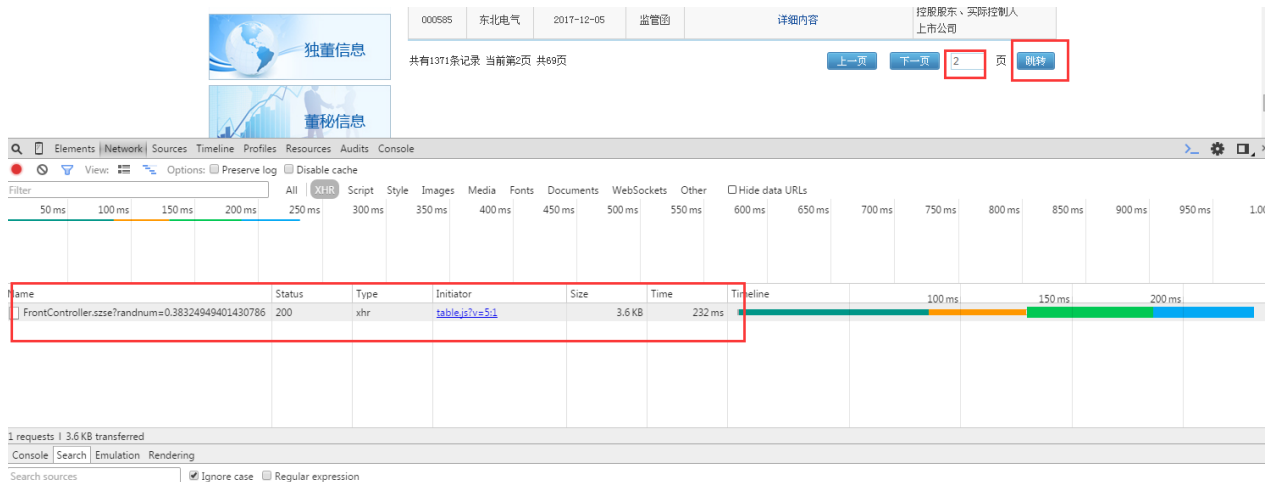
1.需要分析网页在翻页是提交的 post 表单数据

a. 分析网页提交的 post 数据时候，可以用 chrome 浏览器，按 F12



点击 NetWork、XHR

b. 寻找翻页 POST 表单

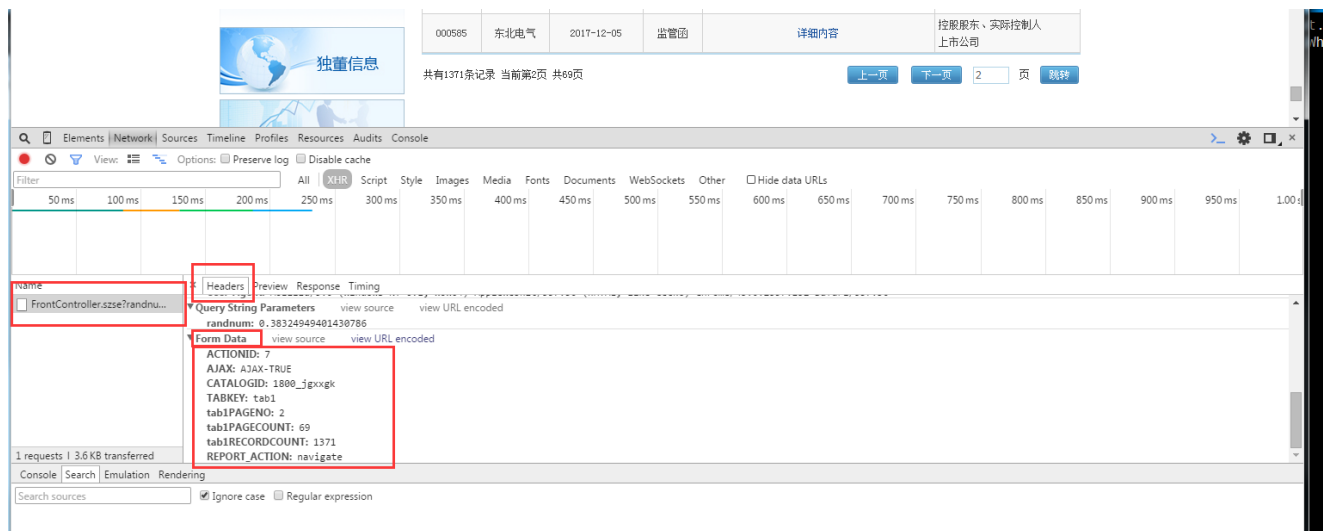


在网页翻页处填上 2，然后点击翻页，此时能看见 POST 请求相关信息。

请求网址：

<http://www.szse.cn/szseWeb/FrontController.sze?randnum=0.38324949401430786>

往下翻看 Headers 能看到 POST 表单信息，Form Data



2.分析服务器的返回的数据

如想下载 PDF , 其对应 URL 为 : <http://www.szse.cn>

后缀 : [UpFiles/jgsy/gkxx_jgsy_00062043644.PDF](http://www.szse.cn/UpFiles/jgsy/gkxx_jgsy_00062043644.PDF)

eg:

http://www.szse.cn/UpFiles/jgsy/gkxx_jgsy_00062043644.PDF

3.使用 urllib2 模块或者 requests 模块发送 post 表单 , 并接受服务器的响应。