裁判文书网，由于爬取人数较多，运营方为了限制爬虫，曾与6月和8月进行了两次大规模改版，并对app进行了停运处理。

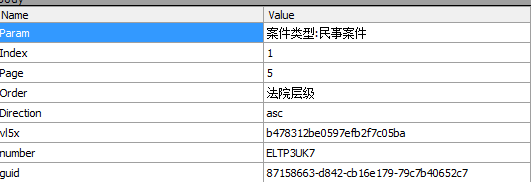
如果采用直接爬取网页或者采用selenium的方式，会有如下问题：

1. 在翻两三页过后，网页会跳出验证码，在正确输入验证码后重新跳回首页；
2. 每页显示条数，默认为5条，最多20条，如果想更改显示条数，则大概率跳出验证码，在正确输入验证码的情况下，跳回网站首页。
3. 最多只能翻25页。

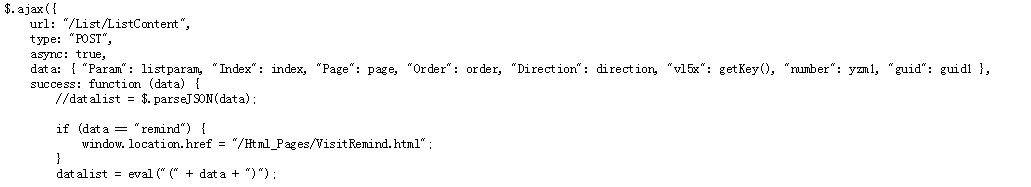
经过抓包分析，发现可以通过构造post请求的方式获取到数据。

路径：/List/ListContent

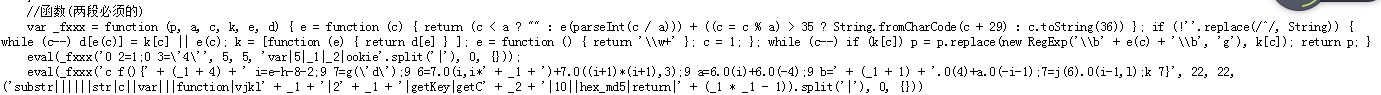
参数：



在及Assets/js8/Lawyee.CPWSW.List.js中可以找到这条函数，并且通过注释可以知道通过post请求可以最多请求100页数据。



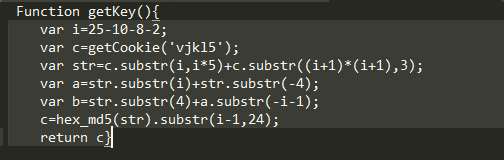
number参数、guid参数通过阅读这段代码，可以发现很容易构造出来。唯独获取vl5x这个参数的getKey()函数不在这段Js中。既然网页会调用这条post，那么回到网页的源代码中去寻找，会找到这样一段代码：



Google可以发现，function (p, a, c, k, e, d)是一段加密JS，将第一个eval中的\_fxxx替换为其定义，利用http://heyongchao.com/eval-function-packed-series-of-encryption-and-decryption-javascript-program.do这个工具进行解码，可以得到解码结果：

var \_1=5;var \_2='ookie'

将第二个eval中的\_1和\_2替换为真实值，并再度进行解码，可以得到如下代码：



从cookie中读取vjkl5值，经过计算就可以得到vl5x的值了。