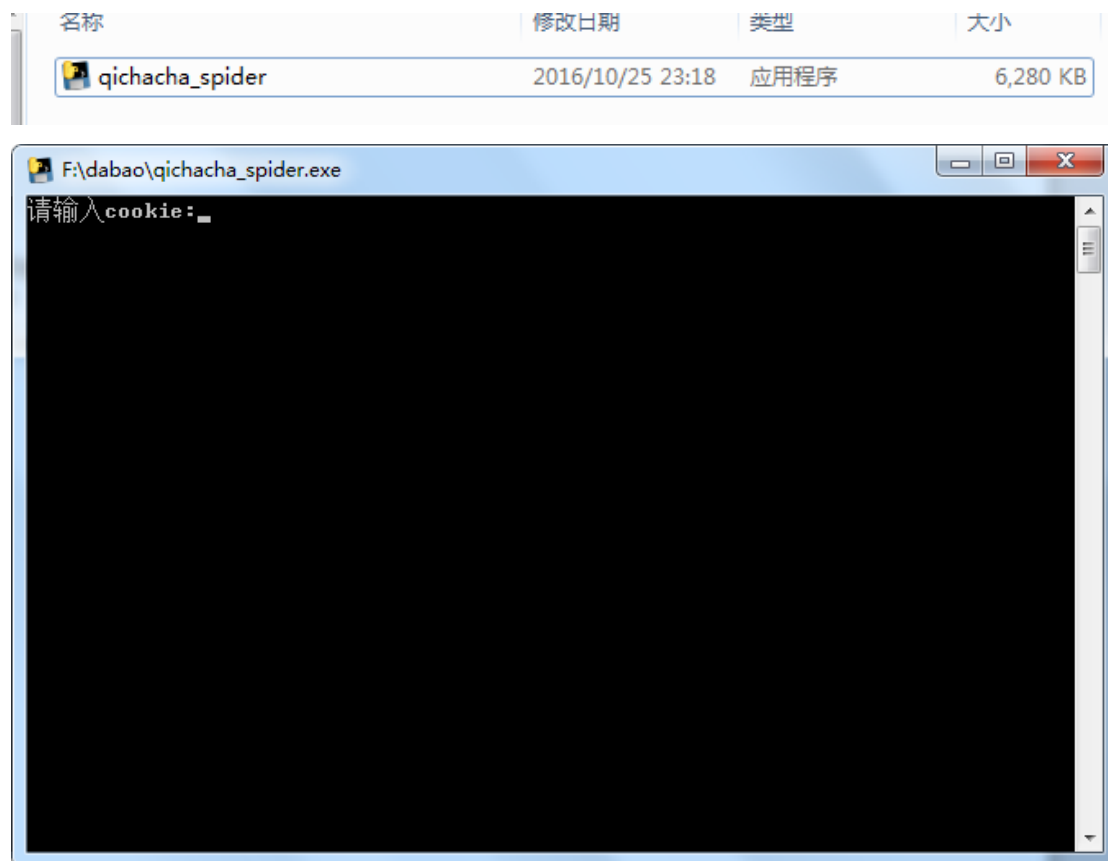


企查查爬虫软件使用指南

本爬虫软件为一个半自动化的数据采集软件，需要用户模拟登录“企查查”网站后，复制到 cookie 值才能进行程序后续的自动化采集工作。而复制 cookie 时的界面、快捷键由于浏览器不同会有所差异，下面截图中的浏览器均是 Chrome（谷歌浏览器），也推荐使用 Chrome，安装包已提供。

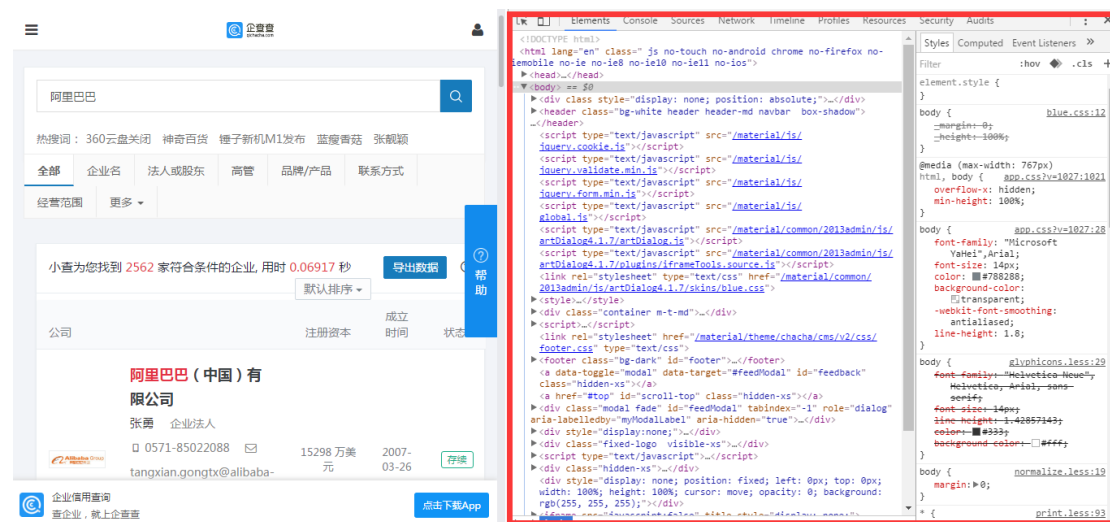
1、打开软件，进入如下黑框界面，这时要求输入 cookie:



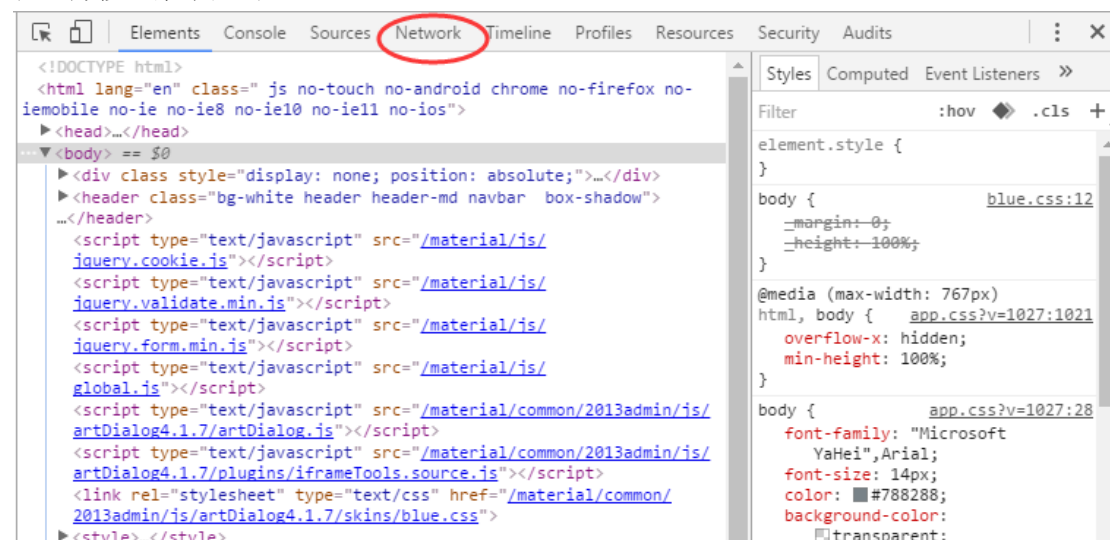
2、获取 cookie: 手动进入企查查官网，用自己的帐号登录系统（最好不要不登录），然后在搜索框中随便输入一个公司名称



这时按 **F12** 打开浏览器页面管理界面，如下图（红色边框区域就是管理界面）：



在上方按钮条中点击“network”：



然后如下图，注意保证红色圆圈内的 All 是灰色（选中）的

Name	Status	Type	Initiator	Size	Time	Timeline - Start Time
17581360	200	xhr	libs.min.js?v...	488 B	35 ms	
17581360	200	xhr	Other	491 B	27 ms	
GetSetting	200	xhr	libs.min.js?v...	386 B	199 ms	

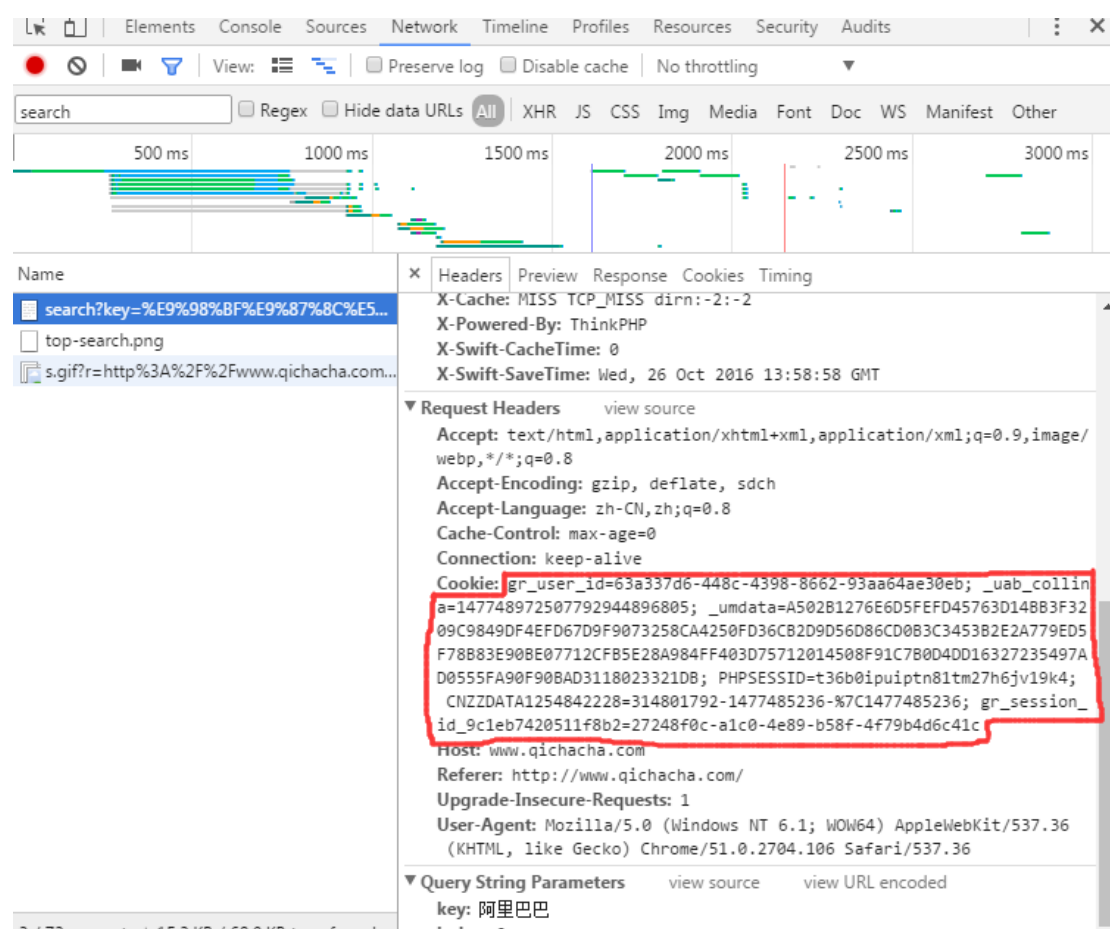
这时点击“刷新”，会发现下方多了很多条数据：

Name	Status	Type	Initiator	Size	Time	Timeline - Start Time
search?key=%E9%98%B...	200	document	Other	14.6 KB	768 ms	
bootstrap.css	304	stylesheet	search?key=...	567 B	17 ms	
icon.css	304	stylesheet	search?key=...	569 B	19 ms	
font-awesome.min.css	304	stylesheet	search?key=...	554 B	16 ms	
app.css?v=1026	200	stylesheet	search?key=...	20.8 KB	507 ms	
jquery.min.js	304	script	search?key=...	567 B	509 ms	
siteconfig.js	304	script	search?key=...	518 B	21 ms	
jquery.scrollTo.min.js	304	script	search?key=...	564 B	402 ms	
font.css	304	stylesheet	search?key=...	526 B	20 ms	
toastr.css	304	stylesheet	search?key=...	552 B	21 ms	
app.js	304	script	search?key=...	582 B	461 ms	
bootstrap.js	304	script	search?key=...	582 B	28 ms	
jquery.slimscroll.min.js	304	script	search?key=...	567 B	462 ms	
app.plugin.js	304	script	search?key=...	582 B	462 ms	
toastr.js	304	script	search?key=...	581 B	401 ms	
custom.js?time=20160...	304	script	search?key=...	582 B	508 ms	
logo4.png	304	png	search?key=...	506 B	494 ms	
0_s.jpg	304	jpeg	search?key=...	346 B	97 ms	
jquery.cookie.js	304	script	search?key=...	581 B	14 ms	

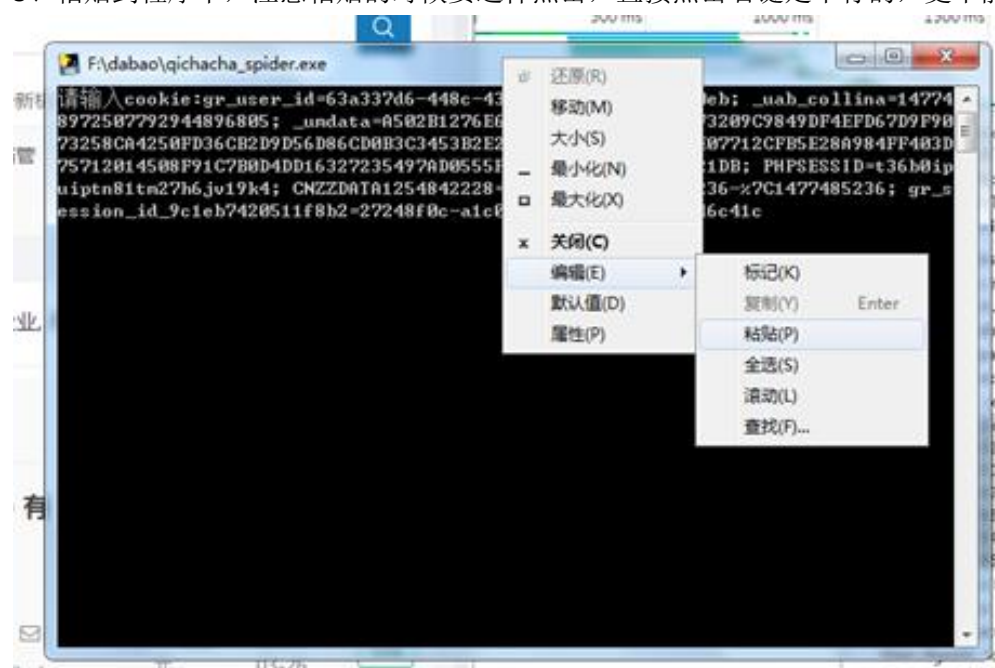
我们的目标就是要找到这个 search 开头的条目，点击它。如果没有找到，可能是没有移动滑块，可以使用上图右侧滑块向上滑动，一般这个 search 就是在第一个，如果不在第一个，可以用上面搜索框快速搜到：

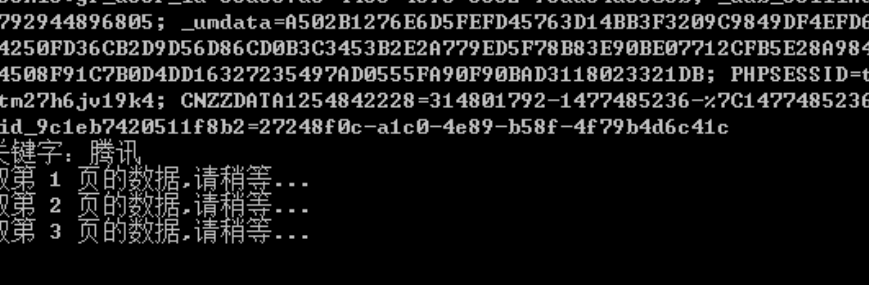
Name	Status	Type	Initiator	Size	Time
search?key=%E9%98%B...	200	document	Other	14.6 KB	768
top-search.png	304	png	search?key=...	503 B	5
s.gif?r=http%3A%2F%2...	200	gif	Other	212 B	224

点开这个 search 以后就有了下图，这时我们继续向下移动滑块，找到以 Cookie: 开头的这项，把它后面的数据复制，如下图红色框中部分：

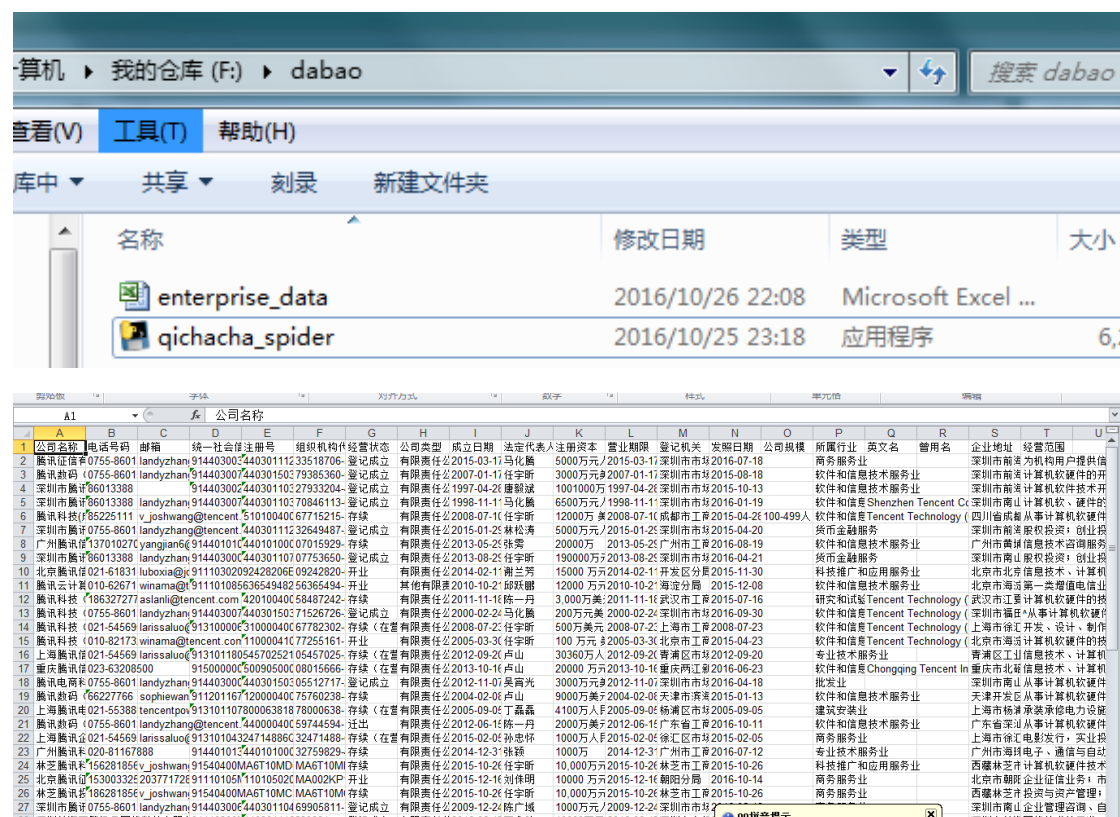


3、粘贴到程序中，注意粘贴的时候要这样点击，直接点击右键是不行的，更不能用 `ctrl+V`



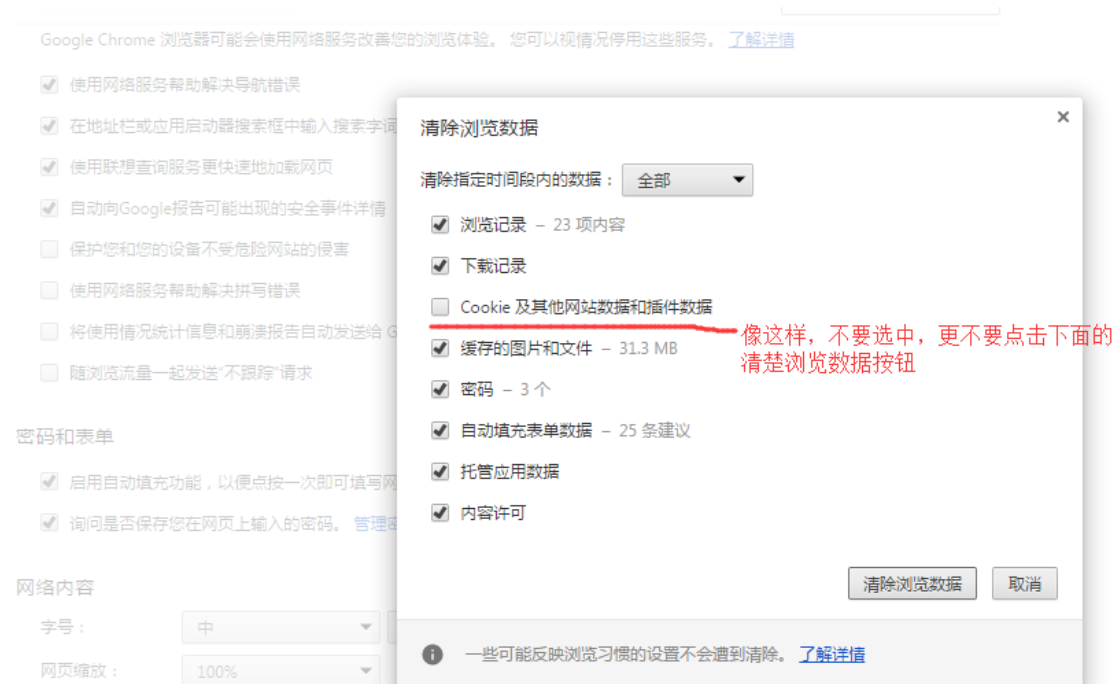


```
F:\dabao\qichacha_spider.exe
请输入cookie:gr_user_id=63a337d6-448c-4398-8662-93aa64ae30eb; _uab_collina=14774
8972507792944896805; _umdata=A502B1276E6D5FEFD45763D14BB3F3209C9849DF4EFD67D9F90
73258CA4250FD36CB2D9D56D86CD0B3C3453B2E2A779ED5F78B83E90BE07712CFB5E28A984FF403D
75712014508F91C7B0D4DD16327235497AD0555FA90F90BAD3118023321DB; PHPSESSID=t36b0ip
uiptn81tm27h6jv19k4; CNZZDATA1254842228=314801792-1477485236-%7C1477485236; gr_s
ession_id_9c1eb7420511f8b2=27248f0c-a1c0-4e89-b58f-4f79b4d6c41c
请输入关键字: 腾讯
正在爬取第 1 页的数据,请稍等...
正在爬取第 2 页的数据,请稍等...
正在爬取第 3 页的数据,请稍等...
```



注意事项:

- 0、模拟登录复制 **cookie** 的操作不需要用一次软件复制一次，这个 **cookie** 的有效期很长，复制一次，存在一个 **txt** 文件中，下次打开软件，继续复制它就可以了，应该可以用好几天。然后保证不要对 **Chrome** 浏览器进行清空，特别是不要选清除 **cookie** 选项，否则 **cookie** 立即失效。



- 1、如果第一次没有 **enterprise_data.xls** 这个 excel 文件程序会帮忙创建，不需要人工创建
- 2、数据是**追加**到这个 excel 表格的，即关了程序后，第二次再打开程序，**不会覆盖**之前采集好的数据，而是紧随其后继续写入
- 3、**请勿在爬取过程中打开 excel 表格**，否则会中断，而下次再爬这个搜索值的时候，数据就会重复了，如果不在乎那无所谓
- 4、支持连续输入，即比如爬完“阿里巴巴”这个搜索值后，程序提示“爬虫结束，请输入：”，这时又可以输入下一个要爬的搜索值“百度”
- 5、**一般不要光输入地名**，输入也没事，程序会继续提示你输入下一个有效搜索值，比如“杭州”“北京”，否则会查不到，这个网站是按照公司名称来搜索的，所以最好这样“杭州 阿里巴巴”、“北京 百度”，以空格分开