

chrome浏览器抓包说明

学习目标

掌握chrome在爬虫中的使用

1.新建无痕窗口

#1.使用隐身窗口，首次打开网站，不会带上cookie，能够观察页面的获取情况，包括对方服务器如何设置cookie在本地

#2.无痕模式下，如果一个地址能够访问，说明该地址的代码请求头中，只需要一个user-agent即可，可观察headers中需要的键值对



您已进入无痕模式

现在，您便可进行私密浏览了。与您共用此设备的其他用户将不会看到您的活动，但您的下载内容、书签和阅读清单项将会保存在此设备上。 [了解详情](#)

Chrome 不会保存以下信息：

- 您的浏览记录
- Cookie 和网站数据
- 在表单中填写的信息

以下各方可能仍会看到您的活动：

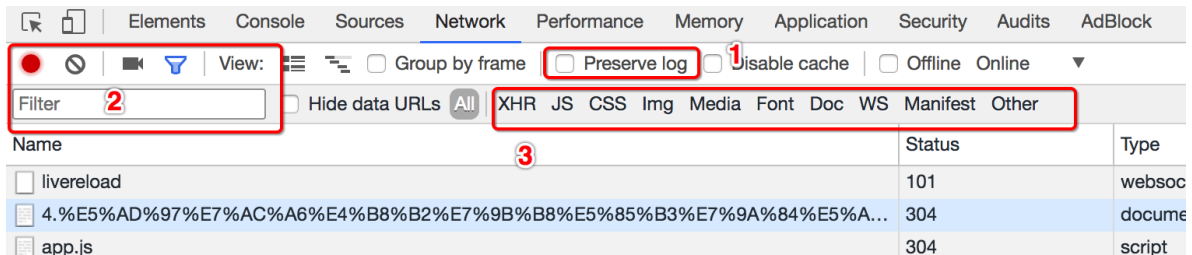
- 您访问的网站
- 您的雇主或您所在的学校
- 您的互联网服务提供商

阻止第三方 Cookie

开启此设置后，网站无法使用 Cookie 来跟踪您在网上的活动。某些网站上的功能可能会无法正常工作。



2.Chrome中network的更多功能



1.Preserve log

log 日志

保留日志：保留之前的请求

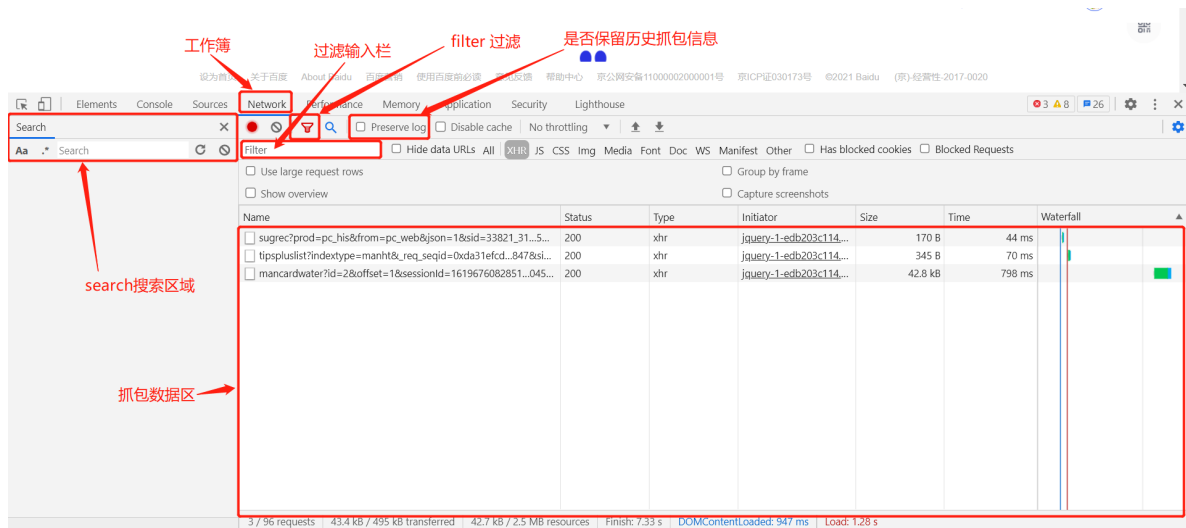
默认情况下，页面发生跳转之后，之前的请求url地址等信息都会消失，勾选**perserve log**后之前的请求都会被保留

View: <input type="checkbox"/> Group by frame <input checked="" type="checkbox"/> Preserve log <input type="checkbox"/> Disable cache <input type="checkbox"/> Offline	
Filter <input type="text"/>	<input type="checkbox"/> Hide data URLs All XHR JS CSS Img Media Font Doc WS Manifest
Name	Status
<input type="checkbox"/> livereload	101
<input checked="" type="checkbox"/> 4.%E5%AD%97%E7%AC%A6%E4%B8%B2%E7%9B%B8%E5%85%B3%E7%9A%84%E5%A...	200
<input checked="" type="checkbox"/> style.css	200
<input checked="" type="checkbox"/> prism-solarizedlight.css	200
<input checked="" type="checkbox"/> expandable-chapters-small.css	200
<input checked="" type="checkbox"/> search.css	200
<input checked="" type="checkbox"/> website.css	200
<input checked="" type="checkbox"/> app.js	200
<input checked="" type="checkbox"/> expandable-chapters-small.js	200
<input checked="" type="checkbox"/> book-summary-scroll-position-saver.js	200
<input checked="" type="checkbox"/> lunr.min.js	200
<input checked="" type="checkbox"/> search.js	200
<input checked="" type="checkbox"/> buttons.js	200
<input checked="" type="checkbox"/> buttons.js	200
<input checked="" type="checkbox"/> plugin.js	200
<input type="checkbox"/> fontawesome-webfont.woff?v=4.1.0	http://loc
<input checked="" type="checkbox"/> livereload.js	200
<input type="checkbox"/> search_index.json?_=1524492166695	200
<input type="checkbox"/> livereload	101
<input checked="" type="checkbox"/> 4.%E5%AD%97%E7%AC%A6%E4%B8%B2%E7%9B%B8%E5%85%B3%E7%9A%84%E5%A...	304
<input checked="" type="checkbox"/> style.css	304
<input checked="" type="checkbox"/> prism-solarizedlight.css	304

2.filter过滤

在url地址很多的时候，可以在filter中输入部分url地址，对所有的url地址起到一定的过滤效果，具体位置在上面第二幅图中的2的位置

3.观察特定种类请求



在上面第二幅图中的3的位置，有很多选项，默认是选择的all，即会观察到所有种类的请求

很多时候处于自己的目的可以选择all右边的其他选项，比如常见的选项：

- XHR：大部分情况表示ajax请求
- JS：js请求
- CSS：css请求
- 搜索数据包：辅助定位（不是万能）

但是很多时候我们并不能保证我们需要的请求是什么类型，特别是我们不清楚一个请求是否为ajax请求的时候，直接选择all,从前往后观察即可，其中js, css, 图片等不去观察即可

不要被浏览器中的一堆请求吓到了，这些请求中除了js, css, 图片的请求外，其他的请求并没有多少个

3 寻找登录接口(新版本规则已更新，思路可借鉴)

回顾之前人人网的爬虫我们找到了一个登陆接口，那么这个接口从哪里找到的呢？

<http://www.renren.com>

3.1 寻找action对的url地址

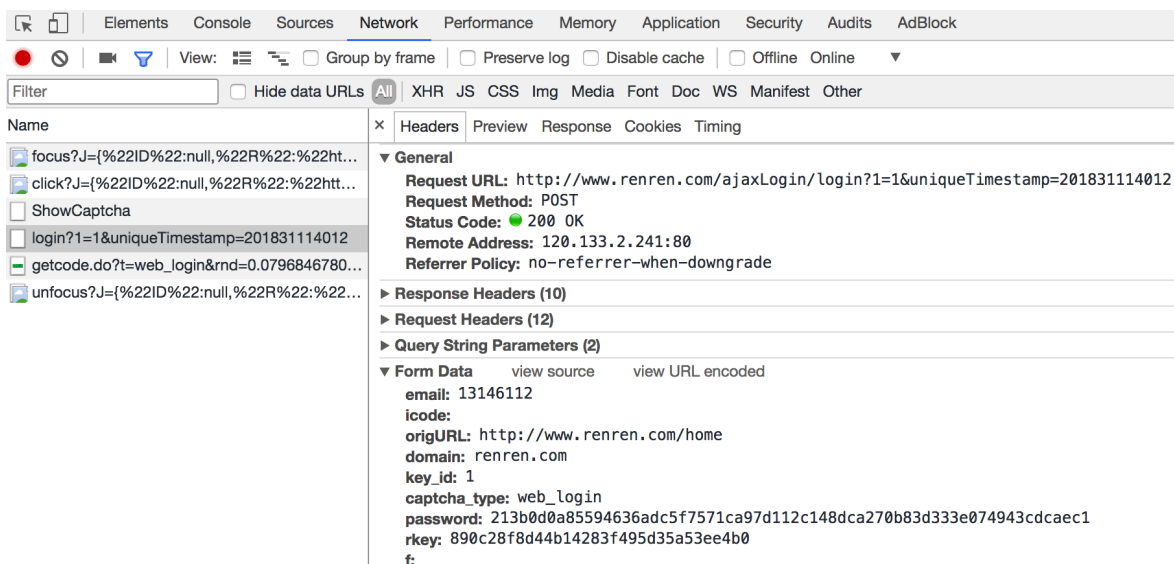
```
▼ <div class="full-page">
  ▼ <div class="login-page clearfix login-wrap">
    ▼ <div class="side-column login-box">
      ▼ <div class="login-panel">
        ▶ <div class="radiusing">...</div>
        ▶ <span id="errorMessage" class="errors_div" style="display:none;"></span>
        ▶ <div class="yellow-error" id="yellow_error">...</div>
        ▶ <div class="yellow-error" id="account_stop">...</div>
        ▶ <div class="yellow-error" id="account_lock">...</div>
        ▼ <form method="post" id="loginForm" class="login-form" action="http://www.renren.com/PLLogin.do"> == $0
          ▶ <dl class="top clearfix">...</dl>
          ▶ <dl class="pwd clearfix">...</dl>
          ▶ <div class="caps-lock-tips" id="capsLockMessage" style="display:none;"></div>
          ▶ <dl class="savepassword clearfix">...</dl>
          ▶ <dl id="code" class="code clearfix" style="display: none;">...</dl>
          ▶ <dl id="codeimg" class="codeimg clearfix" style="display: none;">...</dl>
          ▶ <dl class="bottom">...</dl>
        </form>
      ▶ <div class="regnow">...</div>
      ▶ <p class="third-party-title">...</p>
```

可以发现，这个地址就是在登录的form表单中action对应的url地址，回顾前端的知识点，可以发现就是进行表单提交的地址，对应的，提交的数据，仅仅需要：用户名的input标签中，name的值作为键，用户名作为值，密码的input标签中，name的值作为键，密码作为值即可

思考：

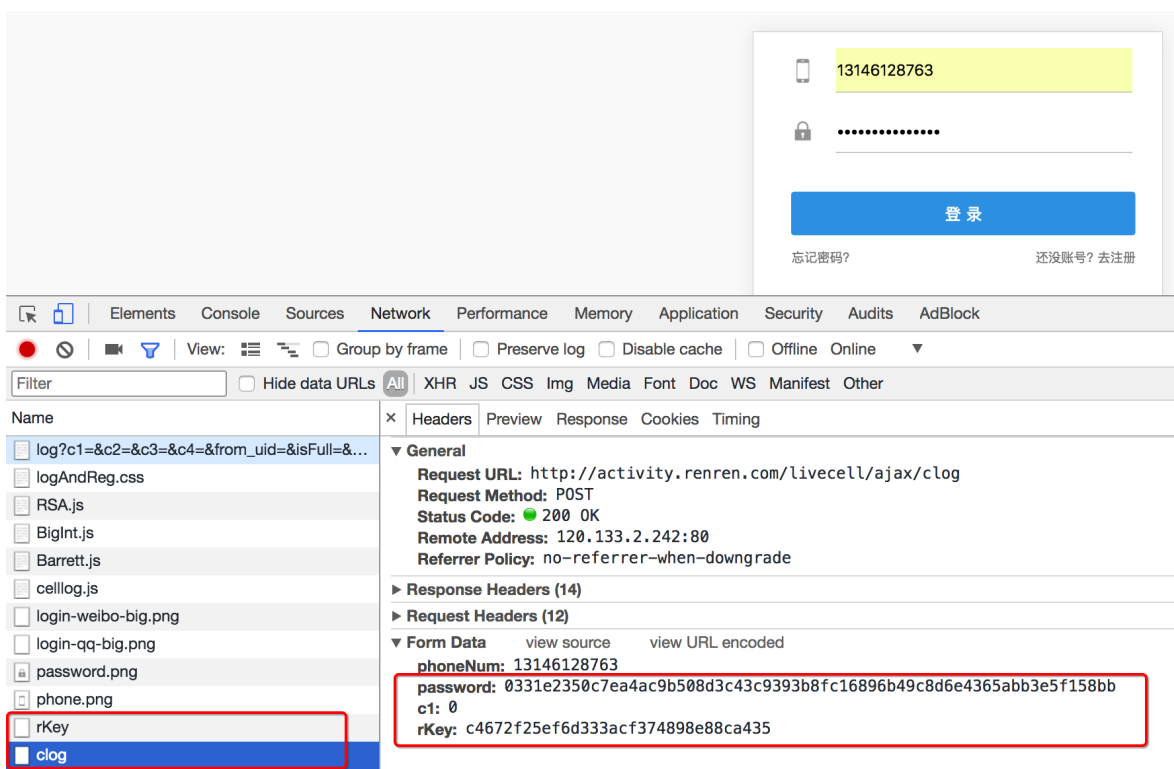
如果action对应的没有url地址的时候可以怎么做？

3.2 通过抓包寻找登录的url地址



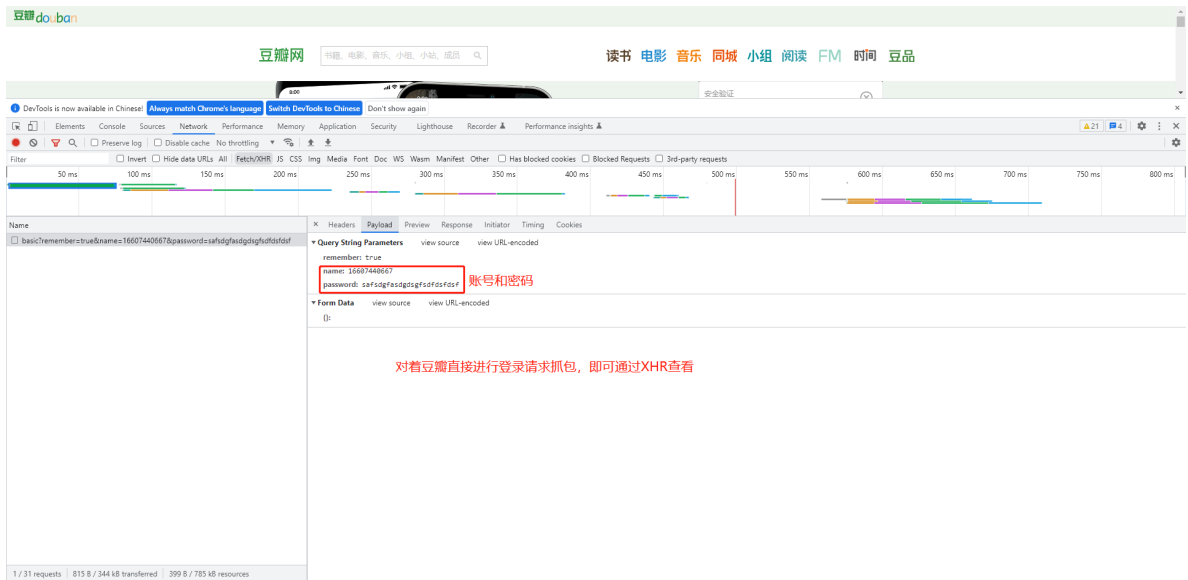
通过抓包可以发现，在这个url地址和请求体中均有参数，比如uniqueTimestamp和rkey以及加密之后的password

这个时候我们可以观察手机版的登录接口，是否也是一样的



可以发现在手机版中，依然有参数，但是参数的个数少一些，这个时候，我们可以使用手机版作为参考

4.寻找豆瓣登录地址



The screenshot shows the Douban website (豆瓣网) with the Chrome DevTools Network tab open. The selected request is a POST to `basic?remember=true&name=16607440667&password=safsdgfsdgsfufdsfdr`. The payload is visible, showing the login credentials:

```
remember: true
name: 16607440667
password: safsdgfsdgsfufdsfdr
```

The password field is highlighted with a red box, and the text "账号和密码" (Username and Password) is written next to it. Below the payload, the "Form Data" section is also visible, showing an empty object: `{}`.

对着豆瓣直接进行登录请求抓包，即可通过XHR查看