



链滴

Java | jsoup 网络爬虫模拟登录思路解析

作者: [haojieli](#)

原文链接: <https://ld246.com/article/1489060479174>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

直奔主题：

本篇文章是给有jsoup抓包基础的人看的.....

今天小编给大家写一篇对于jsoup抓包时需要输入验证码的解决方法之一。大神就绕道，嘿嘿~

任何抓包的基础都是基于Http协议来进行这个抓包行为的，那么，在我们遇到验证码的时候怎么办呢？别着急，继续看

遇到需要验证码的这种情况怎么办：

例如我们要访问www.xxxxx.com这个网站需要你输入验证码，步骤如下：

1) 首先我们二话不说,先访问一下www.xxxxx.com这个网址,目的是什么？

a.获取到cookies

b.为我们下一步获取验证码做好准备

c.利用jsoup对www.xxxxx.com的第一次访问获取到的cookie来进行我们第二次的访问。

2)我们第一次访问完毕后进行第二次访问要做什么呢？

a.利用jsoup带着cookie去访问www.xxxxx.com/img/verifyCode.gif这个地址获取验证码

b.将获取到的验证码存入到本地(一般获取到的是二进制形式的)

c.利用jsoup带着cookie到www.xxxxx.com/img/verifyCode.gif获取到的验证码我们是不是可以登录呢？

3) 第三次访问我们带入帐号+密码+验证码去登录 重要的是不要把cookie忘带了

a.第三次访问www.xxxx.com/login.html?username=haojieli&password=123456&verifyCode=134 在后面跟上cookie的值

解析：

重点是cookie，cookie是会话的首要条件，cookie就相当于你要打电话，你输入的电话要和对方的话号码对得上才能通上话，所以后面获取验证码和登录都必须带上cookie，这是你访问对方服务器的时候识别你的标准。

反之：你到服务器获取验证码时候的cookie为123456789，而你登录的时候的cookie是888888888，你打的电话都不是同一个电话，接电话的肯定不是同一个人，所以，访问的时候注意保持cookie一致！

博文到此结束，感谢您的观看，希望对各位读者有所帮助，如果有什么意见以及建议请在评论留言.....

=====祝各位读者生活愉快=====