

这篇文档中将介绍 Requests 的一些高级特性。为了模拟浏览器的功能，只有之前的知识是不够的，在学习了高级的知识后，我们可以更好的模拟浏览器的功能。完成更多的网页的采集工作。

## 1、状态保持

HTTP协议是无状态的协议。无状态是指协议对于事务处理没有记忆功能。缺少状态意味着，假如后面的处理需要前面的信息，则前面的信息必须重传，这样可能导致每次连接传送的数据量增大。

另一方面，在服务器不需要前面信息时，应答就较快。直观地说，就是每个请求都是独立的，与前面的请求和后面的请求都是没有直接联系的。因此，Cookie和Session存在的作用是进行状态管理。会话对象让你能够跨请求保持某些参数。

Cookie，有时也用其复数形式 [Cookies](#)。类型为“小型文本文件”，是某些网站为了辨别用户身份，进行[Session](#)跟踪而储存在用户本地终端上的数据（通常经过加密），由用户[客户端](#)计算机暂时或永久保存的信息。——百度百科

### 1、1 Cookie

现在的网站中有这样的一种网站类型，也就是需要用户注册以后，并且登陆才能访问的网站，或者说在不登录的情况下不能访问自己的私有数据，例如微博、微信等。

网站记录用户信息的方式就是通过客户端的Cookie值。例如：当我们在浏览器中保存账号和密码的同时，浏览器在我们的电脑上保存了我们的用户信息，并且在下次访问这个页面的时候就会自动的为我们加载Cookie信息。

在需要登陆的网站中，浏览器将Cookie中的信息发送出去，服务器验证Cookie信息，确认登录。既然浏览器在发送请求的时候带有Cookie信息，那么我们的程序同样也要携带Cookie信息。

**Cookie是当你访问某个站点或者特定页面的时候，留存在电脑里的一段文本，它用于跟踪记录网站访问者的相关数据信息，比如：搜索偏好、行为点击、账号、密码等内容。**

## web访问全过程

### 浏览器访问WEB服务器的过程

在用户访问网页时，不论是通过URL输入域名或IP，还是点击链接，浏览器向WEB服务器发出了一个HTTP请求（Http Request），WEB服务器接收到客户端浏览器的请求之后，响应客户端的请求，发回相应的响应信息（Http Response），浏览器解析引擎，排版引擎分析返回的内容，呈现给用户。WEB应用程序在于服务器交互的过程中，HTTP请求和响应时发送的都是一个消息结构。

#### (1) 什么是 cookie

cookie在http请求和http响应的头信息中，cookie是消息头的一种很重要的属性。当用户通过浏览器首次访问一个域名时，访问的WEB服务器会给客户端发送数据，以保持WEB服务器与客户端之间的状态保持，这些数据就是Cookie，它是 Internet 站点创建的,为了辨别用户身份而储存在用户本地终端上的数据，Cookie中的信息一般都是经过加密的，Cookie存在缓存中或者硬盘中，在硬盘中的是一些小文本文件,当你访问该网站时，就会读取对应网站的Cookie信息，Cookie有效地提升了我们的上网体验。一般而言，一旦将 Cookie 保存在计算机上，则只有创建该 Cookie 的网站才能读取它。

#### (2) 为什么需要 cookie

Http协议是一个无状态的面向连接的协议，Http协议是基于tcp/ip协议层之上的协议，当客户端与服务器建立连接之后，它们之间的TCP连接一直都是保持的，至于保持的时间是多久，是通过服务器端来设置的，当客户端再一次访问该服务器时，会继续使用上一次建立的连接，但是，由于Http协议是无状态的，WEB服务器并不知道这两个请求是否同一个客户端，这两次请求之间是独立的。为了解决这个问题，Web程序引入了Cookie机制来维护状态.cookie可以记录用户的登录状态，通常web服务器会在用户登录成功后下发一个签名来标记Cookie的有效性，这样免去了用户多次认证和登录网站。记录用户的访问状态。

比如说有些网站需要登录后才能访问某个页面，在登录之前，你想抓取某个页面内容是不允许的，那么我们可以利用Requests库保存我们登录的Cookie，然后再抓取其他页面就达到目的了。

## cookie 的种类

会话Cookie(Session Cookie): 这个类型的cookie只在会话期间内有效，保存在浏览器的缓存之中，用户访问网站时，会话Cookie被创建，当关闭浏览器的时候，它会被浏览器删除。

持久Cookie(Persistent Cookie): 这个类型的cookie长期在用户会话中生效。当你设置cookie的属性Max-Age为1个月的话，那么在这个月里每个相关URL的http请求中都会带有这个cookie。所以它可以记录很多用户初始化或自定义化的信息，比如什么时候第一次登录及弱登录态等。

Secure cookie：安全cookie是在https访问下的cookie形态，以确保cookie在从客户端传递到Server的过程中始终加密的。

HttpOnly Cookie：这个类型的cookie只能在http(https)请求上传递，对客户端脚本语言无效，从而有效避免了跨站攻击。

第三方cookie：第一方cookie是当前访问的域名或子域名下的生成的Cookie。第三方cookie:第三方cookie是第三方域名创建的Cookie。

## cookie 的构成

Cookie是http消息头中的一种属性，包括：Cookie名字（Name）Cookie的值（Value），Cookie的过期时间（Expires / Max-Age），Cookie作用路径（Path），Cookie所在域名（Domain），使用Cookie进行安全连接（Secure）。前两个参数是Cookie应用的必要条件，另外，还包括Cookie大小（Size，不同浏览器对Cookie个数及大小限制是有差异的）。

Cookie值信息可以在浏览器中复制过来,放到headers中，如下所示：

```
import requests

url = 'https://fanyi.baidu.com/v2transapi?from=zh&to=en'
headers = {
```

```
# 'Cookie': 'BIDUPSID=14D1F44585EF6D4EE5741C3668A7ED09; PSTM=1606978365;
BAIDUID=14D1F44585EF6D4E35EE926D1F67DFDC:FG=1; REALTIME_TRANS_SWITCH=1;
FANYI_WORD_SWITCH=1; HISTORY_SWITCH=1; SOUND_PREFER_SWITCH=1; SOUND_SPD_SWITCH=1;
MCITY=-220%3A274%3A;
BDUSS=p1bw5sbnU5SDNrwi0xLXdSe1Jxa3N4TX54dHpuN2s3cnVZZ0VINUFpdu52N2xnRVFBQUFBJCQA
AAAAAAAAAAEAAAD36OLxx-
C1xr3M0~0AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAI0ykmCNMPJgcw; __yjs_duid=1_86cc2d00a67b06346def825d25716db71620460448943;
BDUSS_BFESS=p1bw5sbnU5SDNrwi0xLXdSe1Jxa3N4TX54dHpuN2s3cnVZZ0VINUFpdu52N2xnRVFBQU
FBJCQAAAAAAAAAAEAAAD36OLxx-
C1xr3M0~0AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAI0ykmCNMPJgcw; BDORZ=B490B5EBF6F3CD402E515D22BCDA1598;
H_PS_PSSID=34099_33848_33607_34135_34118_26350; delPer=0; PSINO=6;
BA_HECTOR=242k8h8l250g80akcq1gd0ji60r; BCLID=6714930875486328860;
BDSFRCVID=eH_OJexroG38EYQeF-JLIGt58FweG7bTDYLEOWXPsp3LGJLVJec6EG0Pts1-dEu-
EHtdogKKLm0THpKF_2uxOjjg8UtVJec6EG0Ptf8g0M5; H_BDCLCKID_SF=tJ4joc8KtDK3qn5zK-
QJbDCs5fQEaC62aKDs3qj1BhcqJ-ovQT3Z3xr0btDLQfoA35PtVUJCBUONHxbewfvph5KU-GQ9y-
cA3Hop0Ko1L15nhmJmb4JWLfIz-lQayMry523i2n6vQpn2OpQ3DRowXPIqbn7P-
p5Z5mAqKl0MLPbtbb0xxj_0D60BjaAjt5Ds-DTbB4oHK--_qnTz-4L_5-_e-xQyetJyaR0j-
hovWJ5TMCoq0no4KRK8QmcwhfcxbnIO-ljltpF-ShPC-tnqbxAhwfbr-fj72jv9BPTx3l02VbcEe-
t2ynLVyxchbPRMW23rw17mWPJvsxA45J7cM4IseboJLft-
0bc4KKJxbnLweIJEjjCaepDyqx5Ka43tHD7yWCvHWU7cOR59K4nnDTkkWJQ-
0DrZBg5xLIKY3t3qsI3P3MOZXMLg5n7Tbb8eBgVZ2UQF5l8-
sq0x0b05DDuOQq_L0xvJ5IOMahv1a17xo-JoQlPK5JkgMx6MqpQJQeQ-5KQN3KJmfbL9bT3YjjISKx-
_tTKOJRTP; BCLID_BFESS=6714930875486328860; BDSFRCVID_BFESS=eH_OJexroG38EYQeF-
JLIGt58FweG7bTDYLEOWXPsp3LGJLVJec6EG0Pts1-dEu-
EHtdogKKLm0THpKF_2uxOjjg8UtVJec6EG0Ptf8g0M5;
H_BDCLCKID_SF_BFESS=tJ4joc8KtDK3qn5zK-QJbDCs5fQEaC62aKDs3qj1BhcqJ-
ovQT3Z3xr0btDLQfoA35PtVUJCBUONHxbewfvph5KU-GQ9y-cA3Hop0Ko1L15nhmJmb4JWLfIz-
lQayMry523i2n6vQpn2OpQ3DRowXPIqbn7P-p5Z5mAqKl0MLPbtbb0xxj_0D60BjaAjt5Ds-DTbB4oHK-
-_qnTz-4L_5-_e-xQyetJyaR0j-hovWJ5TMCoq0no4KRK8QmcwhfcxbnIO-ljltpF-ShPC-
tnqbxAhwfbr-fj72jv9BPTx3l02VbcEe-t2ynLVyxchbPRMW23rw17mWPJvsxA45J7cM4IseboJLft-
0bc4KKJxbnLweIJEjjCaepDyqx5Ka43tHD7yWCvHWU7cOR59K4nnDTkkWJQ-
0DrZBg5xLIKY3t3qsI3P3MOZXMLg5n7Tbb8eBgVZ2UQF5l8-
sq0x0b05DDuOQq_L0xvJ5IOMahv1a17xo-JoQlPK5JkgMx6MqpQJQeQ-5KQN3KJmfbL9bT3YjjISKx-
_tTKOJRTP; Hm_lvt_64ecd82404c51e03dc91cb9e8c025574=1624264265;
Hm_lpv_64ecd82404c51e03dc91cb9e8c025574=1624264265;
__yjs_st=2_Mzg1n2UzODI2Nz1jM2FmYTMwODE0ZDM5NDA3ZmYxNWUzMdVKNWviYTA0NDA0YmYwMTdio
DEXm2QxZDc1MGIxM2E2YTZmMDI3NzI5ZjFjYjMzNzkyN2NjZmU3YmI4MGY3MTRiY2M1N2E4ZTZiYzFmN
2JlMDRjNWJkZjJhZDgxNzZiYTk4NmFiYTA0OWE5OTYwNTBjYWJiMzRlYWFnNGJkNTRjMWEXNTU1ZWZjM
zYxNzhkMzJhMmJmZjYwMDQ4NjgwNDQ5ZDhlZjY5ZjY4NDk2OTdmYTgxMDNmZDUwYmZmOWYwOGVmZTgxY
WFjMTk1NDQ1ODdiNjk4NDM4ODZmODI3M183XzFiOGU1Y2Ix;
ab_sr=1.0.1_NzY1YjQ3MDhmZmE4ZTFjMzcyNjAzMDU5Zju3YjViZTY5MzU0MmYzODRlNzRhNWY3YzRh
Mzc2ZjA3N2Y4YmJkMTRmMTc2MTZiNzmxY2Q4Mjk3MzgwNzU4NDY1MDAwMmQ4OTl1MGI2MDB1ZTI3YzJj
NTQyMDAwNzU1ZTc1NjRmMjM3YmI2NjY5ZDEYmEzZDAxNGVjNWIXOWQ2MTBhMTlmNWZiYWY5MjNkZWZ3
OGI3NDl1ZGVhOGQ3OTQ3M2U5MGY4',
'Host': 'fanyi.baidu.com',
'Origin': 'https://fanyi.baidu.com',
'Referer': 'https://fanyi.baidu.com/?aldtype=16047',
'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36
(KHTML, like Gecko) Chrome/90.0.4430.212 Safari/537.36',
}
data = {
    'from': 'zh',
    'to': 'en',

```

```
{
    'query': '你好',
    'transtype': 'realtime',
    'simple_means_flag': '3',
    'sign': '232427.485594',
    'token': '5e393efd8ddb20a2afa0cfe318606c',
    'domain': 'common'
}
```

# 以字典的形式定义cookies，方便更换cookies

cookies = { # 修改成片段的cookies，替换换行符后 BAIDUID 这个键值对有格式需要调整

```
    'BIDUPSID': '14D1F44585EF6D4EE5741C3668A7ED09',
    'PSTM': '1606978365',
    'BAIDUID': '14D1F44585EF6D4E35EE926D1F67DFDC:FG=1',
    'REALTIME_TRANS_SWITCH': '1',
    'FANYI_WORD_SWITCH': '1',
    'HISTORY_SWITCH': '1',
    'SOUND_PREFER_SWITCH': '1',
    'SOUND_SPD_SWITCH': '1',
    'MCITY': '-220%3A274%3A',
    'BDUSS':
'p1bw5sbnU5SDNrwi0xLXdSe1Jxa3N4TX54dHpun2s3cnvZZ0VINUFpdu52N2xnRVFBQUFBJCQAAAAA
AAAAEAAAD360Lxx-
C1xr3M0~0AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAI0ykmCNmpJgcw',
    '__yjs_duid': '1_86cc2d00a67b06346def825d25716db71620460448943',
    'BDUSS_BFESS':
'p1bw5sbnU5SDNrwi0xLXdSe1Jxa3N4TX54dHpun2s3cnvZZ0VINUFpdu52N2xnRVFBQUFBJCQAAAAA
AAAAEAAAD360Lxx-
C1xr3M0~0AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAI0ykmCNmpJgcw',
    'BDORZ': 'B490B5EBF6F3CD402E515D22BCDA1598',
    'H_PS_PSSID': '34099_33848_33607_34135_34118_26350',
    'delPer': '0',
    'PSINO': '6',
    'BA_HECTOR': '242k8h8l250g80akcq1gd0ji60r',
    'BCLID': '6714930875486328860',
    'BDSFRCVID': 'eH_OJexroG38EYQeF-JLIgt58FweG7bTDYLEowXPsp3LGJLVJec6EG0Pts1-
dEu-EHtdogKKLmOTHpKF_2ux0jjg8UtVJec6EG0Ptf8g0M5',
    'H_BDCLCKID_SF': 'tJ4joc8KtDK3qn5zk-QJbDCs5fQEac62aKDs3qj1BhcqJ-
ovQT3Z3xr0btDLQfoA35PtVUJCBUONHxbewfvph5KU-GQ9y-cA3Hop0Ko1L15nhMJmb4JWLfIz-
1QayMry523i2n6vQpn2OpQ3DRowXPIqbn7P-p5Z5mAqk10MLPbtbb0xxj_0D60BjaAjt5Ds-DTbB4oHK-
_qnTz-4L5-_e-xQyetJyaR0j-h0vWJ5TMCoq0no4KRK8QmcwhfcxbnIO-1j1tpF-ShPC-
tnqbxAhwfbr-fJ72jv9BPTx3102VbcEe-t2ynLVyxchbPRMW23rw17mWPJvsxA45J7cm4IseboJLft-
0bc4KKJxbnLweIJEjjCaePDyqx5Ka43tHD7yWcVHWU7cOR59K4nnDTkkWJQ-
0DrZBg5xLIKY3t3qsI3P3MOZXMLg5n7Tbb8eBgvZ2UQF518-
sq0x0b05DDuOQq_L0xvJ5IOMahv1a17xo-JoQ1PK5Jkgmx6MqpQJQeQ-5KQN3KJmfbL9bT3YjjISKx-
_tTKOJRTP',
    'BCLID_BFESS': '6714930875486328860',
    'BDSFRCVID_BFESS': 'eH_OJexroG38EYQeF-
JLIgt58FweG7bTDYLEowXPsp3LGJLVJec6EG0Pts1-dEu-
EHtdogKKLmOTHpKF_2ux0jjg8UtVJec6EG0Ptf8g0M5',
```

```

'H_BDCLCKID_SF_BFESS': 'tJ4joc8KtDK3qn5zK-QJbDCs5fQEaC62akDs3qj1BhcqJ-
ovQT3Z3xr0btDLQfoA35PtVUJCBuONHxbewfvph5KU-GQ9y-ca3Hop0Ko1L15nhMJmb4JWLFiz-
lQayMry523i2n6vQpn2OpQ3DRoWXPiQbn7P-p5Z5mAqKl0MLPbtbb0xxj_0D60BjaAjt5Ds-DTbB4oHK-
-_qnTz-4L_5-_e-xQyetJyar0j-h0vWJ5TMCoq0no4KRK8QmcwhfcxbnIO-lj1tpF-ShPC-
tnqbxAhwfbr-fJ72jv9BPTx3l02VbcEe-t2ynLVyxchbPRMW23rw17mWPJvsxA45J7CM4IseboJLft-
0bc4KKjxbnLweIJEjjCaepDyqx5Ka43tHD7ywcVHWU7cOR59K4nnDTkkWJQ-
0DrZBg5xLIky3t3qsI3P3MOZXMLg5n7Tbb8eBgVZ2UQF5l8-
sq0x0b05DDuOQq_L0xvJ5IOMahv1a17xo-JoQlPK5JkgMx6MqpQJQeQ-5KQN3KJmfbL9bT3YjjISKx-
_tTKOJRTP',
'Hm_lvt_64ecd82404c51e03dc91cb9e8c025574': '1624264265',
'Hm_lpv_64ecd82404c51e03dc91cb9e8c025574': '1624264265',
'__yjs_st':
'2_Mzg1N2UZODI2Nzljm2FmYTMwODE0ZDM5NDA3ZmYxNWUZMDVknWviYTA0NDA0YmYwMTdiODExM2QxZ
Dc1MGIXM2E2YTZmMDI3NzI5ZjFjYjMzNzkyN2NjZmU3YmI4MGY3MTRiY2M1N2E4ZTZiYzFmN2JlMDRjN
WJkZjJhZDgxNzZiYTk4NmFiYTA0OWE5OTYwNTBjYWJiZmRlYWFMNGJkNTRjMwExNTU1ZWZjMzYxNzhkM
zJhMmJmZjYwMDQ4NjgwNDQ5ZDhlZjY5ZjY4NDk2OTdmYTgxMDNmZDUwYmZmOWYwOGVmZTgxYWFjMTk1N
DQ1ODdiNjk4NDM4ODZmODI3M183XzFiOGU1Y2Ix',
'ab_sr':
'1.0.1_NzY1YjQ3MDhmZmE4ZTFjMzcyNjAzMDU5ZjU3YjYviZTY5MzU0MmYzODRlNzRhNWY3YzRhMzc2Z
jA3N2Y4YmJkMTRmMTc2MTZiNzIxY2Q4Mjk3MzgwNzU4NDY1MDAwMmQ4OTl1MGI2MDBlZTI3YzJjNTQyM
DAwNzU1ZTc1NjRmMjM3YmI2NjY5ZDEyMmEzZDAXNGVjNWixOWQ2MTBhMTlMNWZiYwY5MjNkZWE3OGI3N
Dl1ZGVhOGQ3OTQ3M2U5MGY4',
}

response = requests.post(url=url, headers=headers, data=data, cookies=cookies)
print(response.json())

```

这样就可以随着请求头一起发送出去。当然requests也提供了cookies 关键字参数,供我们提交Cookie信息:

```

import requests
url = xxx
cookies = {'Cookie': '你的Cookie值'}
r = requests.get(url, cookies=cookies)

```

#### 注意事项:

- 在发出请求的时候有两种方式,一种是将Cookie添加到 headers 参数中,另外一种是将Cookie添加到 cookies 参数中,并且都是以字典格式。
- 如果使用了自己的账号,请谨慎操作,不然你的Cookie有可能暴露给别人。
- 在爬取需要登陆的网站的时候,尽量降低爬虫程序的访问频率。谨防账号被封锁。

## 2、会话维持

现代的浏览器都已经有了维持会话这种功能,可以想象一下这样的场景:一个用户在浏览器中登陆了自己的账号,这样他就可以浏览主页了,但是该用户还想浏览该网站的其他页面,如果浏览器没有维持这次会话,那么用户访问该域名下的其他页面是就需要重复登陆。这显然比较麻烦。为了解决这样的问题,需要实现跨请求。

## 2、1 Session

Session：在计算机中,尤其是在网络应用中,称为“会话控制”。Session对象存储特定用户会话所需的属性及配置信息。这样,当用户在应用程序的Web页之间跳转时,存储在Session对象中的变量将不会丢失,而是在整个用户会话中一直存在下去。当用户请求来自应用程序的 Web页时,如果该用户还没有会话,则Web服务器将自动创建一个 Session对象。当会话过期或被放弃后,服务器将终止该会话。**其实维持会话的目的就是共享不同页面的Cookie。**

```
'''
    维持会话
'''
import requests
# 设置Cookie的网址
url_set_cookie = 'http://httpbin.org/cookies/set/sessioncookie/123456789'
# 获取Cookie的网址
url_get_cookie = 'http://httpbin.org/cookies'
# 维持会话实例化
session = requests.Session()
# 发送请求设置Cookie
response_set_cookie = session.get(url_set_cookie)
# 发送请求得到Cookie
response_get_cookie = session.get(url_get_cookie)
# 打印获取的Cookie值
print(response_get_cookie.text)
```

结果是:

```
{
  "cookies": {
    "sessioncookie": "123456789"
  }
}
```

结果显示,在两次无关的请求中维持了相同的Cookie,也就是实现了共享Cookie值。

**注意:**

就算使用了会话,**方法级别的参数也不会被跨请求保持**。下面的例子只会和第一个请求发送Cookie,而非第二个:

```
'''
    维持会话
'''
import requests
# 获取Cookie的网址
url_get_cookie = 'http://httpbin.org/cookies'
# 维持会话实例化
session = requests.Session()
# 发送请求得到Cookie
response_get_cookie = session.get(url_get_cookie, cookies={'form-
my': 'scrapy@qq.com'})
# 打印获取的Cookie值
print(response_get_cookie.text)
# 再次请求不会维持参数级别的值
```

```
r = session.get(url_get_cookie)
print(r.text)
```

结果是:

```
{
  "cookies": {
    "form-my": "scrapy@qq.com"
  }
}
#####
{
  "cookies": {}
}
```

### 案例：青灯论坛模拟注册

```
# http://81.68.68.240:5000/login 登录地址

"""
时间戳：指格林威治时间1970年1月1日0时0分0秒 到 目前为止的总毫秒数
秒级时间戳：10数字
毫秒级时间戳：13数字
微秒级时间戳：16数字
"""

import time
import requests

def get_time():
    """获取时间戳"""
    now_time = str(int(time.time() * 1000))
    print('当前的时间戳为: ', now_time)
    return now_time

# 创建一个会话维持对象
session = requests.Session()

headers = {'User-Agent': 'Mozilla/5.0 (windows NT 10.0; win64; x64)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/86.0.4240.183 Safari/537.36'}
cookies = {'Cookie':
'session=.eJyrVopPy0kszkgtVrKKr1ZSKAFSSrmpxcWJ6a1K0kov1m9_PnP3071Tn-xe-
mLd_pft1z1bu-
Tlwq0v5y5Siq3VIVVDrA7QuqLU4gw1q7TENoJUHaXk_JRUJSS1mZmjI6VaAPqoQyU.YuDpig.eMX-
MKR2-aZgSwjb1B-1voB49A4'}
```

```
"""请求验证码图片保存"""
img_url = 'http://81.68.68.240:5000/passport/getCaptcha?' + get_time()
print(img_url)

img_response = session.get(url=img_url, headers=headers, cookies=cookies)
img_data = img_response.content
print(img_response.cookies.get_dict())
with open('yzm.jpg', mode='wb') as f:
```



```

f.write(img_data)

# 手动输入验证码
code = input('请输入验证码: ')
print('您输入的验证码为: ', code)

"""发送模拟注册请求"""
# 构建请求json参数
json_data = {
    'username': 'admin',
    'password': '123456',
    'captcha': code
}

sms_url = 'http://81.68.68.240:5000/api/v1/passport/login'

sms_response = session.post(url=sms_url, json=json_data)
print(sms_response.json())

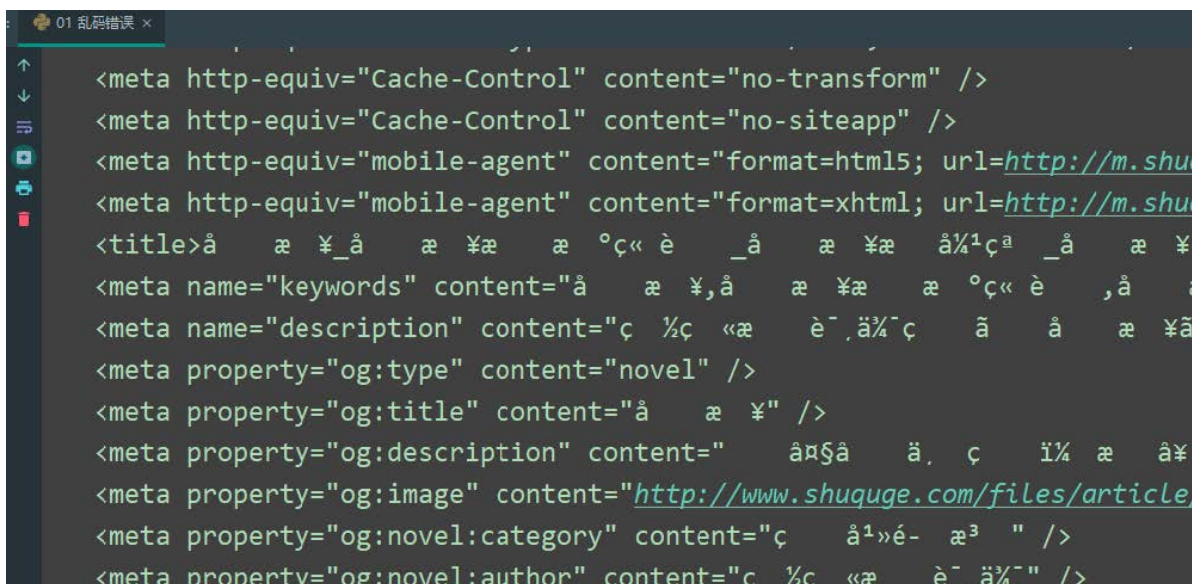
```

注意:

1. 要跳转不同的页面,需要维持相同的会话。
2. 当需要处理带有验证码的网站,也需要维持会话。
3. 登陆以后可以尝试访问其他页面测试是否成功登陆。
4. 如果可以获取Cookie就可以直接使用Cookie,可以不使用Session。
5. 这种方式的模拟登陆对简单网页来说还可以,但是登陆过程复杂的网站往往束手无策,后续我们会学习Selenium来实现登陆。

## 3、requests模块常见异常处理

### 3、1 网页出现乱码



出现乱码的原因是因为网页解码过程中没有设置如何编码

```
response.encoding = response.apparent_encoding
```



### 3、2 请求头参数错误

InvalidHeader: Invalid return character or leading space in header: User-Agent

```
import requests

headers = {
    'User-Agent': ' Mozilla/5.0 (Windows NT 10.0; WOW64)
AppleWebKit/537.36(KHTML, like Gecko) Chrome/84.0.4128.3 Safari/537.36'
}
response = requests.get('http://www.shuquge.com/txt/8659/index.html',
headers=headers)
response.encoding = response.apparent_encoding
html = response.text
print(html)
```

其实很难发现问题在哪，但事实上是因为‘Mozilla’之前多了个空格，把空格删去即可。

### 3、3 得不到数据&参数错误

```
import requests

url = 'https://fanyi.baidu.com/v2transapi?from=zh&to=en'
headers = {
```

```
# 'Cookie': 'BIDUPSID=14D1F44585EF6D4EE5741C3668A7ED09; PSTM=1606978365;
BAIDUID=14D1F44585EF6D4E35EE926D1F67DFDC:FG=1; REALTIME_TRANS_SWITCH=1;
FANYI_WORD_SWITCH=1; HISTORY_SWITCH=1; SOUND_PREFER_SWITCH=1; SOUND_SPD_SWITCH=1;
MCITY=-220%3A274%3A;
BDUSS=p1bw5sbnU5SDNrwi0xLXdSe1Jxa3N4TX54dHpuN2s3cnVZZ0VINUFpdu52N2xnRVFBQUFBJCQA
AAAAAAAAAAEAAAD36OLxx-
C1xr3M0~0AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAI0ykmCNMPJgcw; __yjs_duid=1_86cc2d00a67b06346def825d25716db71620460448943;
BDUSS_BFESS=p1bw5sbnU5SDNrwi0xLXdSe1Jxa3N4TX54dHpuN2s3cnVZZ0VINUFpdu52N2xnRVFBQU
FBJCQAAAAAAAAAAEAAAD36OLxx-
C1xr3M0~0AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAI0ykmCNMPJgcw; BDORZ=B490B5EBF6F3CD402E515D22BCDA1598;
H_PS_PSSID=34099_33848_33607_34135_34118_26350; delPer=0; PSINO=6;
BA_HECTOR=242k8h8l250g80akcq1gd0ji60r; BCLID=6714930875486328860;
BDSFRCVID=eH_OJexroG38EYQeF-JLIGt58FweG7bTDYLEOWXPsp3LGJLVJec6EG0Pts1-dEu-
EHtdogKKLm0THpKF_2uxOjjg8UtVJec6EG0Ptf8g0M5; H_BDCLCKID_SF=tJ4joc8KtDK3qn5zK-
QJbDCs5fQEaC62aKDs3qj1BhcqJ-ovQT3Z3xr0btDLQfoA35PtVUJCBUONHxbewfvph5KU-GQ9y-
cA3Hop0Ko1L15nhmJmb4JWLfIz-lQayMry523i2n6vQpn2OpQ3DRowXPIqbn7P-
p5Z5mAqKl0MLPbtbb0xxj_0D60BjaAjt5Ds-DTbB4oHK--_qnTz-4L_5-_e-xQyetJyaR0j-
hovWJ5TMCoq0no4KRK8QmcwhfcxbnIO-ljltpF-ShPC-tnqbxAhwfbr-fj72jv9BPTx3l02VbcEe-
t2ynLVyxchbPRMW23rw17mWPJvsxA45J7cM4IseboJLft-
0bc4KKJxbnLweIJEjjCaepDyqx5Ka43tHD7yWCvHWU7cOR59K4nnDTkkWJQ-
0DrZBg5xLIKY3t3qsI3P3MOZXMLg5n7Tbb8eBgVZ2UQF5l8-
sq0x0b05DDuOQq_L0xvJ5IOMahv1a17xo-JoQlPK5JkgMx6MqpQJQeQ-5KQN3KJmfbL9bT3YjjISKx-
_tTKOJRTP; BCLID_BFESS=6714930875486328860; BDSFRCVID_BFESS=eH_OJexroG38EYQeF-
JLIGt58FweG7bTDYLEOWXPsp3LGJLVJec6EG0Pts1-dEu-
EHtdogKKLm0THpKF_2uxOjjg8UtVJec6EG0Ptf8g0M5;
H_BDCLCKID_SF_BFESS=tJ4joc8KtDK3qn5zK-QJbDCs5fQEaC62aKDs3qj1BhcqJ-
ovQT3Z3xr0btDLQfoA35PtVUJCBUONHxbewfvph5KU-GQ9y-cA3Hop0Ko1L15nhmJmb4JWLfIz-
lQayMry523i2n6vQpn2OpQ3DRowXPIqbn7P-p5Z5mAqKl0MLPbtbb0xxj_0D60BjaAjt5Ds-DTbB4oHK-
-_qnTz-4L_5-_e-xQyetJyaR0j-hovWJ5TMCoq0no4KRK8QmcwhfcxbnIO-ljltpF-ShPC-
tnqbxAhwfbr-fj72jv9BPTx3l02VbcEe-t2ynLVyxchbPRMW23rw17mWPJvsxA45J7cM4IseboJLft-
0bc4KKJxbnLweIJEjjCaepDyqx5Ka43tHD7yWCvHWU7cOR59K4nnDTkkWJQ-
0DrZBg5xLIKY3t3qsI3P3MOZXMLg5n7Tbb8eBgVZ2UQF5l8-
sq0x0b05DDuOQq_L0xvJ5IOMahv1a17xo-JoQlPK5JkgMx6MqpQJQeQ-5KQN3KJmfbL9bT3YjjISKx-
_tTKOJRTP; Hm_lvt_64ecd82404c51e03dc91cb9e8c025574=1624264265;
Hm_lpv_64ecd82404c51e03dc91cb9e8c025574=1624264265;
__yjs_st=2_Mzg1n2UzODI2Nz1jM2FmYTMwODE0ZDM5NDA3ZmYxNWUzMdVKNWviYTA0NDA0YmYwMTdio
DEXm2QxZDc1MGIxM2E2YTZmMDI3NzI5ZjFjYjMzNzkyN2NjZmU3YmI4MGY3MTRiY2M1N2E4ZTZiYzFmN
2JlMDRjNWJkZjJhZDgxNzZiYTk4NmFiYTA0OWE5OTYwNTBjYWJiMzRlYWFnNGJkNTRjMWEXNTU1ZWZjM
zYxNzhkMzJhMmJmZjYwMDQ4NjgwNDQ5ZDhlZjY5ZjY4NDk2OTdmYTgxMDNmZDUwYmZmOWYwOGVmZTgxY
WFjMTk1NDQ1ODdiNjk4NDM4ODZmODI3M183XzFiOGU1Y2Ix;
ab_sr=1.0.1_NzY1YjQ3MDhmZmE4ZTFjMzcyNjAzMDU5Zju3YjViZTY5MzU0MmYzODRlNzRhNWY3YzRh
Mzc2ZjA3N2Y4YmJkMTRmMTc2MTZiNzmxY2Q4Mjk3MzgwNzU4NDY1MDAwMmQ4OTl1MGI2MDBlZTI3YzJj
NTQyMDAwNzU1ZTc1NjRmMjM3YmI2NjY5ZDEyMmEzZDAxNGVjNWIXOWQ2MTBhMTlmNWZiYWY5MjNkZWZ3
OGI3NDl1ZGVhOGQ3OTQ3M2U5MGY4',
'Host': 'fanyi.baidu.com',
'Origin': 'https://fanyi.baidu.com',
'Referer': 'https://fanyi.baidu.com/?aldtype=16047',
'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36
(KHTML, like Gecko) Chrome/90.0.4430.212 Safari/537.36',
}
data = {
    'from': 'zh',
    'to': 'en',

```

```
{
  'query': '你好',
  'transtype': 'realtime',
  'simple_means_flag': '3',
  'sign': '232427.485594',
  'token': '5e393efd8ddb20a2afa0cfe318606c',
  'domain': 'common'
}
```

# 以字典的形式定义cookies，方便更换cookies

```
cookies = {'cookies': 'BIDUPSID=14D1F44585EF6D4EE5741C3668A7ED09;
PSTM=1606978365; BAIDUID=14D1F44585EF6D4E35EE926D1F67DFDC:FG=1;
REALTIME_TRANS_SWITCH=1; FANYI_WORD_SWITCH=1; HISTORY_SWITCH=1;
SOUND_PREFER_SWITCH=1; SOUND_SPD_SWITCH=1; MCITY=-220%3A274%3A;
BDUSS=p1bw5sbnU5SDNrwi0xLXdSe1Jxa3N4TX54dHpuN2s3cnVZZ0VINUFpdu52N2xnRVFBQUFBJCQA
AAAAAAAAAAEAAAD360Lxx-
C1xr3M0~0AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAI0ykmCNMPJgcw; __yjs_duid=1_86cc2d00a67b06346def825d25716db71620460448943;
BDUSS_BFESS=p1bw5sbnU5SDNrwi0xLXdSe1Jxa3N4TX54dHpuN2s3cnVZZ0VINUFpdu52N2xnRVFBQU
FBJCQAAAAAAAAAAEAAAD360Lxx-
C1xr3M0~0AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAI0ykmCNMPJgcw; BDORZ=B490B5EBF6F3CD402E515D22BCDA1598;
H_PS_PSSID=34099_33848_33607_34135_34118_26350; de1Per=0; PSIN0=6;
BA_HECTOR=242k8h8l250g80akcq1gd0ji60r; BCLID=6714930875486328860;
BDSFRCVID=eH_OJexroG38EYQeF-JLIGt58FweG7bTDYLEOWXPsp3LGJLVJec6EG0Pts1-dEu-
EHtdogKKLm0THpKF_2uxOjjg8UtVJec6EG0Ptf8g0M5; H_BDCLCKID_SF=tJ4joc8KtDK3qn5zK-
QJbDCs5fQEaC62akDs3qj1BhcqJ-ovQT3Z3xr0btDLQfoA35PtVUJCBUONHxbewfvph5KU-GQ9y-
CA3Hop0Ko1L15nhmJmb4JWLfIz-lQayMry523i2n6vQpn2OpQ3DRowXPIqbn7P-
p5Z5MaqK10MLPbtbb0xxj_0D60BjaAjt5Ds-DTB4oHK--_qnTz-4L_5-_e-xQyetJyar0j-
hovwJ5TMCoq0no4KRK8QmcwhfcxbnIO-lj1tpF-ShPC-tnqbxAhwfbr-fj72jv9BPTx3l02VbcEe-
t2ynLVyxchbPRMW23rw17mWPJvsxA45J7cm4IseboJLft-
0bc4KKJxbnLweIJEjjCaepDYqx5Ka43tHD7yWCvHWU7cOR59K4nnDTkkWJQ-
0DrZBg5xLIky3t3qsI3P3MOZXMLg5n7Tbb8eBgvZ2UQF518-
sq0x0bo5DDuOQq_L0xvJ5IOMahv1a17xo-JoQ1PK5Jkgmx6MqpQJQeQ-5KQN3KJmfbL9bT3YjjISKx-
_tTKOJRTP; BCLID_BFESS=6714930875486328860; BDSFRCVID_BFESS=eH_OJexroG38EYQeF-
JLIGt58FweG7bTDYLEOWXPsp3LGJLVJec6EG0Pts1-dEu-
EHtdogKKLm0THpKF_2uxOjjg8UtVJec6EG0Ptf8g0M5;
H_BDCLCKID_SF_BFESS=tJ4joc8KtDK3qn5zK-QJbDCs5fQEaC62akDs3qj1BhcqJ-
ovQT3Z3xr0btDLQfoA35PtVUJCBUONHxbewfvph5KU-GQ9y-CA3Hop0Ko1L15nhmJmb4JWLfIz-
lQayMry523i2n6vQpn2OpQ3DRowXPIqbn7P-p5Z5MaqK10MLPbtbb0xxj_0D60BjaAjt5Ds-DTB4oHK-
-_qnTz-4L_5-_e-xQyetJyar0j-hovwJ5TMCoq0no4KRK8QmcwhfcxbnIO-lj1tpF-ShPC-
tnqbxAhwfbr-fj72jv9BPTx3l02VbcEe-t2ynLVyxchbPRMW23rw17mWPJvsxA45J7cm4IseboJLft-
0bc4KKJxbnLweIJEjjCaepDYqx5Ka43tHD7yWCvHWU7cOR59K4nnDTkkWJQ-
0DrZBg5xLIky3t3qsI3P3MOZXMLg5n7Tbb8eBgvZ2UQF518-
sq0x0bo5DDuOQq_L0xvJ5IOMahv1a17xo-JoQ1PK5Jkgmx6MqpQJQeQ-5KQN3KJmfbL9bT3YjjISKx-
_tTKOJRTP; Hm_lvt_64ecd82404c51e03dc91cb9e8c025574=1624264265;
Hm_lpv_64ecd82404c51e03dc91cb9e8c025574=1624264265;
__yjs_st=2_Mzg1N2UzODI2NzljM2FmYTMwODE0ZDM5NDA3ZmYxNWUzMdVknWviYTA0NDA0YmYwMTdio
DEXm2QxZDc1MGIXm2E2YTzmMDI3NzI5ZjFjYjMzNzkyN2NjZmU3YmI4MGY3MTRiY2M1N2E4ZTZiYzFmN
2JlMDRjNWJkZjJhZDgxNzZiYTk4NmFiYTA0OWE5OTYwNTBjYWJiMzRlYWFnMGJkNTRjMWEXNTU1ZWZjM
zYxNzhkMzJhMmJmZjYwMDQ4NjgwNDQ5ZDhlZjY5ZjY4NDk2OTdmYTgxMDNmZDUwYmZmOWYwOGVmZTgxY
WFjMTk1NDQ1ODdiNjk4NDM4ODZmODI3M183XzfIOGU1Y2Ix;
ab_sr=1.0.1_NzY1YjQ3MDhmZmE4ZTFjMzcyNjAzMDU5Zju3YjYjZTY5MzU0MmYzODRlNzRhNWY3YzRh
Mzc2ZjA3N2Y4YmJkMTRmMTc2MTZiNzmxY2Q4Mjk3MzgwNzU4NDY1MDAwMmQ4OTl1MGI2MDBlZTI3YzJj
NTQyMDAwNzU1ZTc1NjRmMjM3YmI2NjY5ZDEyMmEzZDAXNGVjNWIXOWQ2MTBhMTlmNWZiYWY5MjNkZWE3
OGI3NDl1ZGVhOGQ3OTQ3M2U5MGY4'}
```

```
response = requests.post(url=url, headers=headers, data=data)
print(response.json())
```

"""

请求不到数据的时候，需要考虑如下方面：

headers

User-Agent

Host

Referer

Origin

Cookies

参数

params

data

加密：请求参数，查询参数，请求头字段<逆向解密>

"""

请求到的数据与期待的数据不一样, 这时候肯定是某些参数出现了问题. 就检查是不是缺少了参数或者给错了参数。

## 3、4 目标计算机积极拒绝

```
import requests
```

```
proxy_response = requests.get('http://134.175.188.27:5010/get')
proxy = proxy_response.json()
print(proxy)
```

错误：

```
requests.exceptions.ConnectionError:
HTTPConnectionPool(host='134.175.188.27',port=5010): Max retries exceeded with
url: /get (Caused byNewConnectionError('<urllib3.connection.HTTPConnection
object at 0x0000023AB83AC828>:Failed to establish a new connection: [winError
10061] 由于目标计算机积极拒绝，无法连接。',))
```

- 被识别了
- 网址输入错误了
- 服务器停止提供服务服务器了

## 3、5 链接超时

```
import requests
```

```
proxy_response = requests.get('http://134.175.188.27:5010/get', timeout=0.0001)
proxy = proxy_response.json()
print(proxy)
```

错误：

```
requests.exceptions.ConnectTimeout:
HTTPConnectionPool(host='134.175.188.27',port=5010): Max retries exceeded with
url: /get (Caused by ConnectTimeoutError(<urllib3.connection.HTTPConnection
object at 0x000002045EF9B8D0>, 'Connection to 134.175.188.27 timed out.
(connecttimeout=0.0001)'))
```

### 3、6 异常处理

```
import requests

try:
    proxy_response =
requests.get('http://134.175.188.27:5010/get',timeout=0.0001)
    proxy = proxy_response.json()
    print(proxy)
except:
    pass
```