epython模拟登陆详解

登陆网页，我们要了解下http协议，这里简单介绍一下，以便后面进行模拟登陆。

> http协议属于应用层的面向对象的协议
> 主要功能是用来传输网页，所以传输网页有他特定的规则
> 我们了解它的规则 实现多线程批量注册、投票、签到、发帖、
> 都是可行的、网页除了Flash模块不能操作以外，其他Dom网页元素节点都可以操作。

---

url地址解析：http://host[":"port][abs_path]

第一个host是主机地址，由于域名解析主机的ip，所以输入域名的时候也能找的到服务器

---

port是端口号，浏览器默认80端口所以省略 如：baidu.com:80

---

abs_path：是文件的路径 类似于我们访问文件夹一样

有时候我们会在后面还看到 ？ 问号后面有一些东西然后 = 等号来传值 这个是传递给服务器的参数、比如 访问一篇文章总要知道那篇文章的ID一样 例如以下：
百度后面传递了id属性 值为utf-8

> https://www.baidu.com/s ? ie = utf-8

url我们了解了之后，我们来说下 用户【浏览器】请求服务器

它请求方式有很多比如我们经常说的 get 和 post 其他后续用不到。

get简单来说就是获取属性 具体获取什么数据需要通过url和问号后面传递的参数来定

post也是获取数据，但是可以传递数据到服务器，比如传递账号、密码、验证码、这些。

> 两则还有一点不同的是：get传递参数的时候是在url后面添加问号来传递，后台服务器接受。Django 可以 看不到问号传递参数
> post是隐式提交给服务器，也就是说用户浏览器是看不到的，而且再为了一些安全，有的密码还会通过js一些算法加密，这是后话。

协议头（请求头）
在请求之后呢，他有一些特定的请求头数据。

```
GET /forum.php HTTP/1.1
Host: www.discuz.net
Connection: keep-alive
Cache-Control: max-age=0
Upgrade-Insecure-Requests: 1
User-Agent: Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko)
Chrome/55.0.2883.87 Safari/537.36
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Referer: http://www.discuz.net/member.php
Accept-Encoding: gzip, deflate, sdch
Accept-Language: zh-CN,zh;q=0.8
```

这是一个简单的头信息，在User-Agent中可以看到有 系统版本 浏览器版本及核心
还有Accept-Encoding中 后面使用的gzip请求压缩网页返回，一般写代码里会把他删掉，避免返回压缩后的网页，还要解压缩
ps：有的会自解压缩

了解这些之后呢，我们可以开始实战了。

分析网页

拿一个简单的网页来练手：discuz PHP开源论坛

> 分析：我用的谷歌浏览器，用自带的F12中的Network抓包就可以。

我get请求了首页以及登陆之后抓了一些数据包如下：

> POST登陆的数据包：

```
POST /member.php?mod=logging&action=login&loginsubmit=yes&infloat=yes&lssubmit=yes&inajax=1
HTTP/1.1
Host: www.discuz.net
Connection: keep-alive
Content-Length: 92
Cache-Control: max-age=0
Origin: http://www.discuz.net
Upgrade-Insecure-Requests: 1
User-Agent: Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko)
Chrome/55.0.2883.87 Safari/537.36
Content-Type: application/x-www-form-urlencoded
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Referer: http://www.discuz.net/forum.php
Accept-Encoding: gzip, deflate
Accept-Language: zh-CN,zh;q=0.8
Cookie: t7asq_4ad6_saltkey=ycXR5459; t7asq_4ad6_lastvisit=1482574934; t7asq_4ad6_sendmail=1;
t7asq_4ad6_lastact=1482578535%09plugin.php%09; pgv_pvi=1329620888; pgv_info=ssi=s5442386736

fastloginfield=username&username=python666&password=a123456789&quickforward=yes&handlekey=ls
```

中间请求头我们不用管看最后一行post提交的数据

- username 账号　我登陆前输入的账号

- password 密码 ========= a123456789
- 还有一些参数我们不管

POST成功返回响应头中有set-cookie 最后GET首页中添加这些Cookies来完成登陆，我们来写入python试一下

python模拟登陆__PHP开源论坛

在这里我使用的第三方模块requests

```
pip install requests
```

> 通过cmd 中敲入pip 安装这个模块

```
#!/usr/bin/env python
#-*- coding:utf-8 -*-

import requests

session = requests.Session() #通过这个Session可以直接传递Cookie值不需要手动取出来，get的时候在添加

url="http://www.discuz.net/member.php?
mod=logging&action=login&loginsubmit=yes&infloat=yes&lssubmit=yes&inajax=1"

data = {'fastloginfield':'username','username':'python666','password':
'a123456789','quickforward':'yes','handlekey':'ls'}

session.post(url, data) #使用post请求拿到用户信息的Cookie
#参数一：url是post请求的地址  参数二：是post提交的参数用dict

res = session.get('http://www.discuz.net/forum.php')
#session中会把刚刚返回的Set-Cookie添加到GET请求头中进行访问,将返回的对象用res名字来保存。

print res.content# 通过print打印 get登陆之后返回的页面

#登陆成功之后呢，可以发帖、签到神马的了，模拟登陆没有神马难的，主要是我前面所讲到的要了解http协议，分析网
页数据包，最后需要什么参数，获取之后传递什么参数即可。
```

> 为了避免卡死shell 我们可以使用ipython shell来进行打印，发现还是显示不全，我们还是写到文件吧

写到文件

```
f = open('d:/discuz.txt','w') # 打开一个文件如果没有使用W模式创建，如果有会被删掉再创建
f.write(res.content)# 写进返回的网页源代码
f.flush()   # 刷新写入的数据
f.close()   # 关闭f文件对象
```

我们找到d:/discuz.txt中有一段

```
<a href="home.php?mod=space&amp;uid=3012208" target="_blank" title="访问我的空间">python666</a>
```

> 看到尾部有个我登陆的账号名：python666
> ![enter description here][1]

```
<a href="home.php?mod=spacecp&amp;ac=credit&amp;showcredit=1" id="extcreditmenu"
onmouseover="delayShow(this, showCreditmenu);" class="showmenu">积分: 0</a>
```

> 以及 积分：0 这些获取之后呢，我们可以写GUI界面来操控
> 有兴趣的条友可以去研究下Tk

需要研究Tk 和 继续了解http协议 或 有什么不懂的条友可以加下QQ群:526929231

还有一些很好的技术文章尽情戳知了课堂官方QQ：2156600937