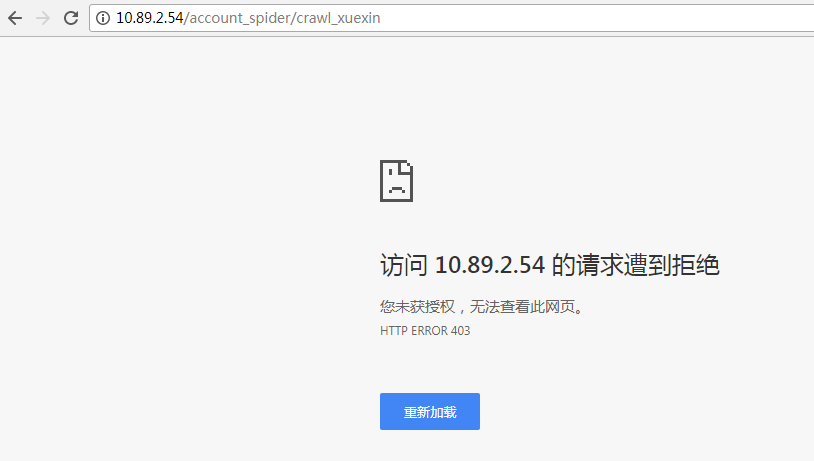
**爬虫页面AccessToken开发接入文档**

访问爬虫页面时，只要请求路径中包含了account\_spider便会自动进入访问授权模式，不管是get还是post方式都需要获取到access token才能正常访问。以学信网爬虫页面为例，页面地址：http://10.89.2.54/account\_spider/crawl\_xuexin，未带授权的access token访问的时候会返回403状态。



为了正常访问到该页面，我们需要先获取access token，具体操作如下：

1. 获取client\_id,client\_secret和public\_key。

client\_id：

EfgBbx2qh6viPsZrNH7eBss3DbcVftYZAlx6tlUu

client\_secret：

O798QPizlm1skRJcZlafVpiZzzUrUW6gGjZOXKe3Umh0Vo1hMkYe2bMihLZfFn8zVkhhqgKgBYAyBl1foFpIvyBahOu6WAzVXUmohjh8yOoJZo9RQ9c5qlgYdzGSZor4

public\_key:

-----BEGIN RSA PUBLIC KEY-----

MIGfMA0GCSqGSIb3DQEBAQUAA4GNADCBiQKBgQCqQ9Fg9J7pVIbcdiX1hiATkc3j

dEHaywEJQs4mbXo6OoJlfGiO+n5/Gz3d/tMXjuW79lzMpXkC33TRY31/Cn8xGIFt

23u1hGihW74JHdYNl3cU0Kc9I+k/8Il3WmicxZz5B8nOqxLXbLhYAlovIiQXaXaq

5KS6ckKdxhAZrGl3ZQIDAQAB

-----END RSA PUBLIC KEY-----

1. 请求access token地址，获取access token(access token有效期暂定10分钟，失效后需要重新获取)。

access token请求地址：http://10.89.2.54/access\_token

请求方式：post

请求参数：请求参数将client\_id、client\_secret和timestamp（时间戳）组装成json字符串格式之后经过rsa加密处理之后再提交

加密前的明文：

{"client\_id": "EfgBbx2qh6viPsZrNH7eBss3DbcVftYZAlx6tlUu","client\_secret": "O798QPizlm1skRJcZlafVpiZzzUrUW6gGjZOXKe3Umh0Vo1hMkYe2bMihLZfFn8zVkhhqgKgBYAyBl1foFpIvyBahOu6WAzVXUmohjh8yOoJZo9RQ9c5qlgYdzGSZor4","timestamp": "1507862762.836200"}

加密后的密文：



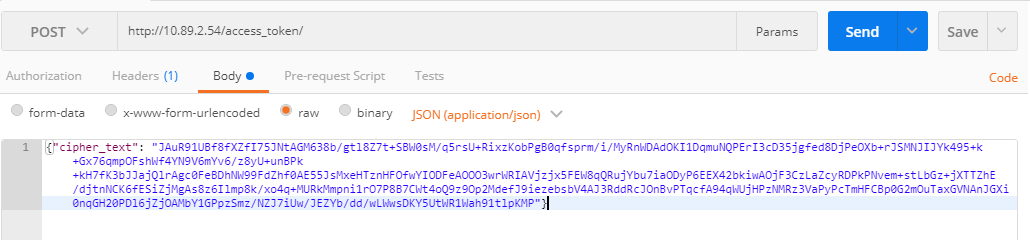
注意：明文加密之后使用了base64位进行编码处理。Python参考代码：

|  |
| --- |
| **def** rsa\_long\_encrypt(pub\_key\_str, msg, length=100):  *"""  单次加密串的长度最大为 (key\_size/8)-11  1024bit的证书用100， 2048bit的证书用 200  """* pubobj = RSA.importKey(pub\_key\_str)  pubobj = PKCS1\_v1\_5.new(pubobj)  res = **b""  for** i **in** range(0, len(msg), length):  cipher\_text = pubobj.encrypt(msg[i:i+length].encode(**"utf-8"**))  res += cipher\_text  **return** base64.b64encode(res) |

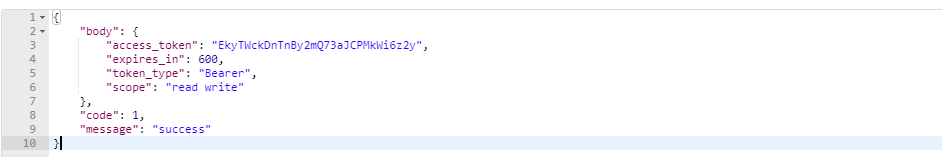
1. 得到密文之后，将密文信息提交到access token请求地址，以post方式提交，同样以json格式提交，如：

|  |
| --- |
| {"cipher\_text": ""} |

Postman示例截图：



1. 请求成功后，便获取到access token，格式如下：

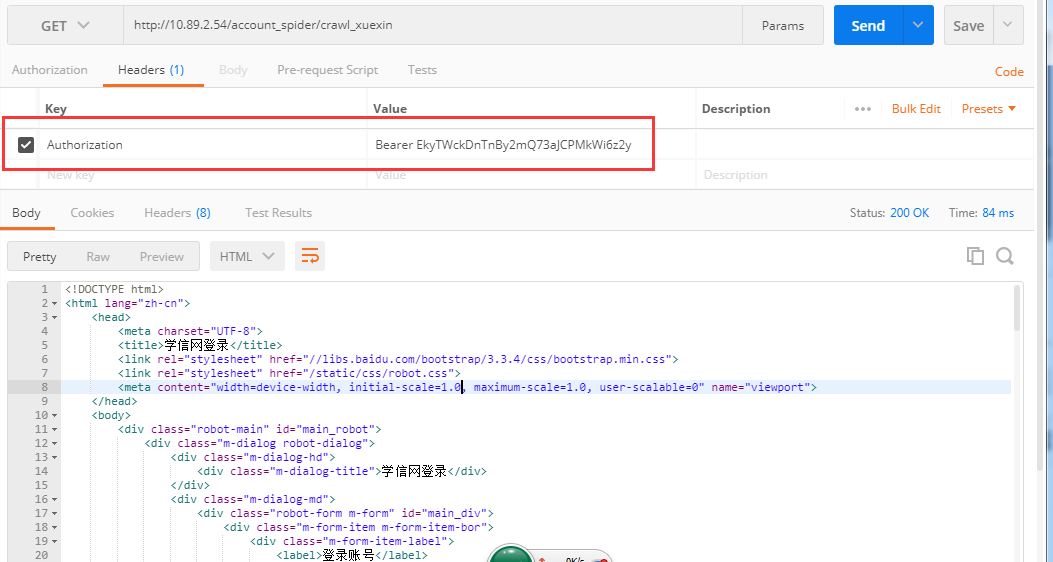


code：状态码。0表示失败，1表示成功

message：处理消息。状态码为1时为success表示成功，状态码为0时包含出错的具体信息。

body里面包含token的详细信息，access\_token为具体的授权码，expires\_in为有效期（单位秒），token\_type为授权类型，scope访问权限。

1. 将access\_type和access\_token放在请求头部里面访问页面，格式Authorization：access\_type access\_token，postman截图如下：



因考虑到app开发用的是H5，不能使用iframe设置header，故又增加了get传递token的方式，在打开页面的时候带上token参数，如：

http://10.89.2.54/account\_spider/crawl\_xuexin?token=fZngEjVWM9XXfyP7MJZu2PTbRSVM7G

至此，便能顺利访问到授权页面。特别注意，目前access\_token时间暂定为10分钟，如果后面业务有需要延长该时间，我可以统一修改。另外，client\_id和client\_secret和public\_key为机密内容，不能被非开发人员知晓，请妥善保管，为了防止别人暴力破解和反编译获取该信息，建议app开发的代码进行混淆并使用360加固。