1、python3调用urllib.urlopen(‘url’)报错：

C:\Users\pc\AppData\Local\Programs\Python\Python36\python.exe "G:/ web\_crawler/urlib.py"

Traceback (most recent call last):

File "C:\Users\pc\AppData\Local\Programs\Python\Python36\lib\urllib\request.py", line 1318, in do\_open

encode\_chunked=req.has\_header('Transfer-encoding'))

File "C:\Users\pc\AppData\Local\Programs\Python\Python36\lib\http\client.py", line 1239, in request

self.\_send\_request(method, url, body, headers, encode\_chunked)

File "C:\Users\pc\AppData\Local\Programs\Python\Python36\lib\http\client.py", line 1285, in \_send\_request

self.endheaders(body, encode\_chunked=encode\_chunked)

File "C:\Users\pc\AppData\Local\Programs\Python\Python36\lib\http\client.py", line 1234, in endheaders

self.\_send\_output(message\_body, encode\_chunked=encode\_chunked)

File "C:\Users\pc\AppData\Local\Programs\Python\Python36\lib\http\client.py", line 1026, in \_send\_output

self.send(msg)

File "C:\Users\pc\AppData\Local\Programs\Python\Python36\lib\http\client.py", line 964, in send

self.connect()

File "C:\Users\pc\AppData\Local\Programs\Python\Python36\lib\http\client.py", line 1400, in connect

server\_hostname=server\_hostname)

File "C:\Users\pc\AppData\Local\Programs\Python\Python36\lib\ssl.py", line 407, in wrap\_socket

\_context=self, \_session=session)

File "C:\Users\pc\AppData\Local\Programs\Python\Python36\lib\ssl.py", line 814, in \_\_init\_\_

self.do\_handshake()

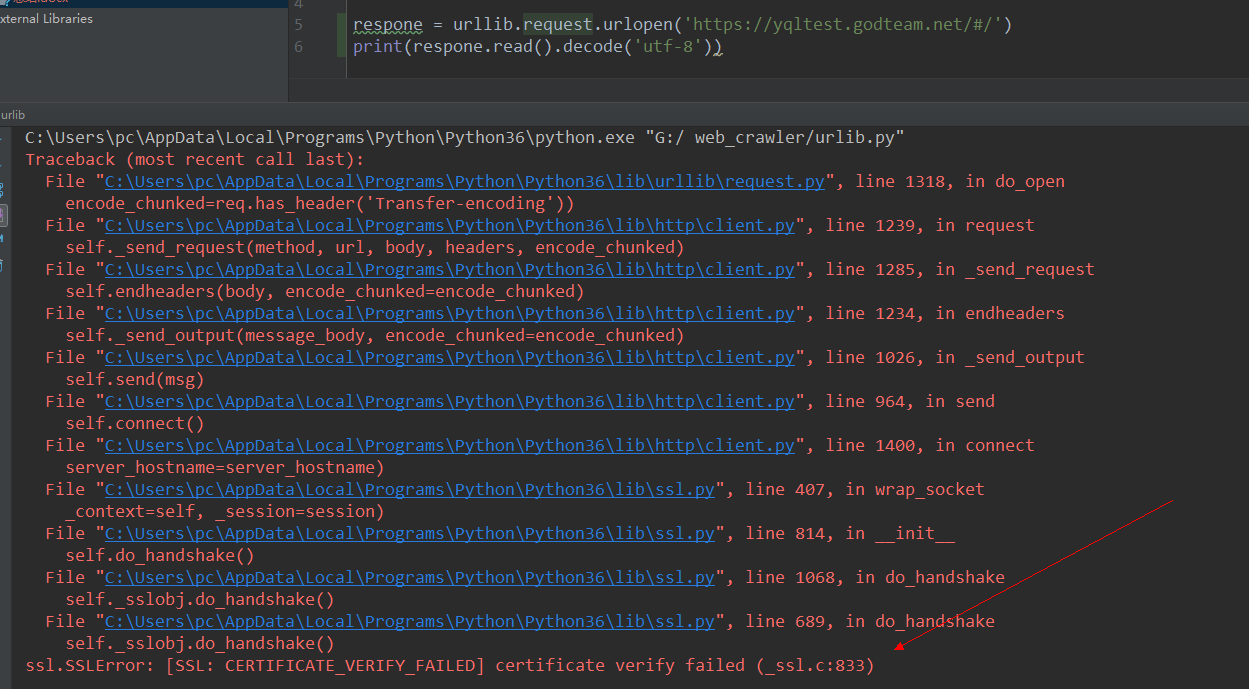
File "C:\Users\pc\AppData\Local\Programs\Python\Python36\lib\ssl.py", line 1068, in do\_handshake

self.\_sslobj.do\_handshake()

File "C:\Users\pc\AppData\Local\Programs\Python\Python36\lib\ssl.py", line 689, in do\_handshake

self.\_sslobj.do\_handshake()

ssl.SSLError: [SSL: CERTIFICATE\_VERIFY\_FAILED] certificate verify failed (\_ssl.c:833)



原因：python2.7升级到python3需要对ssl进行校验

解决方案：引入模块ssl取消全局验证

import ssl

ssl.\_create\_default\_https\_context = ssl.\_create\_unverified\_context

再使用urllib.urlopen('url')就没问题了

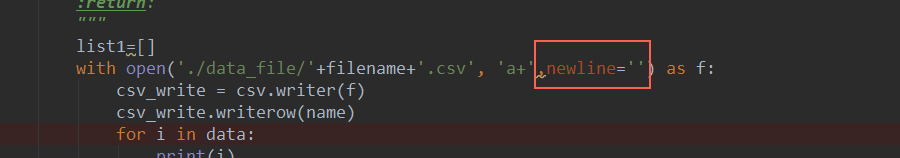
2、报错403，表示爬虫失败，被拦截了

urllib.error.HTTPError: HTTP Error 403: Forbidden

解决方案：添加user-agent,用浏览器F12查看请求头的user-agent

3、写入CSV文件会出现空白行；发现每隔一行会出现空白行

解决：添加newline=‘’参数



4、

Mongodb查询ObjectId数据类型时，需要导入

from bson.objectid import ObjectId

