目录

[测试环境 1](#_Toc466894741)

[安装软件 1](#_Toc466894742)

[QQSpider使用 2](#_Toc466894743)

[QQSpider调试 2](#_Toc466894744)

[运行报错分析 6](#_Toc466894745)

[其他说明 11](#_Toc466894746)

## 测试环境

系统： Windows8.1 64bit

内存： 4G

处理器： Intel® Core™ i5-4210U CPU @ 1.70GHz 2.39GHz

## 安装软件

1. 下载并安装python3

本人安装的 python-3.5.2.exe 版本

<https://www.python.org/downloads/windows/>

1. 安装python 编辑工具： pycharm

本人安装的Community pycharm-community-2016.2.3 版本

<http://www.jetbrains.com/pycharm/download/>

1. 安装数据库 MongoDB

本人安装的 mongodb-win32-x86\_64-2008plus-ssl-3.2.10-signed.msi 版本

<https://www.mongodb.com/>

1. 安装MongoDB数据库可视化工具：MongoBooster

本人安装的 mongobooster-3.1.3.exe 版本

<http://mongobooster.com/>

1. 安装PhantomJS

本人下载的 [phantomjs-2.1.1-windows.zip](https://bitbucket.org/ariya/phantomjs/downloads/phantomjs-2.1.1-windows.zip)  版本，解压缩后，把phantomjs.exe 复制到C:\Python\Python35-32\目录下

<http://phantomjs.org/>

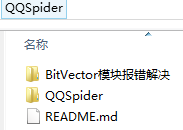
1. 下载QQSpider

<https://github.com/LiuXingMing/QQSpider>

1. QQSpider作者的CSDN blog: <http://blog.csdn.net/bone_ace>

## QQSpider使用

1. 解压缩下载的QQSpider-master.zip

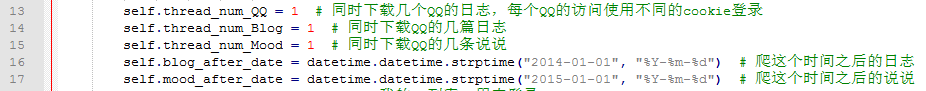


1. 更新myQQ.txt
   1. 打开该文件
   2. 输入自己的QQ号和密码，进入 myQQ.txt 写入QQ账号和密码（用一个空格隔开，不同QQ换行输入）

如：



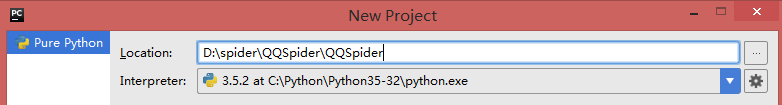
1. 进入 init\_messages.py 进行爬虫参数的配置，例如线程数量的多少、设置爬哪个时间段的日志，哪个时间段的说说，爬多少个说说备份一次等等。



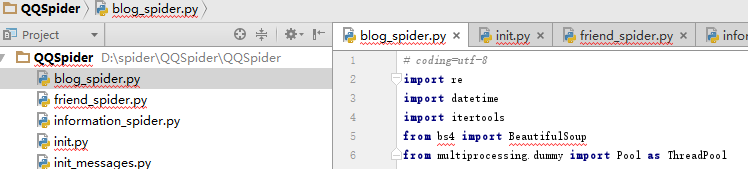
1. 运行 init.py 文件开启爬虫项目

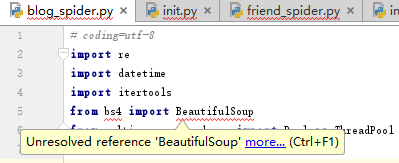
## QQSpider调试

1. 解压缩QQSpider-master.zip到 D:\spider\QQSpider\
2. 删除D:\spider\QQSpider\QQSpider\.idea
3. 打开PyCharm
4. 创建新的project，并引用已有的QQSpider（我解压缩到了D:\spider\QQSpider\QQSpider）



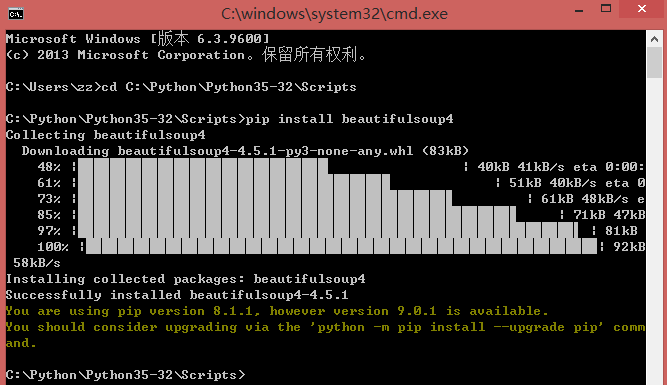
1. 会有很多的报错：
   1. BeautifulSoup



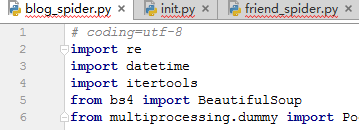


解决方法：

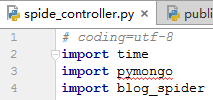
启动Windows dos窗口，并在C:\Python\Python35-32\Scripts目录下运行“pip install beautifulsoup4”



刷新后Beautifulsoup问题消失



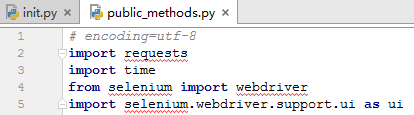
* 1. spide\_controller.py中缺少pymongo 模块



解决：

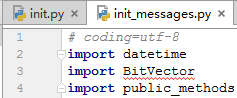
类似Beautifulsoup，运行“pip install pymongo”

* 1. public\_methonds.py 中缺少requests, selenium



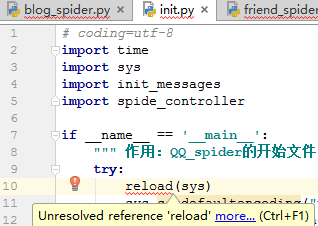
类似Beautifulsoup，运行“pip install requests” “pip install selenium”

* 1. init\_messages.py中缺少BitVector



类似Beautifulsoup，运行“pip install bitvector”

* 1. Init.py中reload函数报错



解决：

引入importlib包，并使用该包下的reload



* 1. print语句和except语句报错：



解决：

把print语句输出内容前后加小括号，如：

print time.ctime(), **"-Now, We start:\n\n"**

改为：

print (time.ctime(), **"-Now, We start:\n\n")**

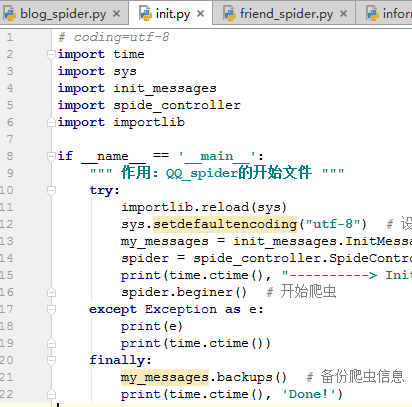
Exception 后的逗号改为 as， 如：

**except** Exception, e:

改为：

**except** Exception **as** e:

修改后效果：



同理修改 **except** KeyboardInterrupt, e:

## 运行报错分析

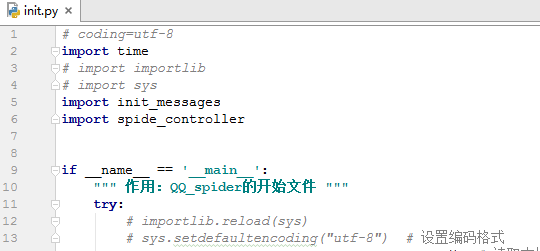
在pycharm中运行init.py

1. 错误：module 'sys' has no attribute 'setdefaultencoding'

|  |
| --- |
| C:\Python\Python35-32\python.exe D:/spider/QQSpider/QQSpider/init.py  module 'sys' has no attribute 'setdefaultencoding'  Traceback (most recent call last):  Fri Nov 11 12:27:54 2016  File "D:/spider/QQSpider/QQSpider/init.py", line 21, in <module>  my\_messages.backups() # 备份爬虫信息  NameError: name 'my\_messages' is not defined  Process finished with exit code 1 |

解决：

注释掉init.py的第3,4,12,13行

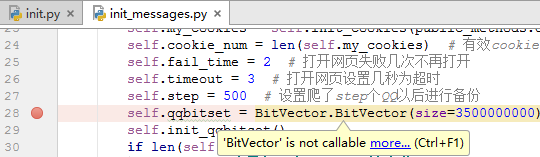


1. 错误： NameError: name 'my\_messages' is not defined

|  |
| --- |
| C:\Python\Python35-32\python.exe D:/spider/QQSpider/QQSpider/init.py  Traceback (most recent call last):  Fri Nov 11 12:51:15 2016  File "D:/spider/QQSpider/QQSpider/init.py", line 22, in <module>  my\_messages.backups() # 备份爬虫信息  NameError: name 'my\_messages' is not defined  Process finished with exit code 1 |

原因：

一般情况是BitVector不能使用，如果加入断点，可以看到执行到此行时会出现异常



异常信息为 MemoryError

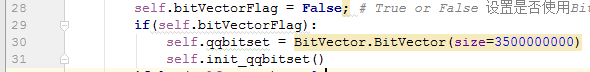


解决方法一：

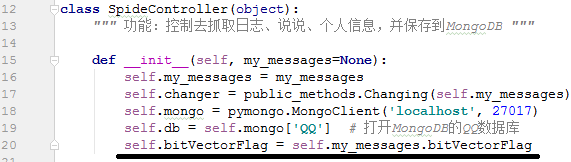
使用 “BitVector模块报错解决”文件夹中的文件替换默认的文件

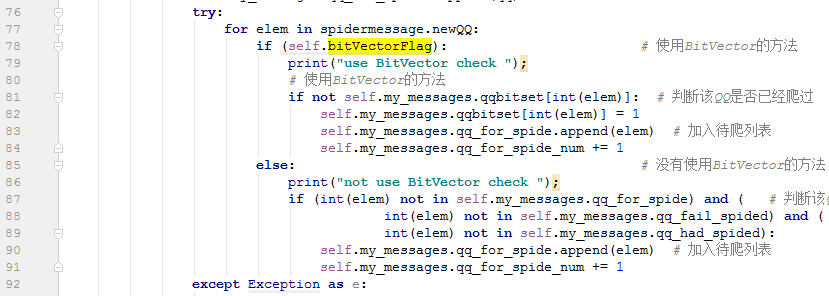
解决方法二：

* 1. 在init\_message.py中加入是否使用BitVector的控制



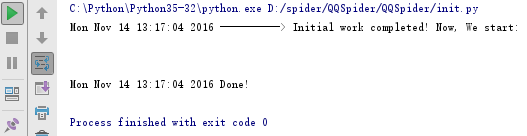
* 1. 修改spider\_controller.py





1. 没爬到结果

|  |
| --- |
| C:\Python\Python35-32\python.exe D:/spider/QQSpider/QQSpider/init.py  Mon Nov 14 13:17:04 2016 ----------> Initial work completed! Now, We start:  Mon Nov 14 13:17:04 2016 Done!  Process finished with exit code 0 |

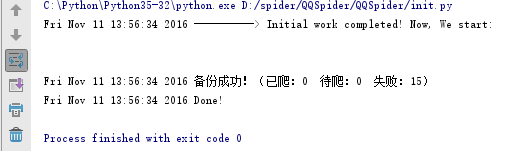


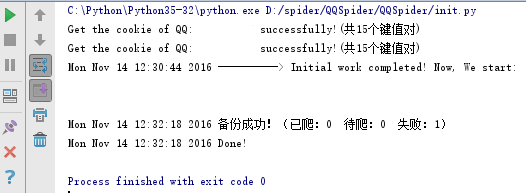
解决：

没显示“已爬”“待爬”“失败”信息，说明QQForSpider.txt为空

1. 没爬到结果

|  |
| --- |
| C:\Python\Python35-32\python.exe D:/spider/QQSpider/QQSpider/init.py  Fri Nov 11 13:56:34 2016 ----------> Initial work completed! Now, We start:  Fri Nov 11 13:56:34 2016 备份成功！（已爬：0 待爬：0 失败：15）  Fri Nov 11 13:56:34 2016 Done!  Process finished with exit code 0 |





解决：

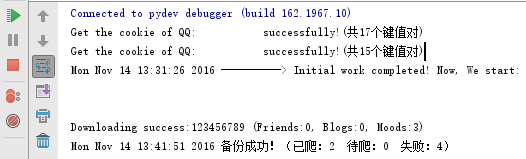
* 1. 只爬取一个qq号，如：123456789 。
  2. 直接打开qq对应的空间地址，检查是否可以访问，并检查是否有说说等内容可以下载

<http://user.qzone.qq.com/123456789>

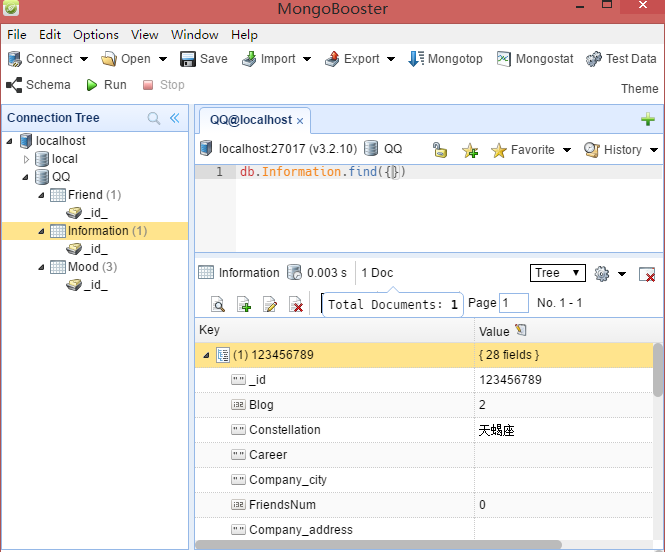


* 1. 检查MongoDB服务是否启动。确保服务启动

1. 程序正常运行起来后，会看到如下信息：

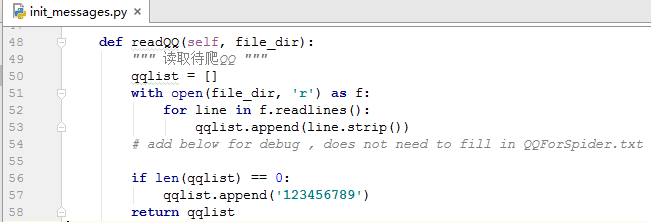


1. MongoBooster中打开MongoDB后，结果类似为：



## 其他说明

1. 为了每次调试时不需要在QQForSpider.txt中输入qq号，我在init\_messages.py中添加一个固定的QQ号：123456789。真正的代码中可以注释掉这几行



1. 本人联系方式

QQ 39679631

欢迎加入QSpider QQ群 563830972

