安装 框架

Pip3

twisted

第三方 Scrapy  
 pwin2

效果

Web前端

Html Css js

小爬虫 🡪 大的爬虫

后续

并行开发 fork & thread

python的正则

socket 编程

实现一个抓取代理服务器的小爬虫

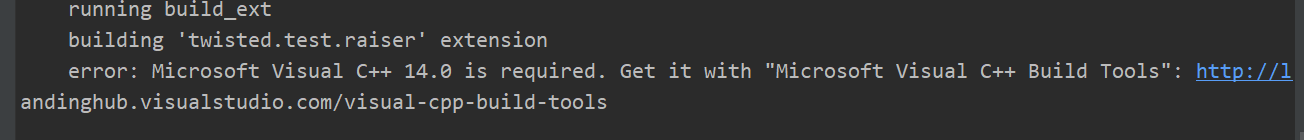
django开发web应用初探

django高级应用

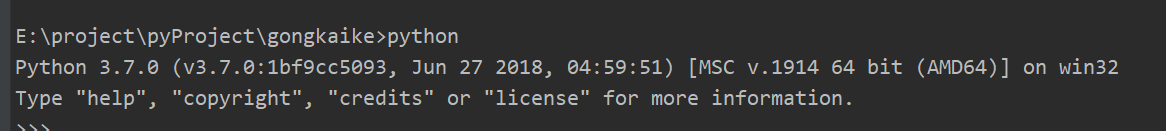
twisted 实现网络应用服务器

twisted的高级应用

python中的web应用



解决



<http://www.lfd.uci.edu/~gohlke/pythonlibs/>

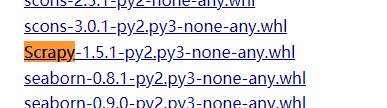
解决办法

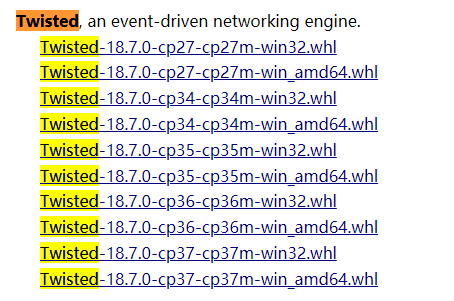
    在http://www.lfd.uci.edu/~gohlke/pythonlibs/有很多用于windows的编译好的Python第三方库，我们下载好对应自己Python版本的库即可。

    (1)在cmd中输入指令python，查看python的版本，如下：

    从上图可以看出可以看出我的Python版本为Python3.5.2-64bit。

    (2)登陆http://www.lfd.uci.edu/~gohlke/pythonlibs/，Ctrl+F搜索Lxml、Twisted、Scrapy，下载对应的版本，例如：lxml-3.7.3-cp35-cp35m-win\_adm64.whl，表示lxml的版本为3.7.3，对应的python版本为3.5-64bit。我下载的版本如下图所示：







    (3)在cmd中输入DOS指令，进入下载好的whl文件夹下，例如我的三个whl文件放在了Scrapy文件夹下：

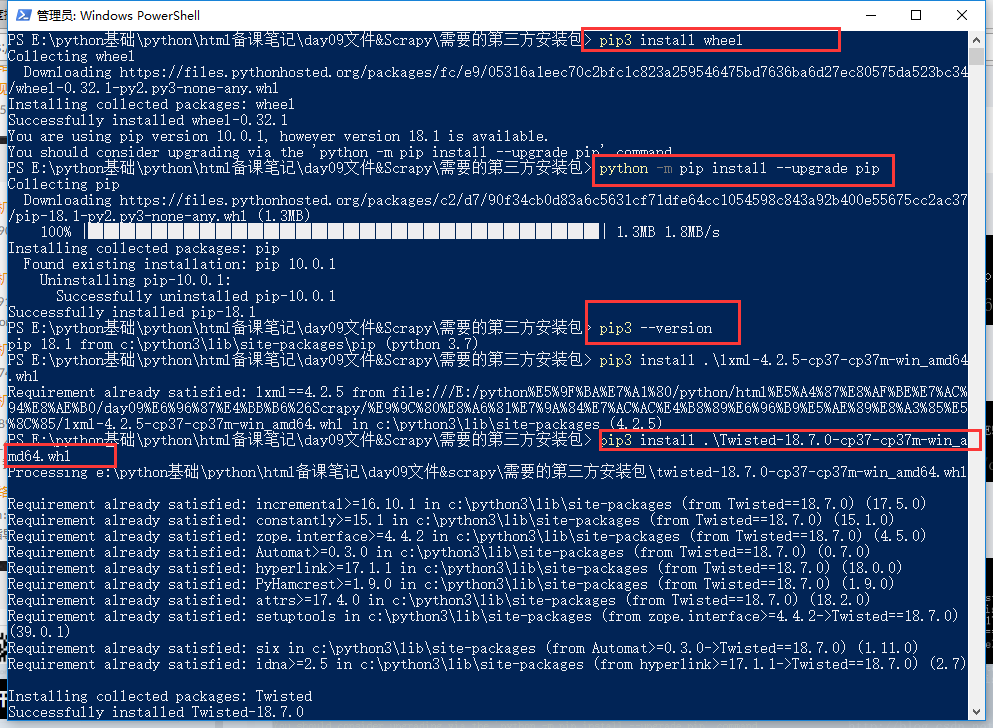
    (4)依次执行如下命令：

a.pip3 install wheel

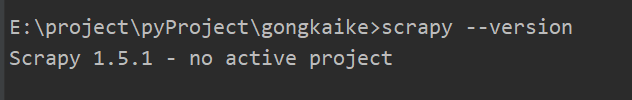
b.pip3 install lxml-3.7.3-cp35-cp35m-win\_amd64.whl

c.pip3 install Twisted-17.1.0-cp35-cp35m-win\_amd64.whl

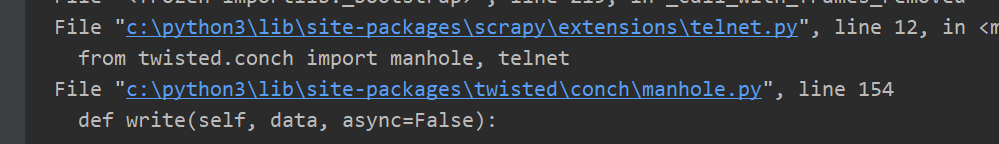
d.pip3 install Scrapy-1.3.2-py2.py3-none-any.whl



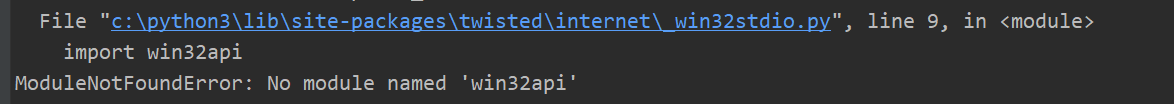
这样Scrapy的安装就完成了，请忽略最后两行让我升级pip的信息。\*.\*



错误:



这里async下面出现报错，因为await async成为了3.7的关键字，不能作为变量名使用。   
进入manhole.py将async全改成async1就阔以了。



python -m pip install pypiwin32

1:测试 : scrapy

2:抓百度的

3:创建项目

4:介绍项目目录

5:创建爬虫网站 cd spider : 执行scrapy genspider 爬虫名 “网站地址”

6:此时命令多了很多

