—————————————————windows下程序使用指南————————

1.将.py程序打包成exe文件具体方法见http://bbs.fishc.com/thread-50485-1-1.html

2.所有的程序一定要在数据库mongodb运行的情况才能使用。

开启数据库方法：切换到数据库的安装位置，比如在windows下的切换到数数据库的安装路径：

C:\Program Files\MongoDB\Server\3.0\bin

然后输入命令行mongod就可以开启数据库的服务器

3.现在开始运行爬虫程序：首先在命令窗口切换到爬虫xywy\_crawl的所在位置例如：

C:\Users\wing1995\Desktop\xywy\_crawl

然后输入命令行scrapy crawl xywy\_spider即可，随时都可以关掉命令窗口终止程序，数据

会安全的进入数据库，如果想要在任何时刻终止或者继续原来没有下载完的数据，可以在开

始的时候输入命令行：

scrapy crawl xywy\_spider -s JOBDIR-crawls/xywy\_spider-1（下次继续下载输入相同的

命令行即可）

这样的好处是可以随时下载任意多的数据，而且接着上次未完成的任务继续下载，具体情况

可以参照scrapy官方文档。

4.所需要的工具版本号：

一、scrapy框架的程序版本号：

1.python 2.7.1 (在运行scrapy的时候使用这个版本)

2.scrapy 2.4 (实际上直接pip就可以了)

二、界面程序pymongo\_to\_db.py程序所需要的版本

1.python 3.4.2 (由于py2版本不支持中文输出，所以数据库的查询界面只能在py3版本中

运行)

2.在python34的路径下pip install pymongo这个外带的模块，版本号电脑会自动匹配

三、mongodb版本号：mongodb 3.0.0

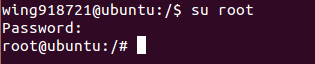
5.有不懂的知识点问google，对程序有意见或者想法的跟我说就是。

—————————————————linux下使用指南———————————————

一、安装mongodb

温馨提示：在这个网站的指示下安装mongodb(都要切换到root权限下才可以):

切换root用户参考这个网站：<http://www.cnblogs.com/clzqwdy/archive/2008/12/05/1348368.html>，

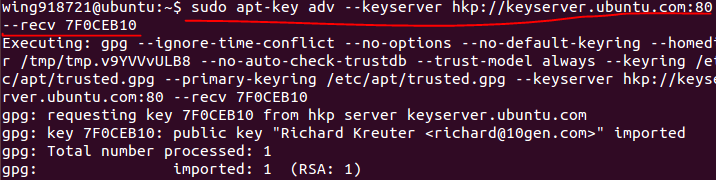


或者直接sudo命令

举例：我的电脑是ubuntu14的64位,系统给我安装了2.4版本的mongodb

【配置APT】1.使用下面的命令导入10gen的公共GPG密钥：

sudo apt-key adv --keyserver keyserver.ubuntu.com --recv 7F0CEB10

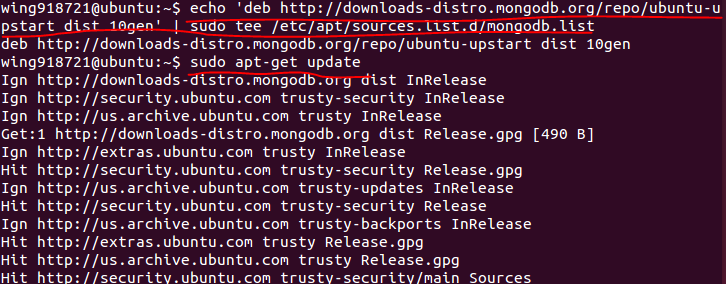


2.创建 10gen.list文件：

echo 'deb http://downloads-distro.mongodb.org/repo/ubuntu-upstart dist 10gen' | sudo tee /etc/apt/sources.list.d/10gen.list

3.更新资料源：

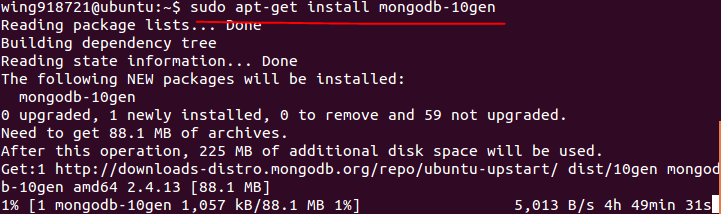
sudo apt-get update



【安装mongodb】

1.使用的是apt-get，因为。。。。。。方便！

sudo apt-get install mongodb-10gen

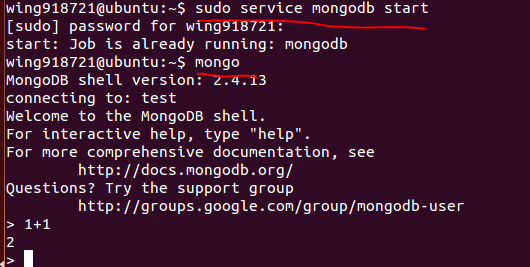


网速慢是硬伤。。。。

2.安装完后，自动就启动了，输入pstree 查一下，应该可以找到mongod

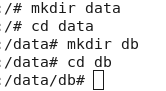
3.测试

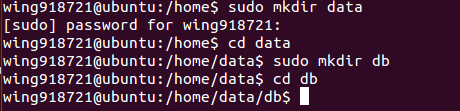
首先输入mongo进入数据库，输入命令行

没问题的话就退出‘Ctrl + c’到终端

4.创建数据库目录

MongoDB的数据存储在data目录的db目录下，但是这个目录在安装过程不会自动创建，所以你需要手动创建data目录，并在data目录中创建db目录。（以下是命令行建立目录）



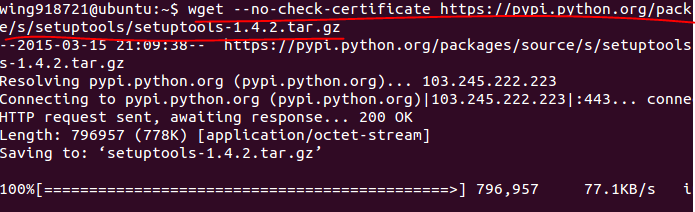


5.输入 sudo mongod启动服务器，接下来的程序都要在数据库服务器启动的情况下执行

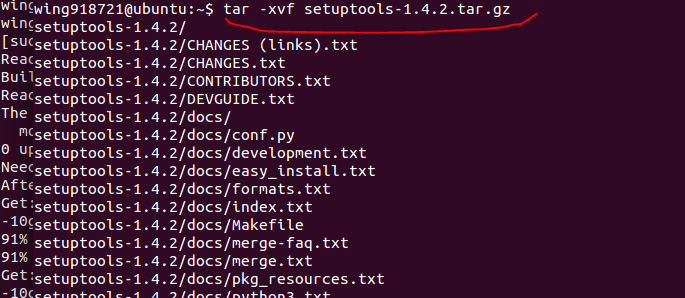
二、安装python的模块

### 【安装pip】（根据不同的版本python选择安装不同版本的工具，系统自带的python版本是2.7.6）

1. 下载setuptools



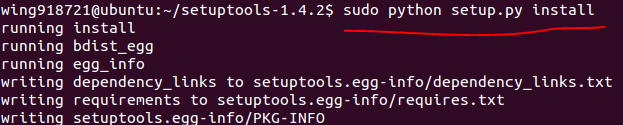
1. 解压setuptools



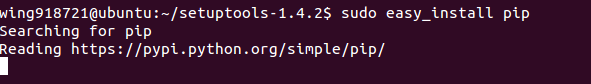
1. 进入setup目录



1. 安装setuptools



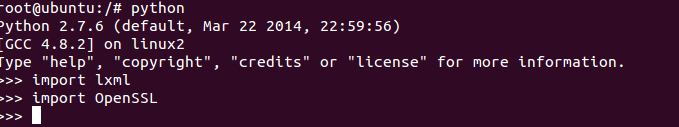
1. 用setuptools安装pip



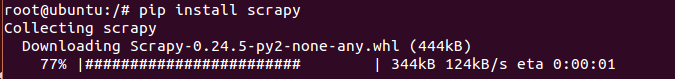
【安装scrapy】

1.搭建scrapy环境需要openssl 和lxml

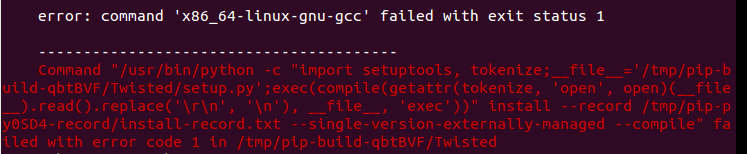
检测一下是否已经安装好



由于ubuntu自己已经安装好了这些，因此直接安装scrapy即可

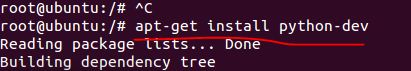


But…

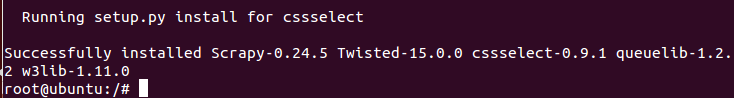


没有安装python-dev导致的结果。

2.安装python-dev



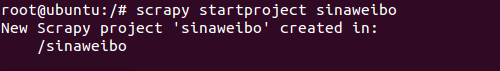
3.ok,再次安装scrapy并检验。



以防万一，再安装一个模块



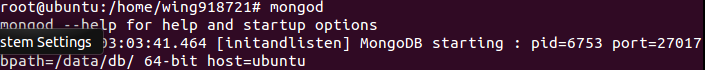
检验一下，创建一个爬虫工程



ok!

【开始爬虫吧】

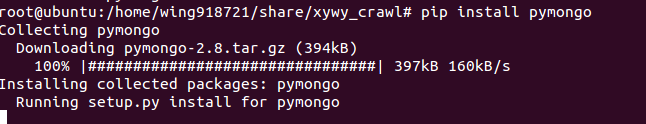
1. 所有的程序一定要在数据库mongodb运行的情况才能使用。



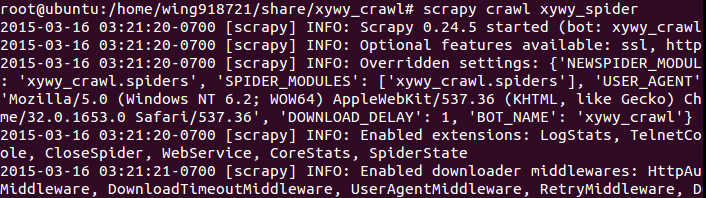
1. 现在开始运行爬虫程序：首先在命令窗口切换到爬虫xywy\_crawl的所在位置



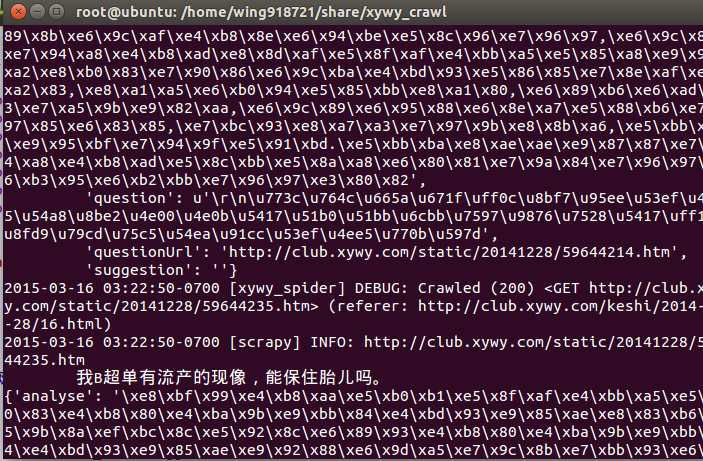
1. 安装没有安装的模块pymongo



4\、开始爬虫



1. 终端输出结果如下（事实证明网速越快，数据下载越多）



1. 由于在python2版本中不支持中文，因此输出的内容不是中文而是UNIcode码，因此检验再安装一次python3版本，为了更好的选择不同版本的python， 选择一样好的编辑器pycharm!

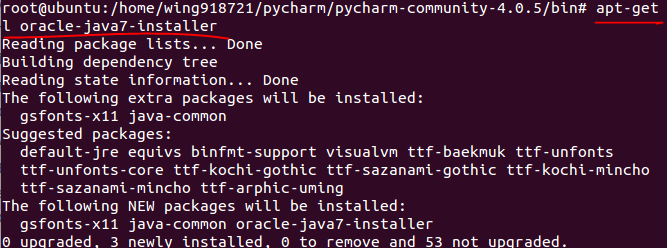
【安装pycharm】

进入官网下载安装包

1.从http://www.jetbrains.com/pycharm/download/下载最新的Pycharm

2.在终端中进入Pycharm的下载目录，输入命令 tar xfz pycharm-\*.tar.gz，将Pycharm文件夹解压到当前目录

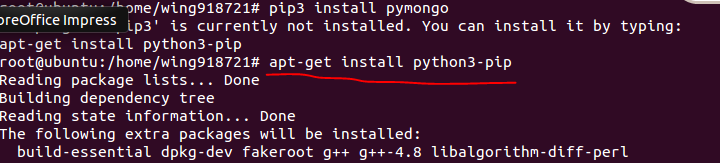
3.配置jdk1.7环境。【考虑到你肯定安装了java，因此大概飘过】

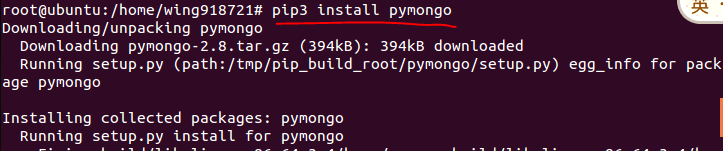


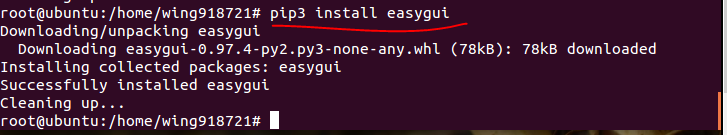
4、具体的pycharm安装教程如下网站<http://jingyan.baidu.com/article/7f766daf5e20944101e1d02b.html>

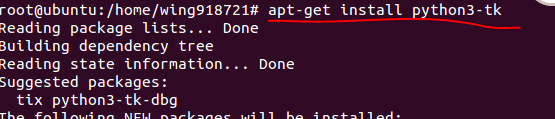
【搭建python3.4中的环境】主要ubuntu14自带了python3.4和2.7版本的

1. 安装pip,easygui,pymongo以及tkinter







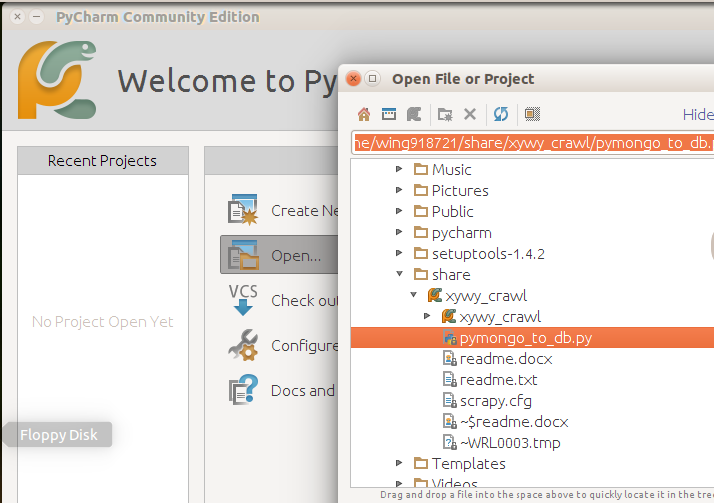


2、在pycharm中运行程序

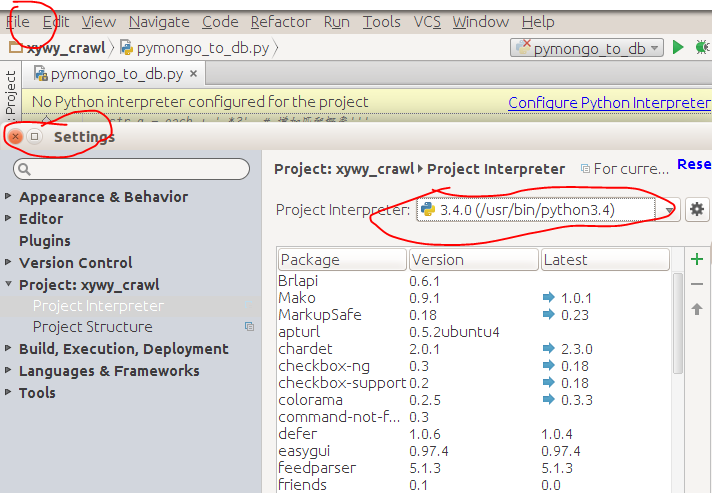
由于pymongo-db.py这个程序需要运行的环境是在py3中，因此先配置环境，打开pycharm，



1. 选择文件



1. 配置环境



运行！！成功

