  因为想自学爬虫，对 “Python那些事” 的公众号中的一篇推送《Python爬虫实战：爬取全站小说排行榜》(链接：<http://mp.weixin.qq.com/s/Qdp3qfZI-rMtEWm5fjXlvw>）进行研读，加上自己的百度知识点来促进自己理解。本文章可能会侵犯到一些博主或者其他作者的知识产权，当然以下内容涉及到的来源我也会备注清楚来源，如若无意中存在侵权行为，望大佬们海涵 ^-^ ！！！

一、python环境安装

python安装及pycharm编辑器：<http://blog.csdn.net/qq_29883591/article/details/52664478>

具体怎么安装python环境我就不赘述了，看以上链接即可。

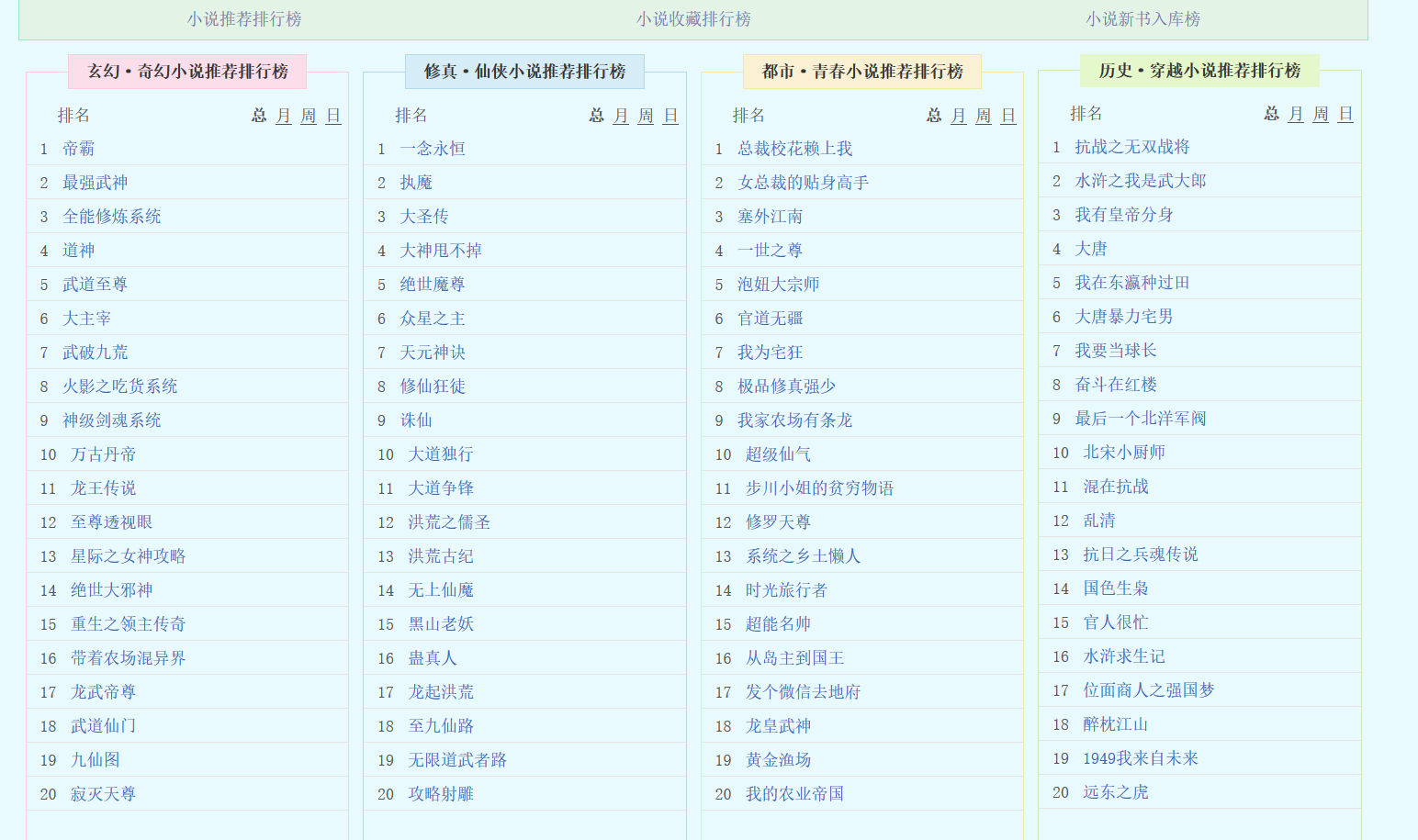
根据公众号给的代码，我模仿者自己重新编写一次，并记录下在这个过程出现的问题。

我的环境： win10、python 3.6.5版本 32位、pycharm 2017.3.1、

二、分析爬取的新笔趣阁网站（<http://www.bxquge.com/paihangbang/>）

1、网站界面：

1）链接：http://www.bxquge.com/paihangbang/



 2）链接：[http://www.bxquge.com/3\_3619/](http://www.bxquge.com/3_3619/#)

3）链接：可以发现其中的规律

楔子：<http://www.bxquge.com/3_3619/10825.html>

第一章：<http://www.bxquge.com/3_3619/10826.html>

第二章：http://www.bxquge.com/3\_3619/10827.html



三、代码分析

创建新的项目,我创建的新项目名为NovelCrawl，公众号的项目文件名为小说【1】（有兴趣可以下载该公众号代码）。

1. item.py

#-\*-coding:utf-8-\*-  
  
#问题1：python3引进scrapy引入不成功？  
import scrapy  
  
  
#问题2：这里的item怎么理解？  
class NovelCrawlItem(scrapy.item):  
 list\_name=scrapy.Field()  
 name=scrapy.Field()  
 title=scrapy.Field()  
 link=scrapy.Field()  
 words=scrapy.Field()

问题1：在File-> Settings-> Project Interpreter中引入scrapy包会报错，是依赖包方面的问题。

解决：手动安装scrapy并引用到虚拟环境

1）Scrapy安装（参考了CSDN博客[Jack-Cui](http://my.csdn.net/c406495762" \t "_blank)的博文【2】）

①直接使用指令pip3 install scrapy，发现有诸多错误。

Failed building wheel for lxml

Microsoft Visual C++ 10.0 is required

Failed building twisted

Unable to find vcvarsall.bat

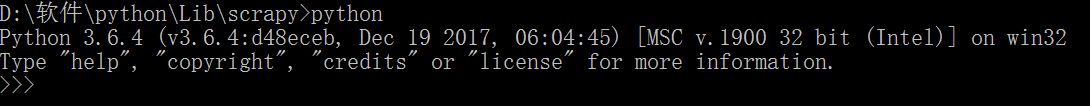
遇到的错误，如下图所示：



②方法：

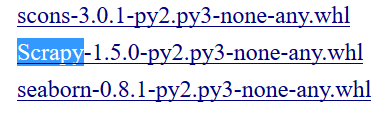
在<https://www.lfd.uci.edu/~gohlke/pythonlibs/>有很多用于windows的编译好的Python第三方库，我们下载好对应自己Python版本的库即可。

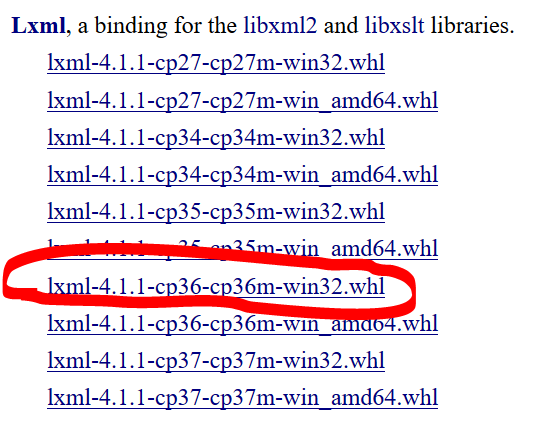
A、在cmd中输入指令python，查看python的版本，如下：

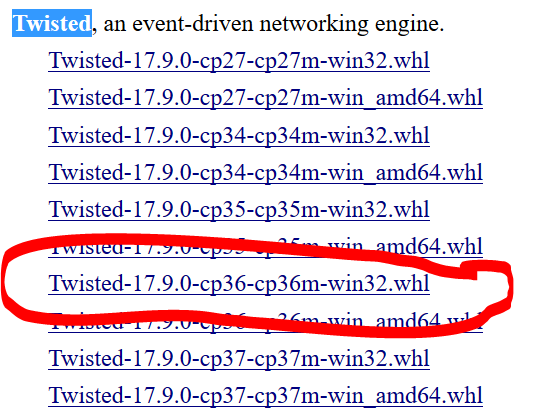


从上图可以看出可以看出我的Python版本为Python3.6.4-32bit。

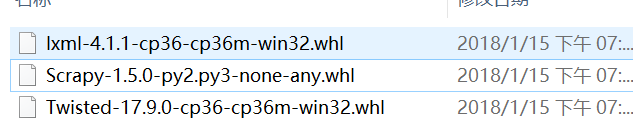
③登陆<https://www.lfd.uci.edu/~gohlke/pythonlibs/>，Ctrl+F搜索Lxml、Twisted、Scrapy，下载对应的版本。cp36代表python3.6，amd64代表python64位



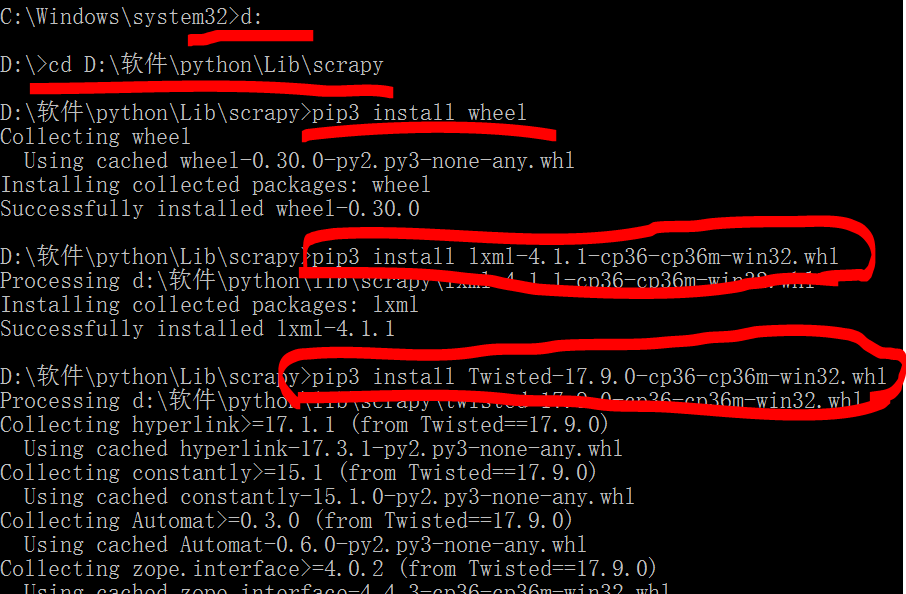


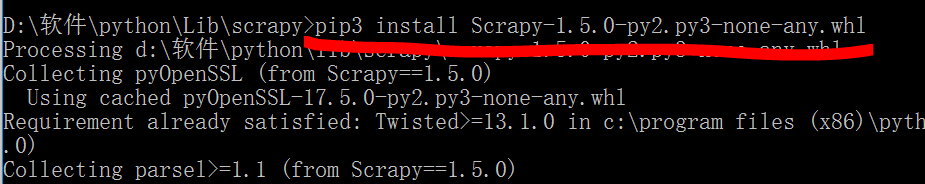


为了方便管理，我创建了该路径文件D:\软件\python\Lib\scrapy，Lib用来存放包。我下载的版本如下，



④在cmd中输入DOS指令，进入D:\软件\python\Lib\scrapy文件夹下，例如我的三个whl文件放在了Scrapy文件夹下，依次执行如下命令：



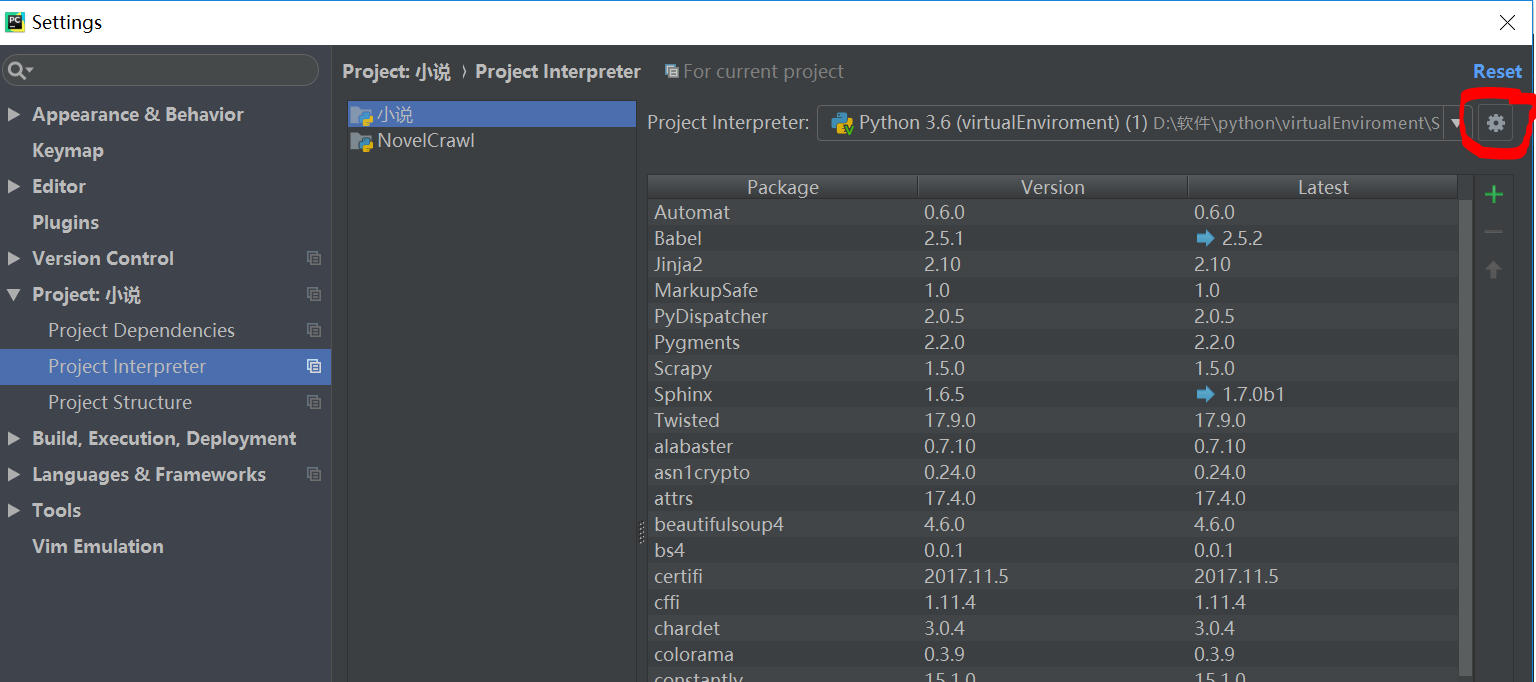


都看到successfully代表都成功了，这样Scrapy的安装就完成了。

⑤Srapy已经安装成功，源博客写需要安装Pywin32,安装出现一些问题，我就没安装，但也不影响一下后面步骤。所以我没去管它了。

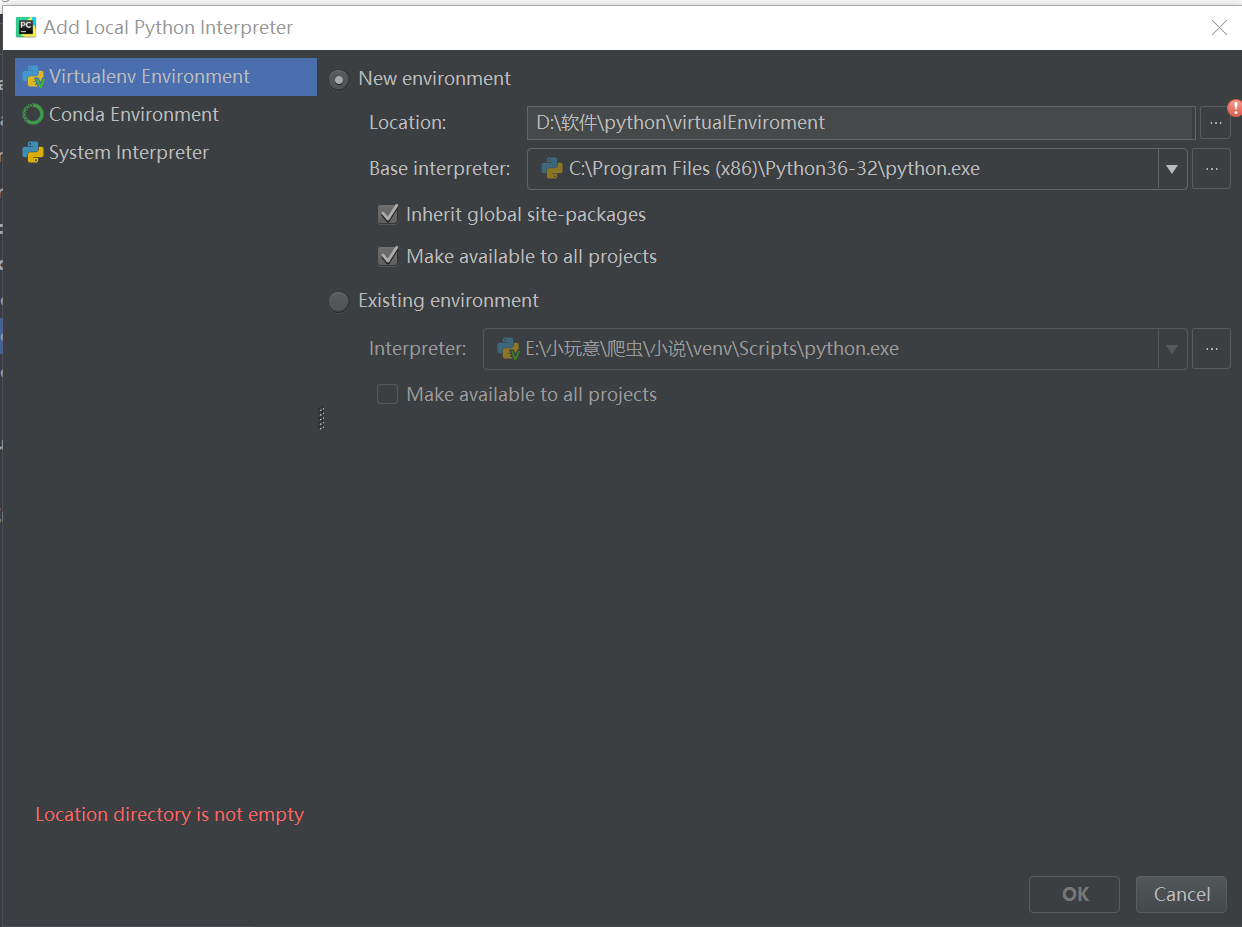
2）创建一个虚拟环境，点击Files-> Settings-> Project Interpreter中的配置icon-> 点击选Add Local Python Interpreter

（虚拟环境：你跑一个项目需要的包可能不同，虚拟环境的好处是根据你的需要配置当前项目需要的包，使得不同项目的不同运行环境区分开来，等等好处，不懂自个百度吧！）

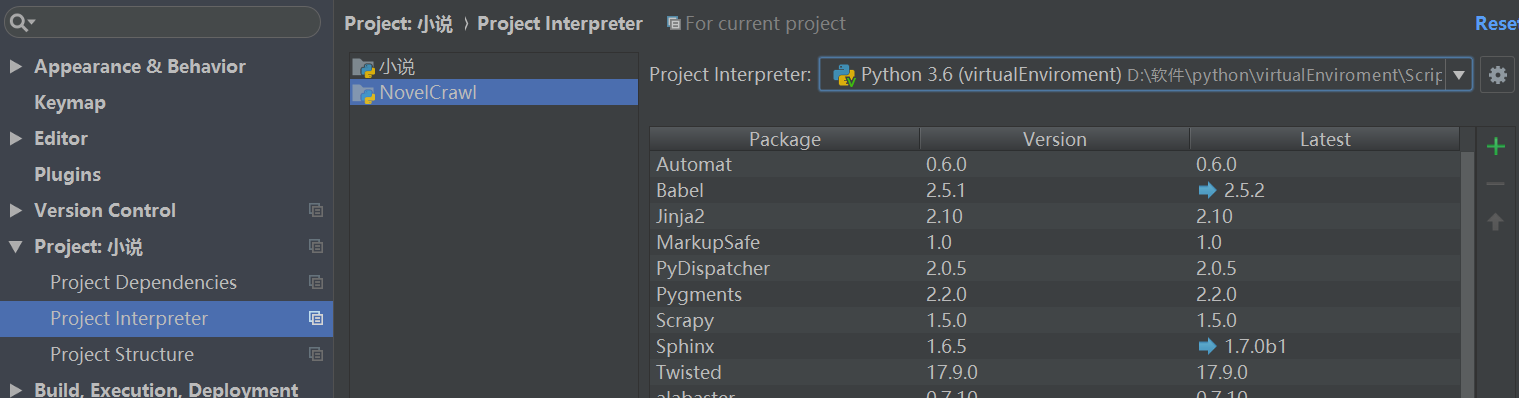


3）因为我要创建的虚拟环境在D:\软件\python 下，所以我现在在此文件夹下已经先创建一个名为virtualEnviroment的空文件夹。

选择New environment, 其中Location选择刚才创建的新文件夹virtualEnviroment，因为我已经在此文件夹下创建了虚拟环境，所以会报错Location directory is not empty.；Base interpreter选择你python安装路径下的python.exe；两个勾按你需求打即可。然后点击OK即可。



4）创建完需要的虚拟环境后，选择 Files-> Settings-> Project Interpreter，选中你的项目，将Project Interpreter设置为你刚才创建的虚拟环境。设置完你会发现你创建的虚拟环境已经自动引用了scrapy包，我也不是很清楚为什么，好像是上面安装scrapy是全局python环境，而不是只应用到某个虚拟环境，因为通过pycharm创建虚拟环境，然后引入包，该包只能在此虚拟环境下才能用，在别的虚拟环境就没办法调用。



问题2：对scrapy.item的理解。

解决方法：可查看<http://blog.csdn.net/sdulsj/article/details/52984619>

1. mongoReading.py

参考链接：

【1】公众号“Python那些事”提供的源代码

           百度云盘地址为：https://pan.baidu.com/s/1slkeYoH，链接为：jz2e。

【2】python3安装scrapy：作者[Jack-Cui](http://my.csdn.net/c406495762) 博文<http://blog.csdn.net/c406495762/article/details/60156205>