若用豆瓣元安装报错则去[www.lfd.uci.edu/~gohlke/pythonlibs/](http://www.lfd.uci.edu/~gohlke/pythonlibs/)找对应python版本的包下下来然后进去下的路径使用workon环境然后pip install 这个下的文件，然后再总的重新安装

Scrapy的学习

Spider

Item

Item loader

Pipeline

Feed export

CrawlSpider

防止被网站禁止：

1. 图片验证码
2. Ip访问频率限制
3. User-agent随机切换

Scrapy进阶 – 1、scrapy的原理

2、基于scrapy的中间件开发

1、动态网站的抓取处理

2、将selenuim 和 phantomjs集成到scrapy

3、scrapy log 配置

4、email发送

5、scrapy信号

之后基于scrapy二次开发比如redis-分为几个服务器进行分布式爬虫（这里面集成bloomfilter到scrapy-redis中）

Elasticsearch+django实现一个搜索网站

具体的是、

数据库：mysql、redis、elasticsearch

开发环境：virtualenv

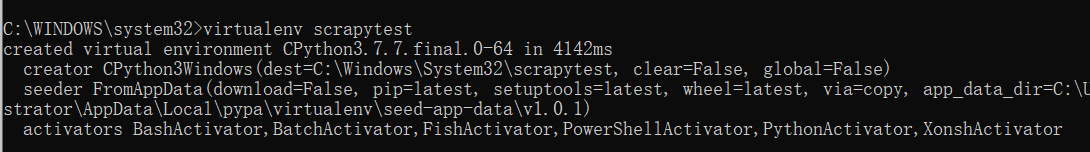
一大点：virtualenv和virtualenvwrapper安装和配置

好处：将我们的开发环境隔离并不影响

-i <https://pypi.douban.com/simple/>

下面是它的使用

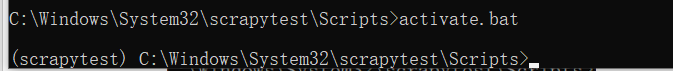
virtualenv scrapytest将虚拟环境新建到当前的目录下面



Then加入来python



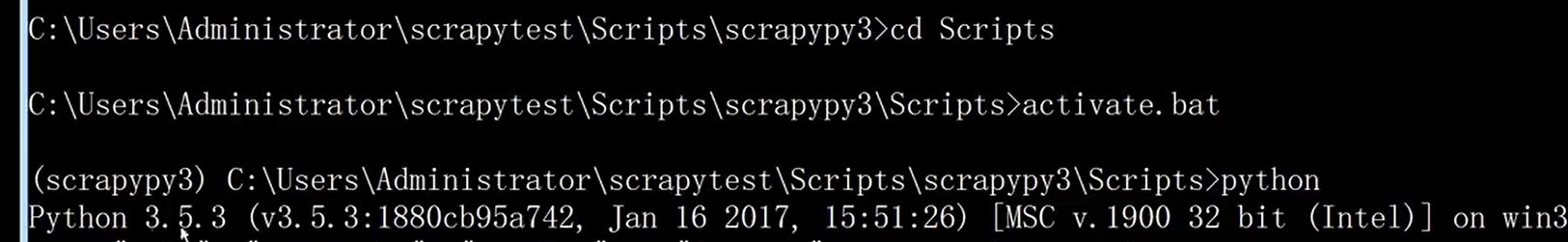
然后bat，加入此环境



deactivate.bat为退出

加注virtualenv可以指定python版本进入python的安装环境



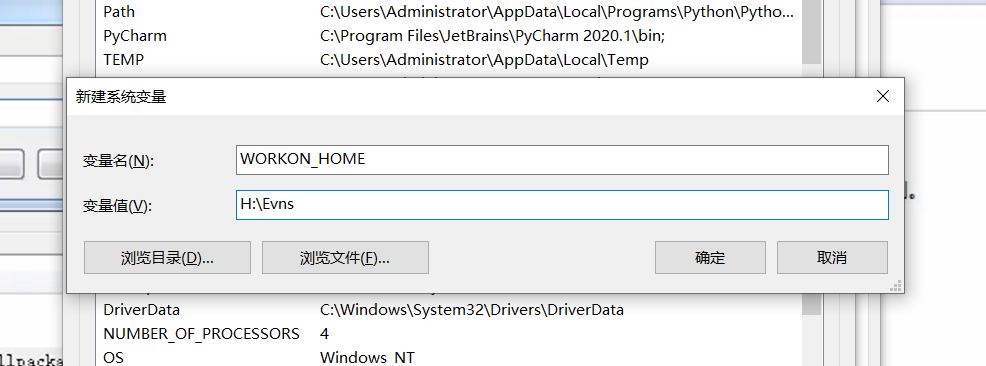


如不换上面的版本则先deactivate.bat后再装virtualenvwrapper（注意装这个的时候一定要退出当前目录装，若安装失败则pip3 install virtualenvwrapper-win -i https://pypi.douban.com/simple/）



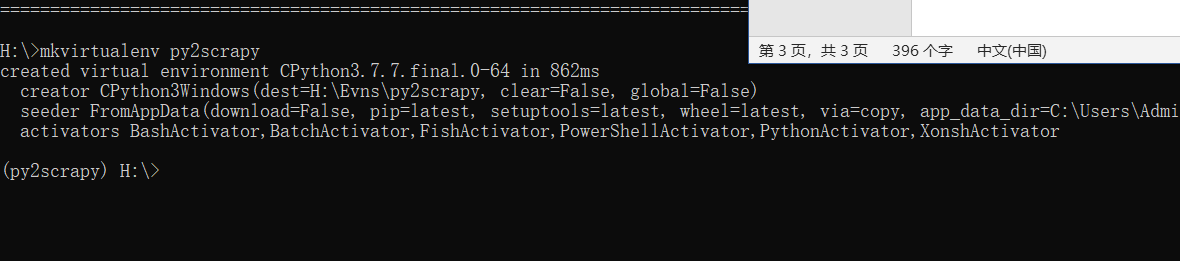
之后用的workon的命令可以配置环境变量

先再e盘新建evns文件夹用

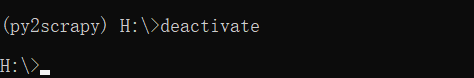


之后

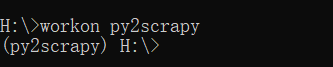
mkvirtualenv py2scrapy创建在envs下



之后若想要不激活就直接deactivate



之后要进去的时候就直接workon scrapxxxx文件包进去就行了



然后在里面装scrapy

若要指定python版本新建

mkvirtualenv --python=就是python那个路径 文件名