

# 豆果爬虫Windows环境搭建

---

## 安装需要的程序包

---

- Python3.4.3 > <https://pan.baidu.com/s/1pK8KDcv>
- pip9.0.1 > <https://pan.baidu.com/s/1mhNdRN6>
- 编辑器pycharm > <https://pan.baidu.com/s/1i4Nkdk5>
- pywin32 > <http://pan.baidu.com/s/1pKZiZWZ>
- pyOpenSSL > <http://pan.baidu.com/s/1hsgOQJq>
- windows\_sdk > <http://pan.baidu.com/s/1hrM6iRa>
- phantomjs > <http://pan.baidu.com/s/1nvHm5AD>

## 安装过程

---

### 安装基础环境

1. 安装Python安装包，一路Next
2. 将Python的安装目录添加到环境变量Path中
3. win + r 输入Cmd打开命令行窗口，输入Python 测试是否安装成功

### 安装pip

pip的作用相当于linux的yum，安装之后可以采用命令行的方式在线安装一些依赖包

1. 解压pip压缩包到某一目录（推荐与Python基础环境目录同级）
2. cmd窗口进入pip解压目录
3. 输入 python [setup.py](#) install 进行安装，安装过程中将会在Python目录的scripts目录下进行
4. 将pip的安装目录 C:\Python34\Scripts; 配置到环境变量path中
5. cmd命令行输入pip list 或者 pip --version 进行检验

## 安装Scrapy

Scrapy是一个比较成熟的爬虫框架，使用它可以进行网页内容的抓取，但是对于windows并不友好，我们需要一些类库去支持它

1. 安装pywin32: 一路next即可
2. 安装wheel: 安装scrapy时需要一些whl文件的安装，whl文件的安装需要预先配置wheel文件。在cmd下使用pip安装：`pip install wheel`
3. 安装PyOpenSSL: 下载完成PyOpenSSL后，进入下载所在目录，执行安装：`pip install pyOpenSSL` (注意，执行安装的wheel文件名一定要tab键自动弹出，不要手动敲入)
4. 安装lxml: 直接使用pip在线安装 `pip install lxml`

在Windows的安装过程中，一定会出现“error: Microsoft Visual C++ 10.0 is required (Unable to find vcvarsall.bat).”的问题，也就是无法找到相对应的编译包。一般的做法是下载VisualStudio来获得Compiler，但是我们不这样做。

下载windows-sdk后，执行安装操作，如果安装成功，那么这个问题就解决了。如果失败，那么需要先把安装失败过程中的2个编译包卸载。他们分别为：Microsoft Visual C++ 2010 x86 Redistributable、Microsoft Visual C++ 2010 x64 Redistributable（可以使用360或者腾讯管家来卸载）

卸载完成之后，在安装确认过程中，不要勾选Visual C++ compiler，这样他第一次就能安装成功。安装成功之后，再次点击sdk进行安装，这时候又需要把Visual C++ compiler勾选上，再次执行安装。完成以上操作后，就不会出现Microsoft Visual C++ 10.0 is required的问题了。

如果在安装过程中出现“failed building wheel for xxx”的问题，那么需要手动下载wheel包进行安装，所有的安装文件都可以在<http://www.lfd.uci.edu/~gohlke/pythonlibs/>里找到，找到需要的包并下载完成后执行`pip install xxxx`即可。

5. 安装Scrapy: `pip install Scrapy`, 安装完成后可以再命令行窗口输入Scrapy进行验证。