

要使用 Scrapy，首先需要安装 Python3 和 Scrapy，然后即可使用，下面介绍 Python3 和 Scrapy 的安装。

一、Python3的安装

既然要用 Python3 开发爬虫，那么第一步一定是安装 Python3，本节会介绍 Windows、Linux、Mac 三大平台下安装 Python3 的过程。

1. 相关链接

- 官方网站: <http://python.org>
- 下载地址: <https://www.python.org/downloads>
- 第三方库: <https://pypi.python.org/pypi>
- 官方文档: <https://docs.python.org/3>
- 中文教程: <http://www.runoob.com/python3/python3-tutorial.html>
- Awesome Python: <https://github.com/vinta/awesome-python>
- Awesome Python 中文版: <https://github.com/jobbole/awesome-python-cn>

2. Windows下的安装

Windows 下安装 Python3 的方式有两种，一种是通过 Anaconda 安装，Anaconda 提供了 Python 的科学计算环境，里面自带了 Python 以及常用的库，如果选用了此种方式后面的环境配置方式会更加简便，另一种是直接下载安装包安装，即标准的安装方式。下面会依次介绍这两种安装方式，任选其一即可。

Anaconda安装

Anaconda的 官方下载链接为: <https://www.continuum.io/downloads>, 选择 Python3 版本的安装包下载即可，如图所示：

Download for Your Preferred Platform



Anaconda 4.4.0 For Windows Graphical Installer

Python 3.6 version *

64-Bit (437 MB) ②

DOWNLOAD

Download 32-bit (362 MB)

Python 2.7 version *

64-Bit (430 MB) ②

DOWNLOAD

Download 32-bit (354 MB)

如果下载速度过慢可以选择使用清华大学镜像，下载列表链接

为：<https://mirrors.tuna.tsinghua.edu.cn/anaconda/archive/>，使用说明链接

为：<https://mirrors.tuna.tsinghua.edu.cn/help/anaconda/>，可以选择需要的版本进行下载，速度相比官网会快很多。

下载完成之后直接双击运行安装包安装即可，安装完成之后Python3的环境就配置好了。

安装包安装

推荐直接下载可执行文件安装包安装，到官方网站下载Python3安装包。

<https://www.python.org/downloads/>

到目前为止，Python 的最新版本是 3.6.2，最新版本以官网为准，如 Python3.6.2 的下载链接为：<https://www.python.org/downloads/release/python-362/>，下载页面如图所示：

Files					
Version	Operating System	Description	MD5 Sum	File Size	PGP
Gzipped source tarball	Source release		2d0fc9f3a5940707590e07f03ecb08b9	22540566	SIG
XZ compressed source tarball	Source release		692b4fc3a2ba0d54d1495d4ead5b0b5c	16872064	SIG
Mac OS X 64-bit/32-bit installer	Mac OS X	for Mac OS X 10.6 and later	6dd08e7027d2a1b3a2c02cfacbe611ef	27511848	SIG
Windows help file	Windows		69082441d723060fb333dca8815105e	7986690	SIG
Windows x86-64 embeddable zip file	Windows	for AMD64/EM64T/x64, not Itanium processors	708496ebbe9a730d19d5d288afd216f1	6926999	SIG
Windows x86-64 executable installer	Windows	for AMD64/EM64T/x64, not Itanium processors	ad69fdacde90f2ce8286c279b11ca188	31392272	SIG
Windows x86-64 web-based installer	Windows	for AMD64/EM64T/x64, not Itanium processors	a055a1a0e938e74c712a1c495261ae6c	1312520	SIG
Windows x86 embeddable zip file	Windows		8dff09a1b19b7a7dcb915765328484cf	6320763	SIG
Windows x86 executable installer	Windows		3773db079c173bd6d8a631896c72a88f	30453192	SIG
Windows x86 web-based installer	Windows		f58f019335f39e0b45a0ae68027888d7	1287064	SIG

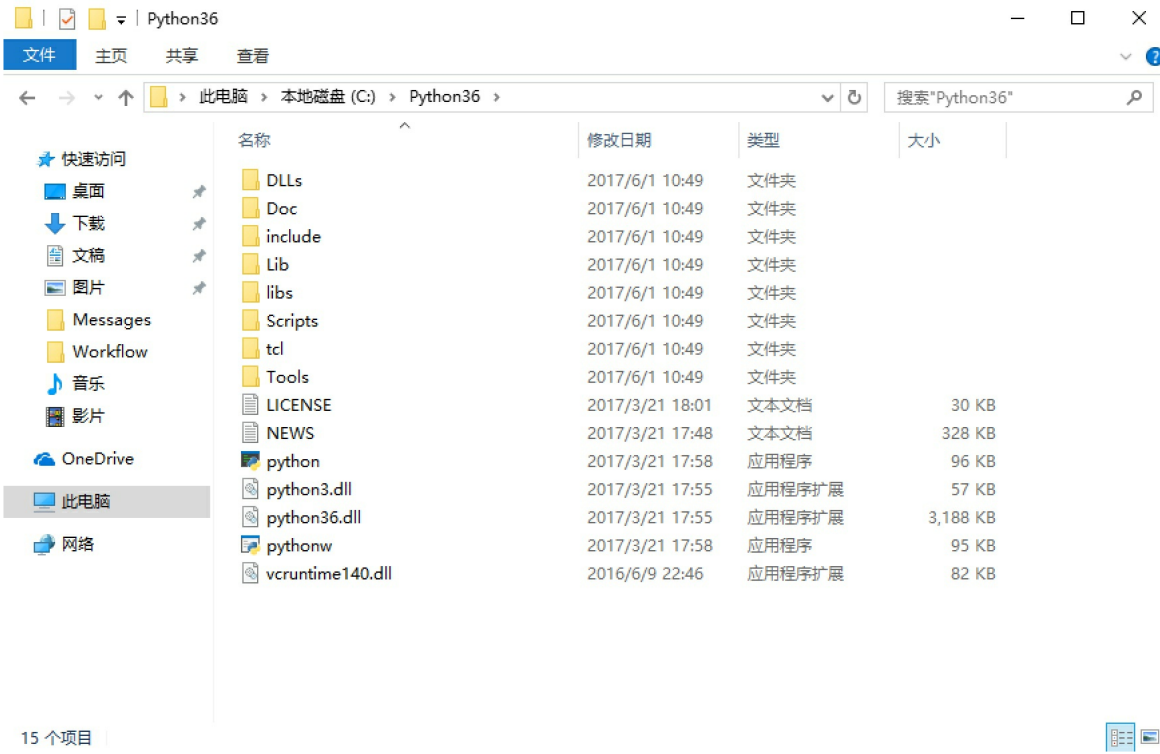
64 位系统可以下载 Windows x86-64 executable installer，32 位系统可以下载 Windows

x86 executable installer。

下载完成之后，直接双击运行 Python 安装包，使用图形界面安装，设置好 Python 的安装路径，完成后将 Python3和 Python3 的 Scripts 目录配置到环境变量即可。

环境变量的配置，此处以 Win10 系统为例进行演示。

假如我安装后的 Python3 路径为 C:\Python36，从资源管理器中打开该路径，如图所示：



将该路径复制下来。

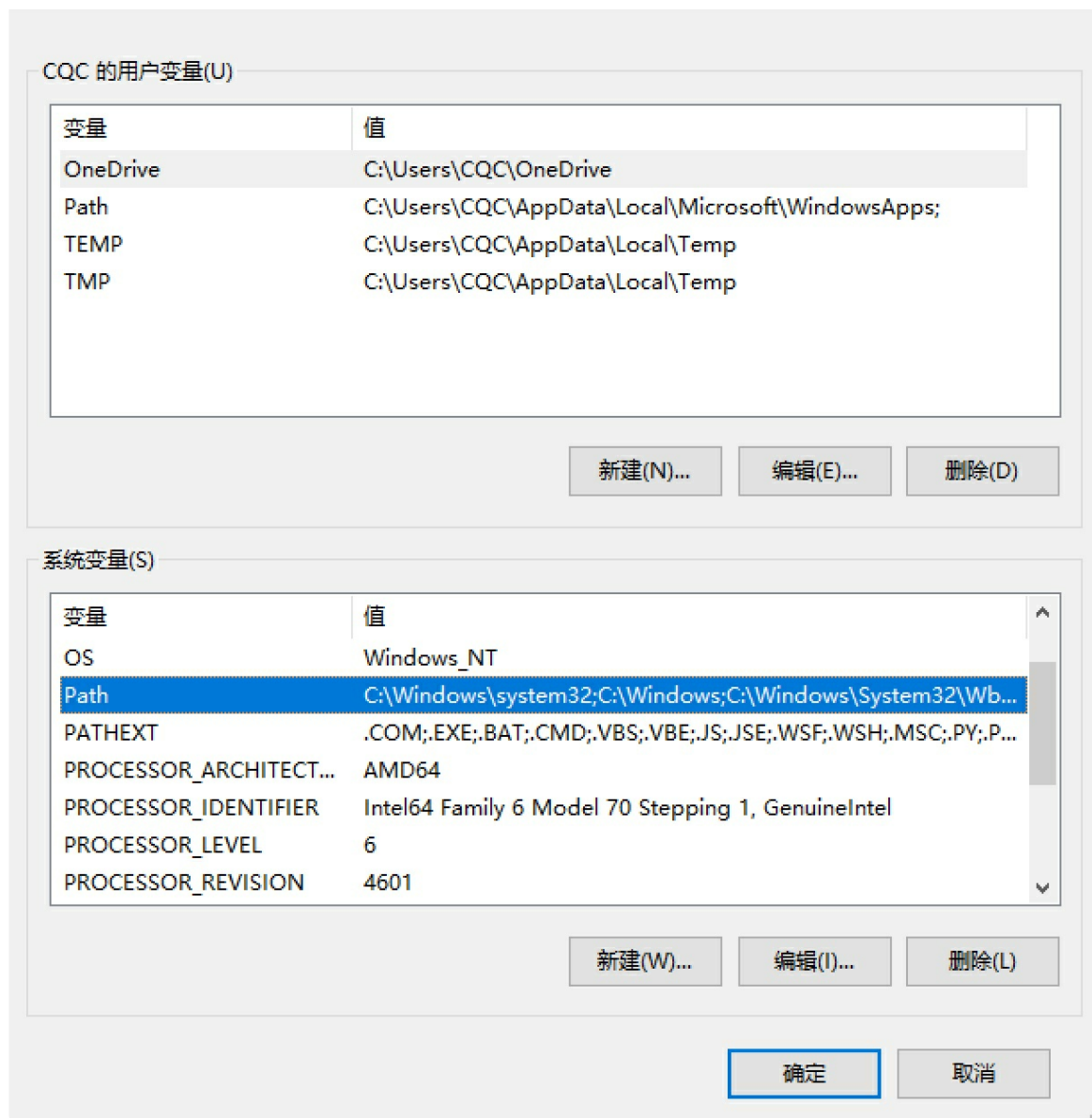
随后打开电脑-属性，如图所示：



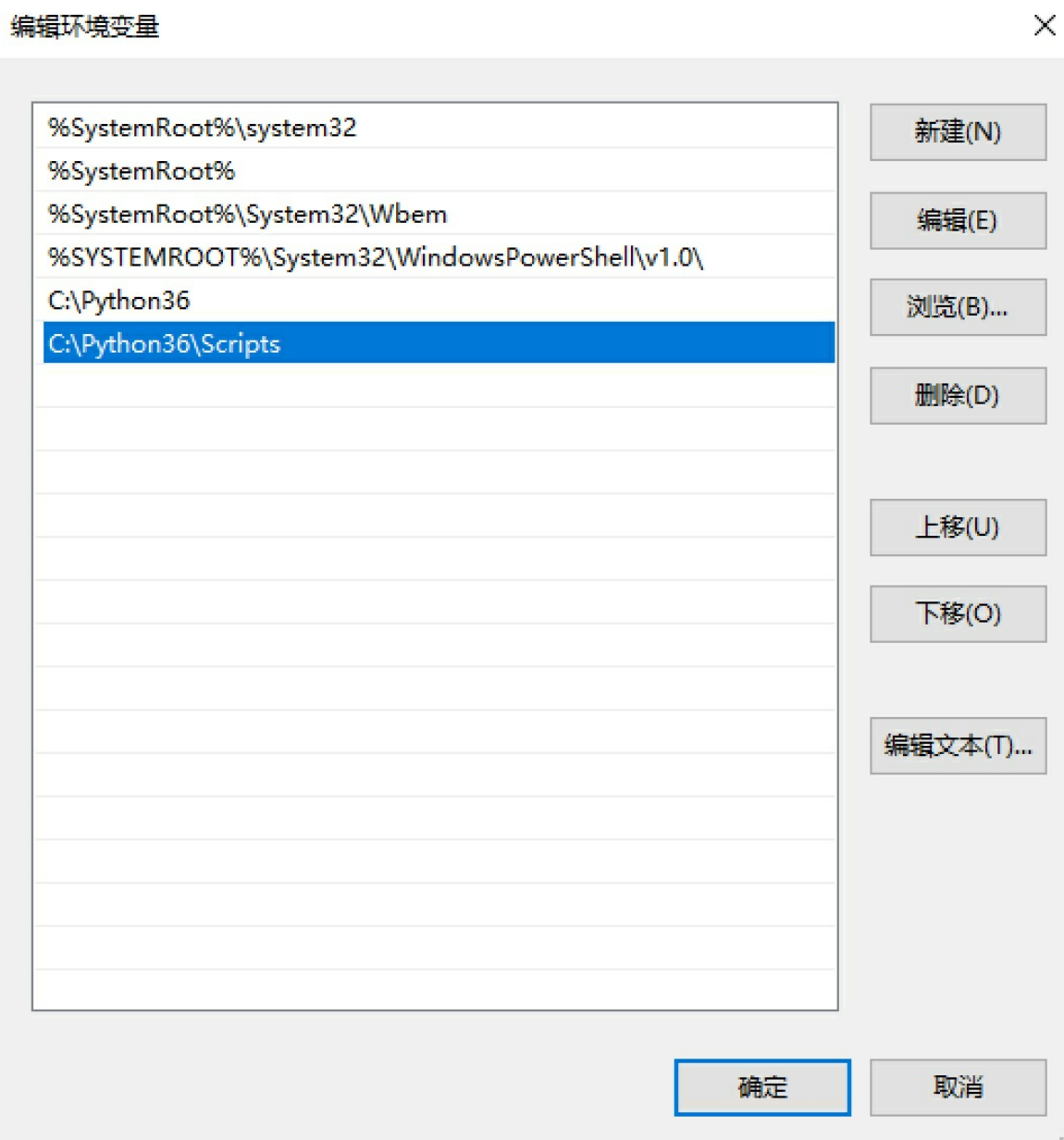
点击左侧的高级系统设置，即可看到在弹出的窗口中下方有环境变量按钮，如图所示：



点击环境变量，找到系统变量下的 Path 变量，随后点击编辑按钮，如图所示：



随后点击新建，新建一个条目，将刚才拷贝的 C:\Python36 复制进去，当然此处的路径就是你的 Python3 安装目录，请自行替换，然后再把 C:\Python36\Scripts 路径复制进去，如图所示：

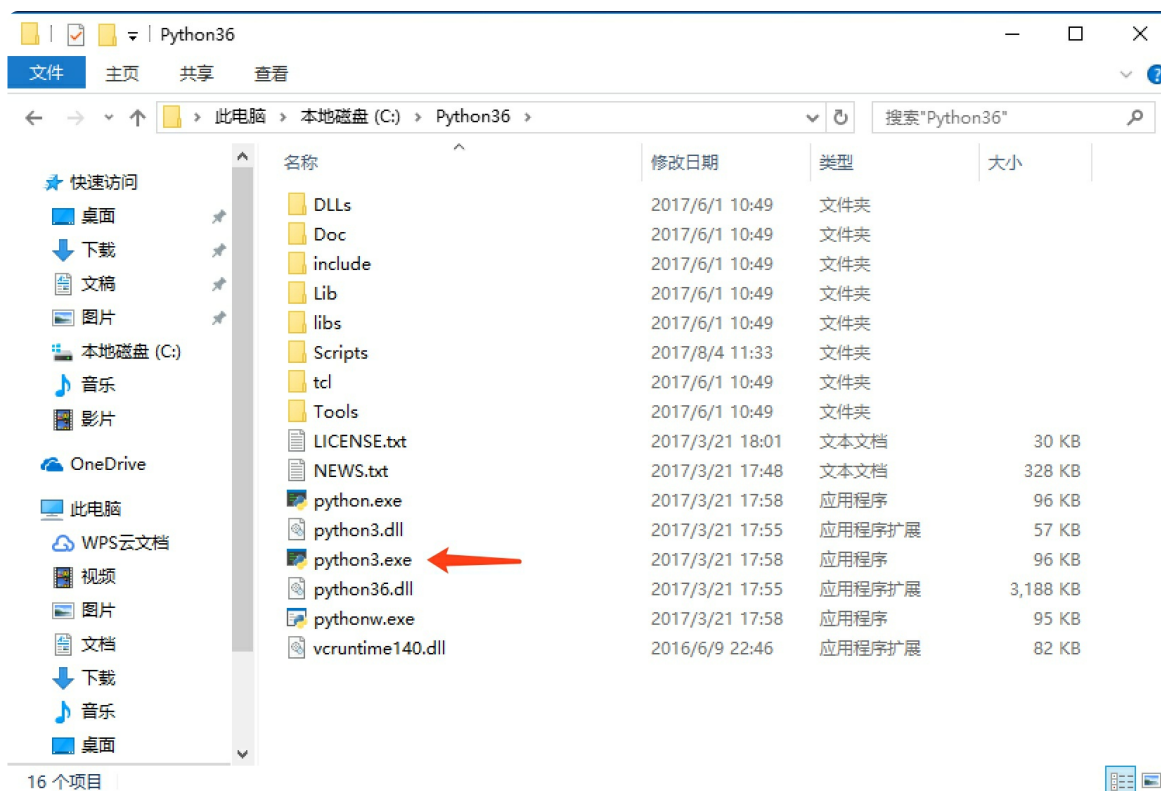


最后点击确定即可完成环境变量的配置。

配置好环境变量之后，我们就可以直接在命令行直接执行环境变量路径下的可执行文件了，如 python、pip 等命令。

添加别名

以上两种安装方式任选其一即可完成安装，但如果我们之前安装过 Python2 的话，可能会导致版本冲突问题，比如在命令行下输入 python 就不知道是调用的 Python2 还是 Python3 了，为了解决这个问题，建议将安装目录中的 python.exe 复制一份，命名为 python3.exe，这样便可以调用 python3 命令了，实际上和 python 命令是完全一致的，这样可以更好地区分 Python 版本，当然如果没有安装过 Python2 的话也建议添加此别名，添加完毕之后如图所示：



对于 Pip 来说，安装包中自带了 pip3.exe 可执行文件，我们也可以直接使用 pip3 命令，无需额外配置。

测试验证

安装完成之后我们可以通过命令行测试一下安装是否成功，在开始菜单搜索 cmd，找到命令提示符，就进入了命令行模式，输入 python 测试一下能否成功调用 python，如果添加了别名的话可以输入 python3 测试，在这里输入的是 python3，如图所示：

```
选择命令提示符

(c) 2017 Microsoft Corporation. 保留所有权利。

C:\Users\CQC>python3
Python 3.6.1 (v3.6.1:69c0db5, Mar 21 2017, 17:54:52) [MSC v.1900 32 bit (Intel)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> print('Hello World')
Hello World
>>> exit()

C:\Users\CQC>pip3 -V
pip 9.0.1 from c:\python36\lib\site-packages (python 3.6)

C:\Users\CQC>
```

类似输出结果如下：


```
$ python3
Python 3.6.1 (v3.6.1:69c0db5, Mar 21 2017, 17:54:52) [MSC v.1900 32 bi
Type "help", "copyright", "credits" or "license" for more information.
>>> print('Hello World')
Hello World
>>> exit()
$ pip3 -V
pip 9.0.1 from c:\python36\lib\site-packages (python 3.6)
```

如果出现类似上面的提示，则证明 Python3 和 Pip3 均安装成功，如果提示命令不存在，那么请检查下环境变量的配置情况。

3. Linux下的安装

Linux 下安装方式有多种，命令安装、源码安装、Anaconda 安装。

使用源码安装需要自行编译，时间较长。推荐使用系统自带命令或 Anaconda 安装，简单高效。在这里对各种安装方式分别予以讲解。

命令行安装

不同的 Linux 发行版本安装方式又有不同，在此分别予以介绍。

CentOS、RedHat

如果是 CentOS 或 RedHat 版本，使用 yum 命令安装即可。

下面列出了 Python3.5 和 3.4 两个版本的安装方法，可以自行选择。

Python3.5 版本：

```
sudo yum install -y https://centos7.iuscommunity.org/ius-release.rpm
sudo yum update
sudo yum install -y python35u python35u-libs python35u-devel python3
5u-pip
```

执行完毕之后便可以成功安装 Python3.5 及 Pip3。

Python3.4 版本：

```
sudo yum groupinstall -y development tools
sudo yum install -y epel-release python34-devel libxslt-devel libxml2-devel openssl-devel
sudo yum install -y python34
sudo yum install -y python34-setuptools
sudo easy_install-3.4 pip
```

执行完毕之后便可以成功安装 Python3.4 及 Pip3。

Ubuntu、Debian、Deepin

首先安装 Python3，使用 apt-get 安装即可，在安装前还需安装一些基础库，命令如下：

```
sudo apt-get install -y python3-dev build-essential libssl-dev libffi-dev libxml2 libxml2-dev libxslt1-dev zlib1g-dev libcurl4-openssl-dev
sudo apt-get install -y python3
```

执行完如上命令之后就可以成功安装好 Python3 了。

然后还需要安装 Pip3，仍然使用 apt-get 安装即可，命令如下：

```
sudo apt-get install -y python3-pip
```

执行完毕之后便可以成功安装 Python3 及 Pip3。

源码安装

如果命令行安装方式有问题，还可以下载 Python3 源码进行安装。

源码下载地址为：<https://www.python.org/ftp/python/>，可以自行选用想要的版本进行安装，在此以 Python3.6.2 为例进行说明，安装路径设置为 /usr/local/python3。

首先创建安装目录，命令如下：

```
sudo mkdir /usr/local/python3
```

随后下载安装包并解压进入，命令如下：

```
wget --no-check-certificate https://www.python.org/ftp/python/3.6.2/Python-3.6.2.tgz
tar -xzvf Python-3.6.2.tgz
cd Python-3.6.2
```

接下来编译安装，所需时间可能较长，请耐心等待，命令如下：

```
sudo ./configure --prefix=/usr/local/python3
sudo make
sudo make install
```

安装完成之后创建 Python3 链接，命令如下：

```
sudo ln -s /usr/local/python3/bin/python3 /usr/bin/python3
```

随后下载 Pip 安装包并安装，命令如下：

```
wget --no-check-certificate https://github.com/pypa/pip/archive/9.0.1.tar.gz
tar -xzvf 9.0.1.tar.gz
cd pip-9.0.1
python3 setup.py install
```

安装完成后再创建 Pip3 链接，命令如下：

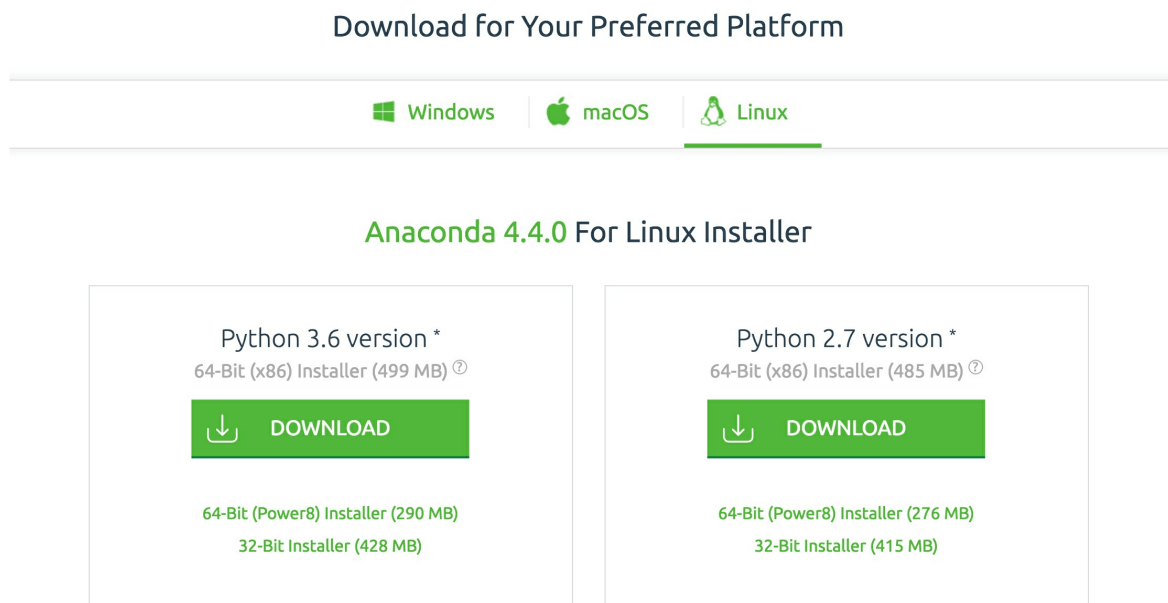
```
sudo ln -s /usr/local/python3/bin/pip /usr/bin/pip3
```

这样就成功安装好了 Python3 及 Pip3。

Anaconda安装

Anaconda 同样支持 Linux，Anaconda 的官方下载链接

为：<https://www.continuum.io/downloads>，选择 Python3 版本的安装包下载即可，如图所示：



如果下载速度过慢同样可以使用清华镜像，参考 Windows 部分的介绍，在此不再赘述。

测试验证

命令行测试 Python3 和 Pip3 是否安装成功。

```
$ python3
Python 3.5.2 (default, Nov 17 2016, 17:05:23)
Type "help", "copyright", "credits" or "license" for more information.
>>> exit()
$ pip3 -V
pip 8.1.1 from /usr/lib/python3/dist-packages (python 3.5)
```

如出现类似上面的提示，则证明 Python3 和 Pip3 安装成功。

4. Mac下的安装

在 Mac 下同样有多种安装方式，如 Homebrew、安装包安装、Anaconda 安装等，推荐使用 Homebrew 安装。

Homebrew安装

Homebrew 是 Mac 平台下强大的包管理工具，首先安装 Homebrew，官方网站是：<https://brew.sh/>。

执行如下命令即可安装 Homebrew：

```
ruby -e "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/master/install)"
```

安装完成后便可以使用 brew 命令安装 Python3 和 Pip3 了。

```
brew install python3
```

命令执行完成之后发现 Python3 和 Pip3 均已经安装成功。

安装包安装

可以到官方网站下载 Python3 安装包。链接为：<https://www.python.org/downloads/>，页面如图 所示：

Files					
Version	Operating System	Description	MD5 Sum	File Size	PGP
Gzipped source tarball	Source release		e1a36bffffd1d3a780b1825daf16e56c	22580749	SIG
XZ compressed source tarball	Source release		2c68846471994897278364fc18730dd9	16907204	SIG
Mac OS X 64-bit/32-bit installer	Mac OS X	for Mac OS X 10.6 and later	86e6193fd56b1e757fc8a5a2bb6c52ba	27561006	SIG
Windows help file	Windows		e520a5c1c3e3f02f68e3db23f74a7a90	8010498	SIG
Windows x86-64 embeddable zip file	Windows	for AMD64/EM64T/x64	0fdfe9f79e0991815d6fc1712871c17f	7047535	SIG
Windows x86-64 executable installer	Windows	for AMD64/EM64T/x64	4377e7d4e6877c248446f7cd6a1430cf	31434856	SIG
Windows x86-64 web-based installer	Windows	for AMD64/EM64T/x64	58ffad3d92a590a463908dfedbc68c18	1312496	SIG
Windows x86 embeddable zip file	Windows		2ca4768fdbadf6e670e97857bfab83e8	6332409	SIG
Windows x86 executable installer	Windows		8d8e1711ef9a4b3d3d0ce21e4155c0f5	30507592	SIG
Windows x86 web-based installer	Windows		ccb7d66e3465eaf40ade05b76715b56c	1287040	SIG

Mac 平台下可以选择下载 Mac OS X 64-bit/32-bit installer，下载完成之后打开安装包按照提示安装即可。

Anaconda安装

Anaconda 同样支持 Mac，Anaconda 的官方下载链接为：<https://www.continuum.io/downloads>，选择 Python3 版本的安装包下载即可，如图所示：

Download for Your Preferred Platform

Windows

macOS

Linux

Anaconda 4.4.0 For macOS Graphical Installer

Python 3.6 version *
Graphical Installer (442 MB) ⓘ



DOWNLOAD

Command-Line Installer (380 MB) ⓘ

Python 2.7 version *
Graphical Installer (438 MB) ⓘ



DOWNLOAD

Command-Line Installer (375 MB) ⓘ

如果下载速度过慢同样可以使用清华镜像，参考 Windows 部分的介绍，在此不再赘述。

测试验证

打开终端，在命令行测试 Python3 和 Pip3 是否成功安装，如图所示：

```
CQC — -zsh — 80x24
[➔ ~ python3
Python 3.6.1 (default, Apr  4 2017, 09:40:21)
[GCC 4.2.1 Compatible Apple LLVM 8.1.0 (clang-802.0.38)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> exit()
[➔ ~ pip3 -V
pip 9.0.1 from /usr/local/lib/python3.6/site-packages (python 3.6)
[➔ ~ ]
```

如出现上面的提示，则证明 Python3 和 Pip3 安装成功。

二、Scrapy的安装

Scrapy 是一个十分强大的爬虫框架，依赖的库比较多，至少需要依赖库有 Twisted 14.0, lxml 3.4, pyOpenSSL 0.14。而在不同平台环境又各不相同，所以在安装之前最好确保把一些基本库安装好，尤其是 Windows。本节介绍一下 Scrapy 在不同平台的安装方法。

1. 相关链接

- 官方网站: <https://scrapy.org>
- 官方文档: <https://docs.scrapy.org>
- PyPi: <https://pypi.python.org/pypi/Scrapy>
- GitHub: <https://github.com/scrapy/scrapy>
- 中文文档: <http://scrapy-chs.readthedocs.io>

2. Anaconda安装

这种方法是一种比较简单的安装 Scrapy 的方法（尤其是对 Windows 来说），如果你的 Python 是使用 Anaconda 安装的，或者还没有安装 Python 的话，可以使用该方法安装，简单省力，当然如果你的 Python 不是通过 Anaconda 安装的，可以继续查看下文中各平台直接安装方法。

Anaconda 的安装方式可以查看前文中 Python3 的安装方式一节，在此不再赘述。

如果已经安装好了 Anaconda，那么可以通过 conda 命令安装 Scrapy，安装命令如下：

```
conda install Scrapy
```

运行之后便可以完成 Scrapy 的安装。

3. Windows下的安装

如果你的 Python 不是使用 Anaconda 安装的，可以参考如下方式来一步步完成 Scrapy 的安装。

安装LXML

LXML 的安装过程请参见前文 LXML 安装方法讲解，在此不再赘述，此库非常重要，请一定要安装成功。

安装pyOpenSSL

官方网站下载 Wheel 文件，<https://pypi.python.org/pypi/pyOpenSSL#downloads>，如当

前最新版本名称是 pyOpenSSL-17.2.0-py2.py3-none-any.whl， 直接下载即可， 如图所示：

Changes:

- Fixed a bug causing Context.set_default_verify_paths() to not work with cryptography manylinux1 wheels on Python 3.x. [#665](#)
- Fixed a crash with (EC)DSA signatures in some cases. [#670](#)

[Full changelog.](#)

File	Type	Py Version	Uploaded on	Size
pyOpenSSL-17.2.0-py2.py3-none-any.whl (md5, pgp) 	Python Wheel	py2.py3	2017-07-20	51KB
pyOpenSSL-17.2.0.tar.gz (md5, pgp)	Source		2017-07-20	166KB

Author: Hynek Schlawack

Documentation: [pyOpenSSL package documentation](#)

Home Page: <https://pyopenssl.org/>

License: Apache License, Version 2.0

Categories

Development Status :: 6 - Mature

Intended Audience :: Developers

下载后利用 Pip 安装即可：

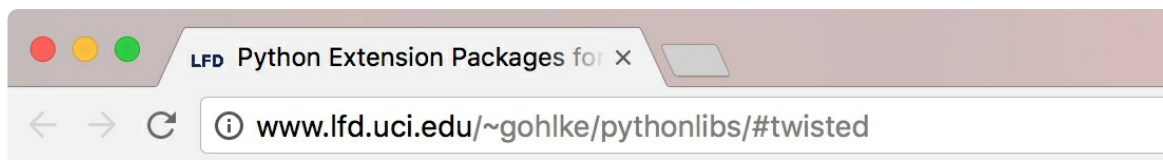
```
pip3 install pyOpenSSL-17.2.0-py2.py3-none-any.whl
```

安装Twisted

到 <http://www.lfd.uci.edu/~gohlke/pythonlibs/#twisted> 下载 Wheel 文件， 利用 Pip 安装即可。

如 Python 3.6 版本， Windows 64 位系统， 当前最新版本为

Twisted-17.5.0-cp36-cp36m-win_amd64.whl， 直接下载即可， 如图所示， 最新版本以网站为准。



Twisted, an event-driven networking engine.

[Twisted-17.5.0-cp27-cp27m-win32.whl](#)

[Twisted-17.5.0-cp27-cp27m-win_amd64.whl](#)

[Twisted-17.5.0-cp34-cp34m-win32.whl](#)

[Twisted-17.5.0-cp34-cp34m-win_amd64.whl](#)

[Twisted-17.5.0-cp35-cp35m-win32.whl](#)

[Twisted-17.5.0-cp35-cp35m-win_amd64.whl](#)

[Twisted-17.5.0-cp36-cp36m-win32.whl](#)

[Twisted-17.5.0-cp36-cp36m-win_amd64.whl](#)



然后 Pip 安装即可：

```
pip3 install Twisted-17.5.0-cp36-cp36m-win_amd64.whl
```

安装PyWin32

从官方网站下载对应版本的安装包即可，链接

为：<https://sourceforge.net/projects/pywin32/files/pywin32/Build%202021/>，如图所示：

[Home](#) / [Browse](#) / [Python for Windows Extensions](#) / Files

Python for Windows Extensions

Python extensions for Windows

Brought to you by: [mhammond](#)[Summary](#)[Files](#)[Reviews](#)[Support](#)[Wiki](#)[Mailing Lists](#)[Tickets ▾](#)[Donate](#) Looking for the latest version? [Download README.txt \(951 Bytes\)](#)[Home](#) / [pywin32](#) / Build 221

Name ▴	Modified ▴	Size ▴	Downloads / Week ▴
↑ Parent folder			
pywin32-221.zip	2017-03-29	7.3 MB	581
pywin32-221.win-amd64-py3.7.exe	2017-03-29	9.5 MB	440
pywin32-221.win-amd64-py3.6.exe	2017-03-29	9.5 MB	1,229
pywin32-221.win-amd64-py3.5.exe	2017-03-29	9.5 MB	438
pywin32-221.win-amd64-py3.4.exe	2017-03-29	8.8 MB	144
pywin32-221.win-amd64-py3.3.exe	2017-03-29	8.8 MB	24
pywin32-221.win-amd64-py3.2.exe	2017-03-29	7.5 MB	10
pywin32-221.win-amd64-py3.1.exe	2017-03-29	7.5 MB	12
pywin32-221.win-amd64-py2.7.exe	2017-03-29	7.5 MB	1,611
pywin32-221.win-amd64-py2.6.exe	2017-03-29	7.5 MB	66

如 Python 3.6 版本可以选择下载 `pywin32-221.win-amd64-py3.6.exe`，下载完毕之后双击安装即可。

注意这里使用的是 Build 221 版本，随着时间推移，版本肯定会继续更新，最新的版本可以查看：<https://sourceforge.net/projects/pywin32/files/pywin32/>，查找最新的版本安装即可。

安装Scrapy

安装好了以上的依赖库，最后安装 Scrapy 就非常简单了，依然使用 Pip，命令如下：

```
pip3 install Scrapy
```

等待命令结束没有报错，就证明 Scrapy 已经安装好了。

4. Linux下的安装

在 Linux 上的安装方式依然分为两类平台介绍。

CentOS、RedHat

首先确保一些依赖库已经安装，运行如下命令：

```
sudo yum groupinstall -y development tools
sudo yum install -y epel-release libxslt-devel libxml2-devel openssl
-devel
```

最后利用 Pip 安装 Scrapy 即可，运行如下命令：

```
pip3 install Scrapy
```

Ubuntu、Debian、Deepin

首先确保一些依赖库已经安装，运行如下命令：

```
sudo apt-get install build-essential python3-dev libssl-dev libffi-d
ev libxml2 libxml2-dev libxslt1-dev zlib1g-dev
```

然后利用 Pip 安装 Scrapy 即可，运行如下命令：

```
pip3 install Scrapy
```

运行完毕之后即可完成 Scrapy 的安装。

5. Mac下的安装

在 Mac 下首先也是进行依赖库的安装。

在 Mac 上构建 Scrapy 的依赖库需要 C 编译器以及开发头文件，它一般由 Xcode 提供，运行如下命令安装即可：

```
xcode-select --install
```

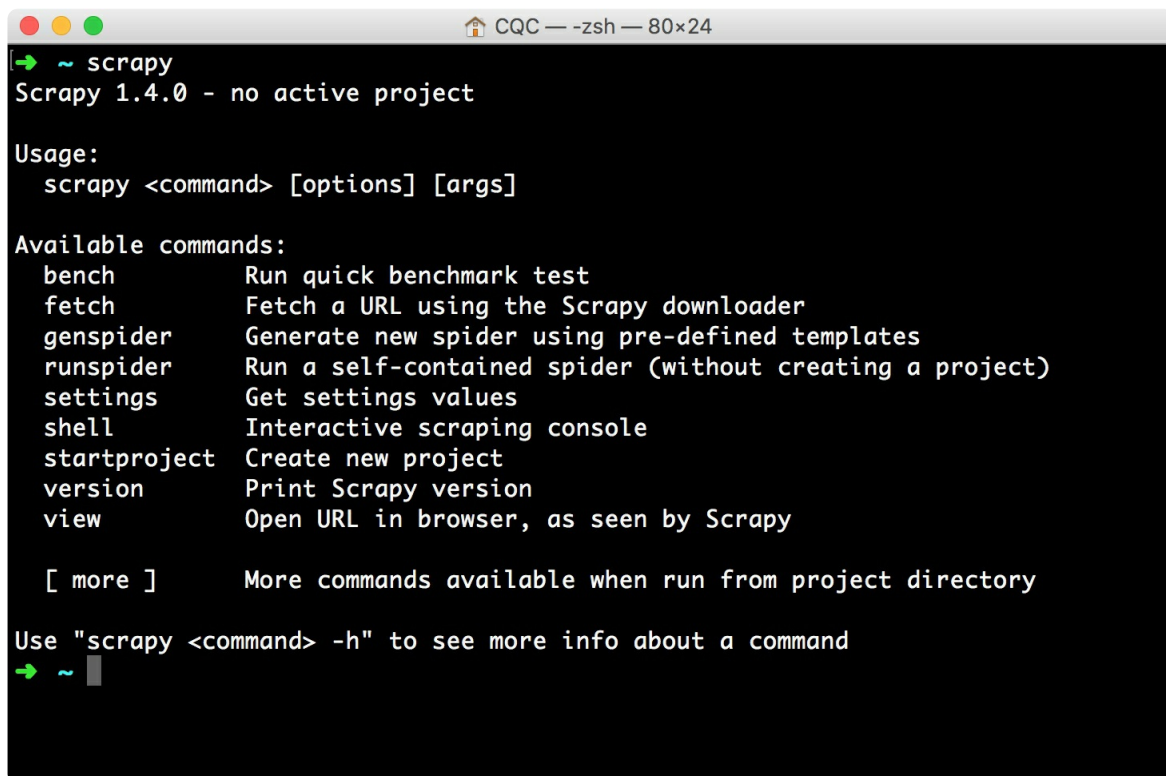
随后利用 Pip 安装 Scrapy 即可，运行如下命令：

```
pip3 install Scrapy
```

运行完毕之后即可完成 Scrapy 的安装。

6. 验证安装

安装之后，在命令行下输入 scrapy，如果出现类似下方的结果，就证明 Scrapy 安装成功，如图所示：



```
CQC — -zsh — 80x24
[➔ ~ scrapy
Scrapy 1.4.0 - no active project

Usage:
  scrapy <command> [options] [args]

Available commands:
  bench          Run quick benchmark test
  fetch          Fetch a URL using the Scrapy downloader
  genspider       Generate new spider using pre-defined templates
  runspider      Run a self-contained spider (without creating a project)
  settings       Get settings values
  shell          Interactive scraping console
  startproject    Create new project
  version        Print Scrapy version
  view           Open URL in browser, as seen by Scrapy

  [ more ]       More commands available when run from project directory

Use "scrapy <command> -h" to see more info about a command
➔ ~ █
```

7. 常见错误

pkg_resources.VersionConflict: (six 1.5.2 (/usr/lib/python3/dist-packages), Requirement.parse('six>=1.6.0'))

six 包版本过低，six包是一个提供兼容 Python2 和 Python3 的库，升级 six 包即可：

```
sudo pip3 install -U six
```

c/_cffi_backend.c:15:17: fatal error: ffi.h: No such file or directory

这是在 Linux 下常出现的错误，缺少 Libffi 这个库。什么是 libffi？“FFI”的全名是 Foreign Function Interface，通常指的是允许以一种语言编写的代码调用另一种语言的代码。而 Libffi 库只提供了最底层的、与架构相关的、完整的“FFI”。

安装相应的库即可。

Ubuntu、Debian：

```
sudo apt-get install build-essential libssl-dev libffi-dev python3-dev
```

CentOS、RedHat：

```
sudo yum install gcc libffi-devel python-devel openssl-devel
```

Command "python setup.py egg_info" failed with error code 1 in /tmp/pip-build/cryptography/

这是缺少加密的相关组件，利用Pip 安装即可。

```
pip3 install cryptography
```

ImportError: No module named 'packaging'

缺少 packaging 这个包，它提供了 Python 包的核心功能，利用 Pip 安装即可。

ImportError: No module named '_cffi_backend'

缺少 cffi 包，使用 Pip 安装即可：

```
pip3 install cffi
```

ImportError: No module named 'pyparsing'

缺少 pyparsing 包，使用 Pip 安装即可：

```
pip3 install pyparsing appdirs
```

三、使用

下载代码 <http://114.215.100.72:8095/summary/ScrapyUniversal.git>

示例运行：

```
python3 run.py xinhua
```

运行之后便会出现爬取结果。