

# Scrapy

Instalando o Scrapy Framework

# Scrapy - Instalação

*Scrapy* é um *framework Python* para raspagem de dados em grande escala. *Scrapy* fornece todas as ferramentas que você precisa para extrair dados de sites de forma eficiente, processá-los como quiser e armazená-los em sua estrutura e formatos preferidos.

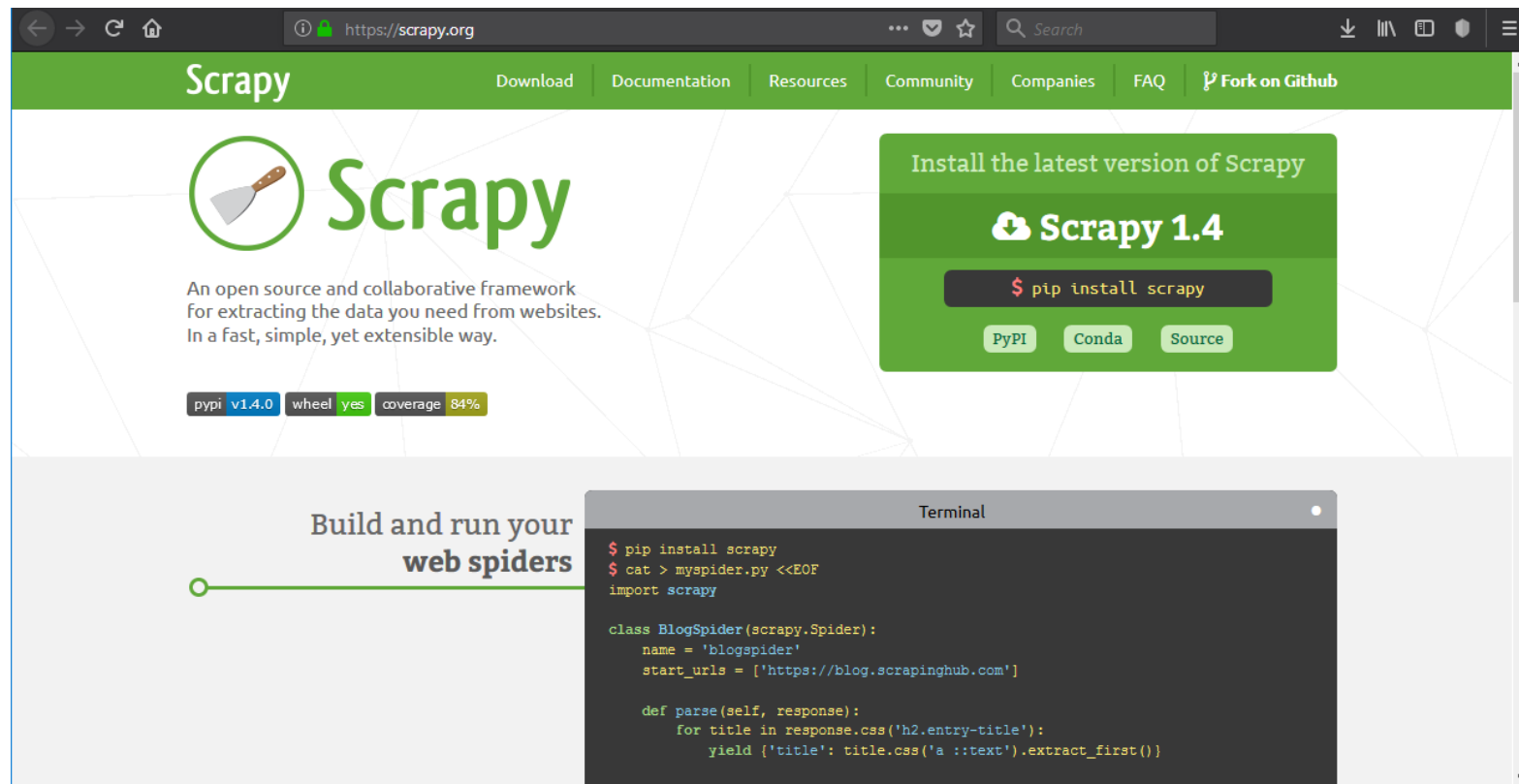
## **Definição retirada do site do projeto.**

*Scrapy* é um *framework* colaborativo e *open source* para extrair os dados que você precisa de websites.

De forma rápida, simples, porém, extensível.

# Scrapy - Instalação

Site do Projeto: <https://scrapy.org>



# Scrapy - Instalação

## Instalando o Scrapy no Windows 10

O *Scrapy* é escrito em *Python* puro e depende de alguns pacotes como o *lxml*, *parsel*, *w3lib*, *twisted* e *cryptography* e *pyOpenSSL*.

A maioria das dependências são instaladas automaticamente.

Porém, algumas dependências teremos que instalar antes de usar o *Scrapy*.


# Scrapy - Instalação

O *lxml* precisa de algumas bibliotecas C, sendo assim, começaremos instalando o “*Microsoft Visual C++ 14.0*” que está disponível em:

<http://landinghub.visualstudio.com/visual-cpp-build-tools>

# Scrapy - Instalação

Caso não tenha instalado o Microsoft Visual C++ 14.0, será exibido este erro quando tentar instalar o *Scrapy*:



```
Administrator: Command Prompt

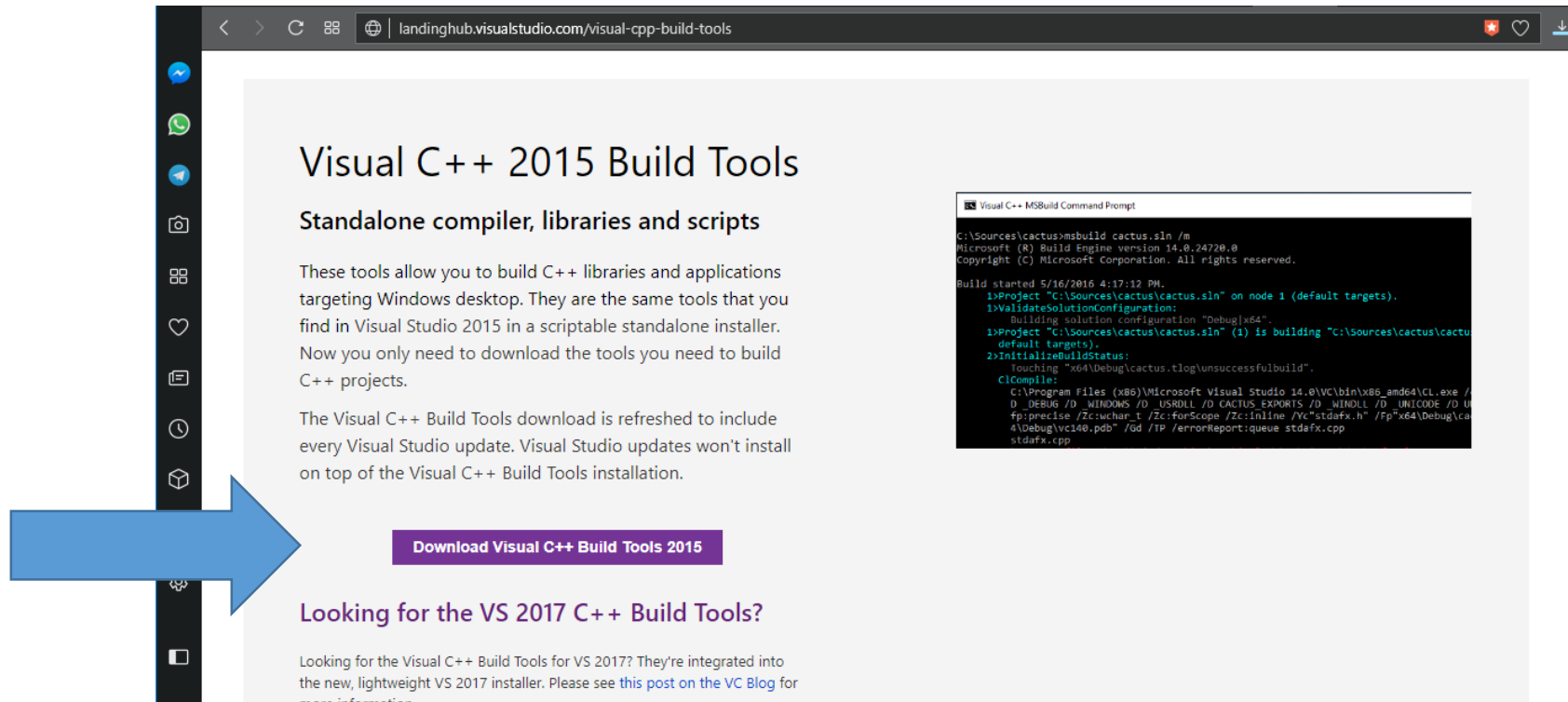
running build_ext
building 'twisted.test.raiser' extension
error: Microsoft Visual C++ 14.0 is required. Get it with
"Microsoft Visual C++ Build Tools": http://landinghub.visualst
udio.com/visual-cpp-build-tools

-----
Command "c:\evaldo\ferramentasdesenvolvimento\python\python36\
python.exe -u -c "import setuptools, tokenize;__file__='C:\\Us
ers\\evaldo\\AppData\\Local\\Temp\\pip-build-sn_j8bxd\\Twisted
\\setup.py';f=getattr(tokenize, 'open', open)(__file__);code=f
.read().replace('\\r\\n', '\\n');f.close();exec(compile(code, __f
ile__, 'exec'))" install --record C:\\Users\\evaldo\\AppData\\Loca
l\\Temp\\pip-ludc7j4_-record\\install-record.txt --single-version
-externally-managed --compile" failed with error code 1 in C:\\
Users\\evaldo\\AppData\\Local\\Temp\\pip-build-sn_j8bxd\\Twisted\\

C:\\WINDOWS\\system32>
```

# Scrapy - Instalação

Faça o download do Microsoft Visual C++ 14.0.



The screenshot shows a web browser window at the URL `landinghub.visualstudio.com/visual-cpp-build-tools`. The page title is "Visual C++ 2015 Build Tools" with the subtitle "Standalone compiler, libraries and scripts". The main text describes these tools as standalone compilers for building C++ libraries and applications on Windows desktop, noting they are the same as those in Visual Studio 2015 but available as a standalone installer. A large blue arrow points from the left towards a purple button labeled "Download Visual C++ Build Tools 2015". Below this button, there is a section titled "Looking for the VS 2017 C++ Build Tools?" which mentions that for VS 2017, the tools are integrated into the new installer and refers to a blog post for more information. An inset image on the right shows a Visual C++ MSBuild Command Prompt window displaying the output of a build command, including project configuration and compilation details.

Visual C++ 2015 Build Tools

Standalone compiler, libraries and scripts

These tools allow you to build C++ libraries and applications targeting Windows desktop. They are the same tools that you find in Visual Studio 2015 in a scriptable standalone installer. Now you only need to download the tools you need to build C++ projects.

The Visual C++ Build Tools download is refreshed to include every Visual Studio update. Visual Studio updates won't install on top of the Visual C++ Build Tools installation.

[Download Visual C++ Build Tools 2015](#)

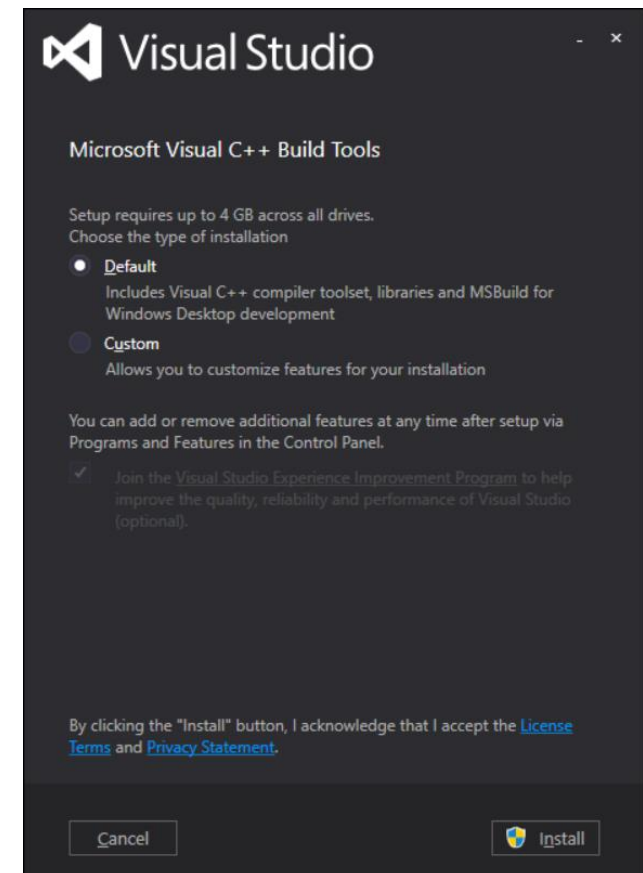
Looking for the VS 2017 C++ Build Tools?

Looking for the Visual C++ Build Tools for VS 2017? They're integrated into the new, lightweight VS 2017 installer. Please see [this post on the VC Blog](#) for more information.

# Scrapy - Instalação

Execute o arquivo de instalação:  
*visualcppbuildtools\_full.exe*

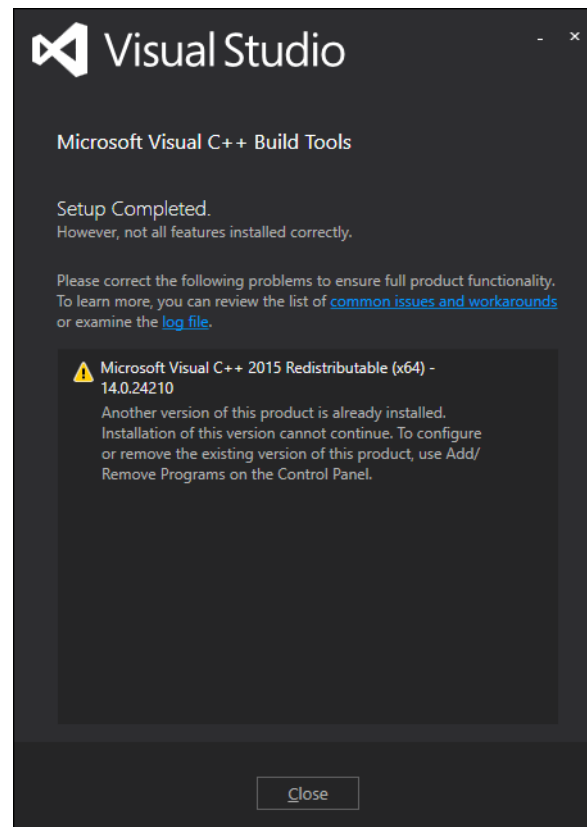
Será exibida a tela de instalação, deixe marcado “*Default*”, clique em “*Install*”, se aparecer uma mensagem solicitando confirmação, responda “sim” e aguarde a instalação.





# Scrapy - Instalação

Instalação do Microsoft Visual C++ 14.0 finalizada



# Scrapy - Instalação

Agora podemos instalar utilizando o *pip*:

*\$ pip install scrapy*

```
Command Prompt - pip install scrapy
C:\Users\evaldo>pip install scrapy
Collecting scrapy
  Downloading Scrapy-1.4.0-py2.py3-none-any.whl (248kB)
    100% |████████████████████| 256kB 563kB/s
Collecting service-identity (from scrapy)
  Downloading service_identity-17.0.0-py2.py3-none-any.whl
Collecting cssselect>=0.9 (from scrapy)
  Downloading cssselect-1.0.1-py2.py3-none-any.whl
Collecting pyOpenSSL (from scrapy)
  Downloading pyOpenSSL-17.4.0-py2.py3-none-any.whl (52kB)
    100% |████████████████████| 61kB 520kB/s
Collecting parsel>=1.1 (from scrapy)
  Downloading parsel-1.2.0-py2.py3-none-any.whl
Collecting queuelib (from scrapy)
  Downloading queuelib-1.4.2-py2.py3-none-any.whl
Collecting six>=1.5.2 (from scrapy)
  Downloading six-1.11.0-py2.py3-none-any.whl
Collecting w3lib>=1.17.0 (from scrapy)
  Downloading w3lib-1.18.0-py2.py3-none-any.whl
Collecting PyDispatcher>=2.0.5 (from scrapy)
  Downloading PyDispatcher-2.0.5.tar.gz
Collecting lxml (from scrapy)
  Downloading lxml-4.1.1-cp36-cp36m-win_amd64.whl (3.5MB)
    99% |████████████████████| 3.5MB 402kB/s eta 0:00:01
```

← Instalando

Instalação finalizada →

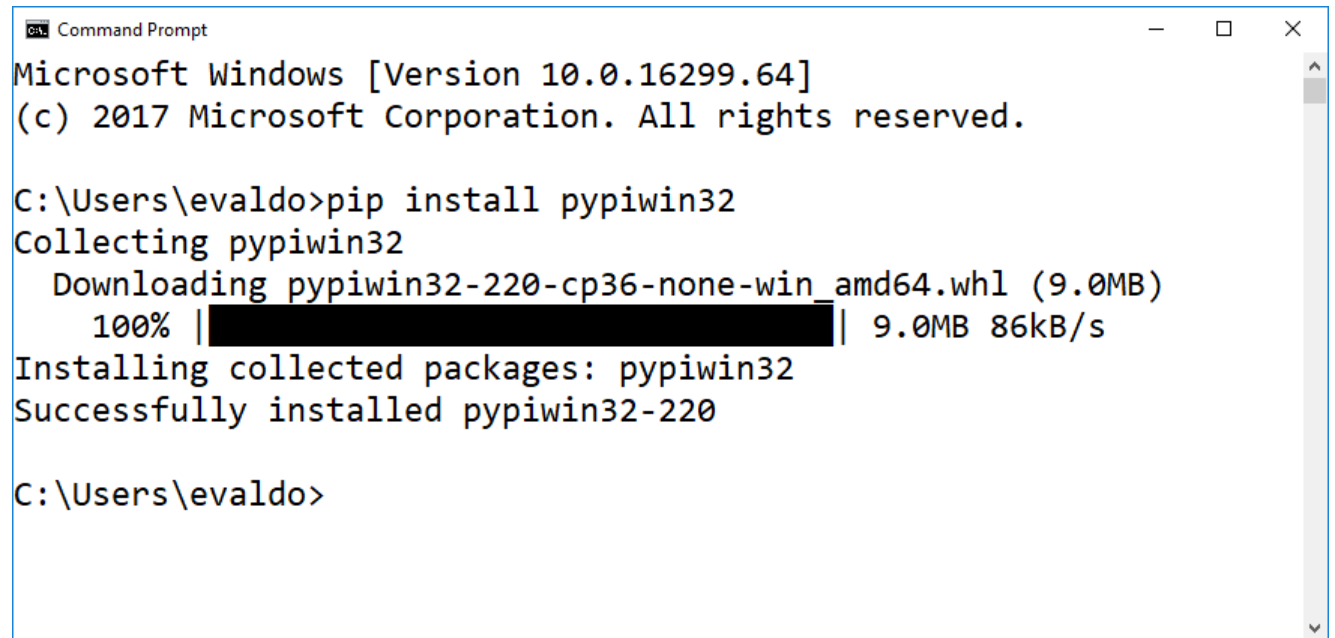
```
Command Prompt
lvimento\python\python36\lib\site-packages (from service-identity->scrapy)
Requirement already satisfied: attrs in c:\evaldo\ferramentasdesenvolvimento\python\python36\lib\site-packages (from service-identity->scrapy)
Requirement already satisfied: pyasn1 in c:\evaldo\ferramentasdesenvolvimento\python\python36\lib\site-packages (from service-identity->scrapy)
Requirement already satisfied: setuptools in c:\evaldo\ferramentasdesenvolvimento\python\python36\lib\site-packages (from zope.interface>=4.0.2->Twisted>=13.1.0->scrapy)
Requirement already satisfied: idna>=2.1 in c:\evaldo\ferramentasdesenvolvimento\python\python36\lib\site-packages (from cryptography>=1.9->pyOpenSSL->scrapy)
Requirement already satisfied: asn1crypto>=0.21.0 in c:\evaldo\ferramentasdesenvolvimento\python\python36\lib\site-packages (from cryptography>=1.9->pyOpenSSL->scrapy)
Requirement already satisfied: cffi>=1.7; platform_python_implementation != "PyPy" in c:\evaldo\ferramentasdesenvolvimento\python\python36\lib\site-packages (from cryptography>=1.9->pyOpenSSL->scrapy)
Requirement already satisfied: pycparser in c:\evaldo\ferramentasdesenvolvimento\python\python36\lib\site-packages (from cffi>=1.7; platform_python_implementation != "PyPy"->cryptography>=1.9->pyOpenSSL->scrapy)
Installing collected packages: Twisted, scrapy
  Running setup.py install for Twisted ... done
Successfully installed Twisted-17.9.0 scrapy-1.4.0

C:\Users\evaldo>
```

# Scrapy - Instalação

Para utilizar o Scrapy também devemos instalar o *pypiwin32* com o comando a seguir:

*\$ pip install pypiwin32*



```
Command Prompt
Microsoft Windows [Version 10.0.16299.64]
(c) 2017 Microsoft Corporation. All rights reserved.

C:\Users\evaldo>pip install pypiwin32
Collecting pypiwin32
  Downloading pypiwin32-220-cp36-none-win_amd64.whl (9.0MB)
    100% |████████████████████████████████████████| 9.0MB 86kB/s
Installing collected packages: pypiwin32
Successfully installed pypiwin32-220

C:\Users\evaldo>
```

# Scrapy - Instalação

Verificando se foi instalado corretamente.

```
Command Prompt - python

C:\Users\evaldo>python
Python 3.6.3 (v3.6.3:2c5fed8, Oct  3 2017,
18:11:49) [MSC v.1900 64 bit (AMD64)] on
win32
Type "help", "copyright", "credits" or "li
cense" for more information.
>>> import scrapy
>>> █
```

```
Command Prompt

C:\Users\evaldo>scrapy version
Scrapy 1.4.0

C:\Users\evaldo> █
```

# Scrapy - Instalação

## Instalando o *Scrapy* no Linux

Para instalar o *Scrapy* no Linux execute o comando

```
$ sudo pip install scrapy
```

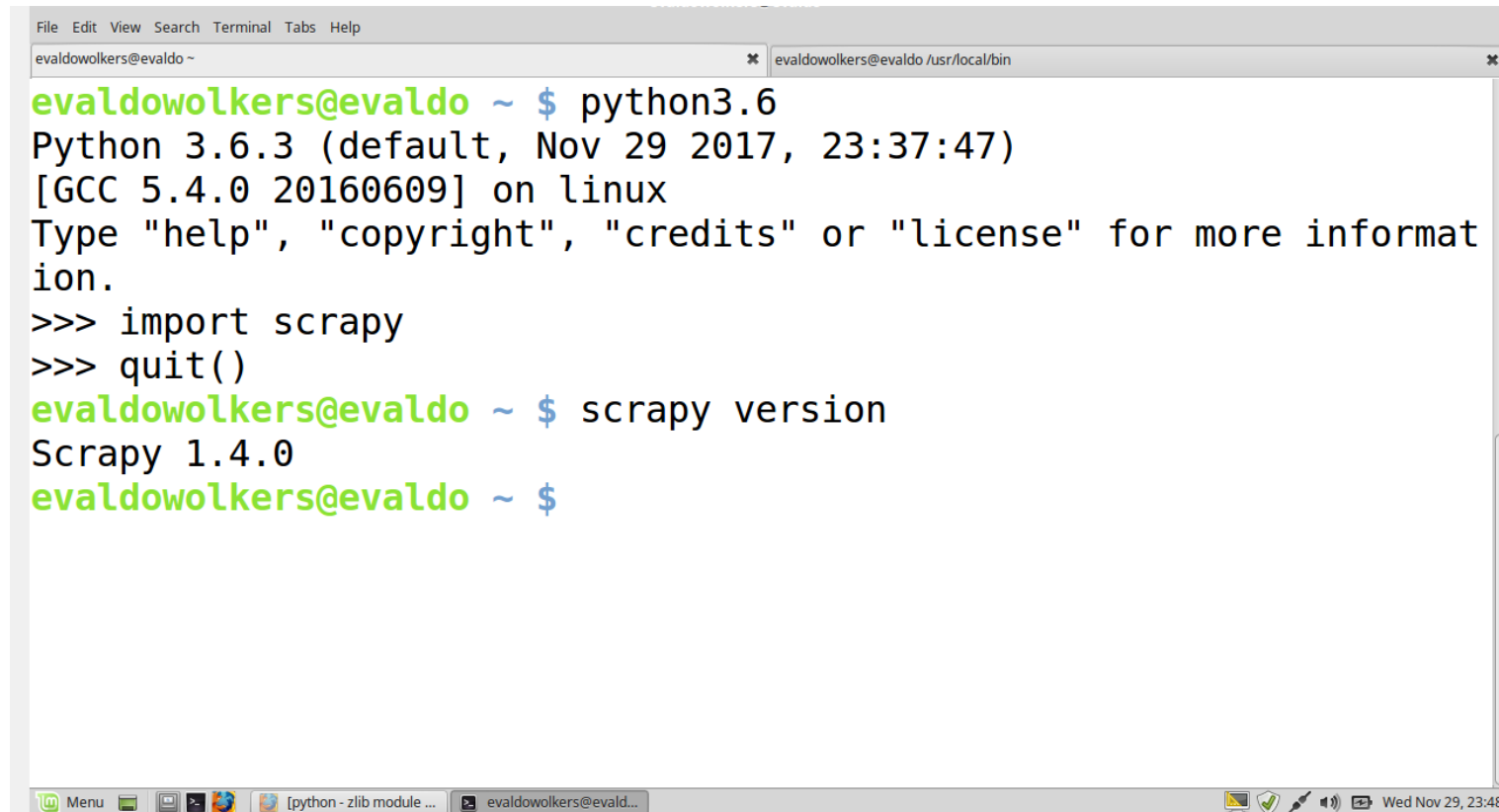
Caso tenha mais de uma versão poderá ter que especificar por exemplo:

```
$ sudo pip3.6 install scrapy
```

Vai depender da versão do Python que esteja instalada e também da distribuição do Linux que estiver utilizando.

# Scrapy - Instalação

Verificando se foi instalado corretamente.

A screenshot of a terminal window with a light gray title bar and menu bar. The title bar contains 'File Edit View Search Terminal Tabs Help'. The menu bar has 'evaldowolkers@evaldo ~' and 'evaldowolkers@evaldo /usr/local/bin'. The terminal content shows a user running 'python3.6', which outputs 'Python 3.6.3 (default, Nov 29 2017, 23:37:47) [GCC 5.4.0 20160609] on linux'. The user then enters a Python prompt '>>>' and runs 'import scrapy' and 'quit()'. After exiting Python, the user runs 'scrapy version' at the shell prompt, which outputs 'Scrapy 1.4.0'. The terminal ends with the shell prompt. The bottom status bar shows a menu icon, a taskbar with icons for a terminal, a file manager, and a web browser, and system icons for network, volume, and battery, along with the date and time 'Wed Nov 29, 23:48'.

```
evaldowolkers@evaldo ~ $ python3.6
Python 3.6.3 (default, Nov 29 2017, 23:37:47)
[GCC 5.4.0 20160609] on linux
Type "help", "copyright", "credits" or "license" for more informat
ion.
>>> import scrapy
>>> quit()
evaldowolkers@evaldo ~ $ scrapy version
Scrapy 1.4.0
evaldowolkers@evaldo ~ $
```

# FIM