

NLTK Installation

Step 1: Visit <http://www.nltk.org/install.html>

Step 2: Click **Install NLTK URL**

Windows

These instructions assume that you do not already have Python installed on your machine.

32-bit binary installation

1. Install Python 3.5: <http://www.python.org/downloads/> (avoid the 64-bit versions)
2. Install Numpy (optional): <http://sourceforge.net/projects/numpy/files/NumPy/> (the version that specifies python3.5)
3. Install NLTK: <http://pypi.python.org/pypi/nltk>
4. Test installation: Start>Python35, then type `import nltk`

Installing Third-Party Software

Step 3: Click **nltk-3.2.5.win32.exe (md5)**

nltk 3.2.5

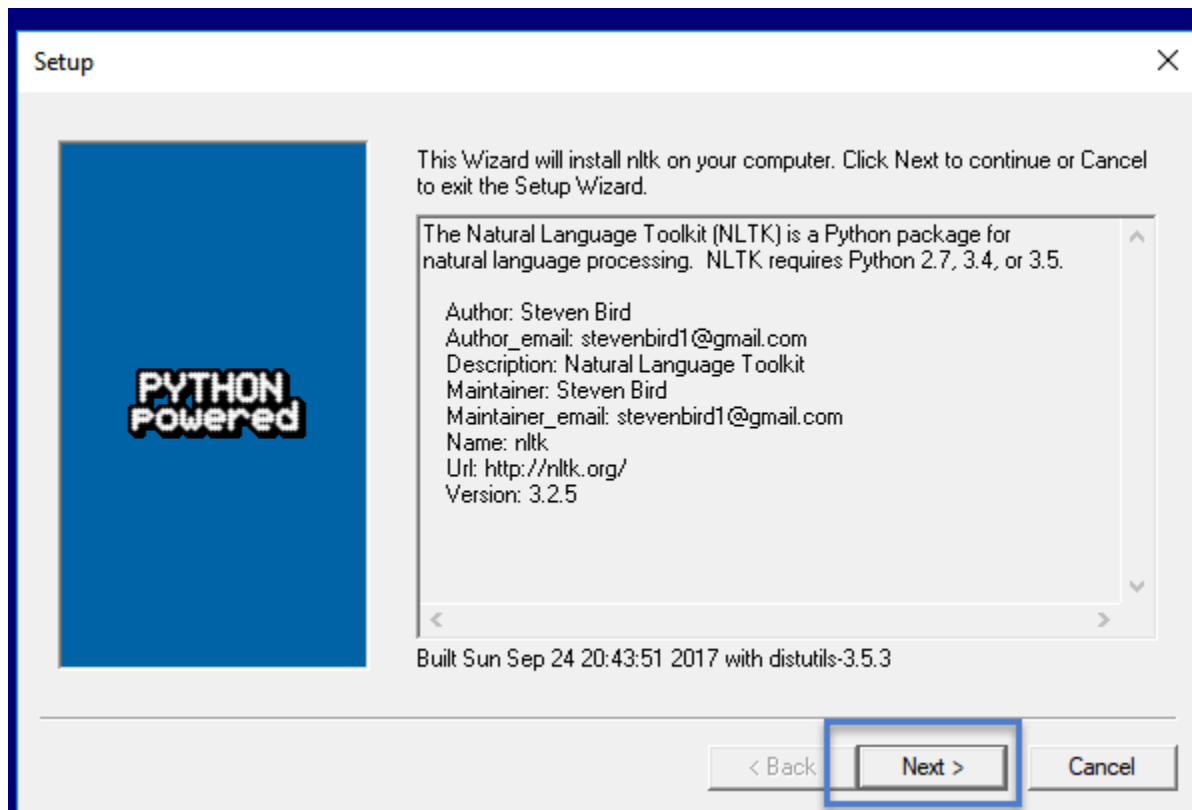
Natural Language Toolkit

Downloads

The Natural Language Toolkit (NLTK) is a Python package for natural language processing. NLTK requires Python 3.4, or 3.5.

File	Type	Py Version
nltk-3.2.5.tar.gz (md5)	Source	
nltk-3.2.5.win32.exe (md5)	MS Windows installer	any

Step 4: Click nltk-3.2.5 win32.exe(md5) file



Step 5: Click **Next**

Step 6: Click **Finish**

Step 7: Open command prompt and go to C:\Python34\Scripts

Step 8: **Enter** easy_install pip

Step 9: **Enter** pip install pyyaml nltk

```
Command Prompt
Collecting pyparsing!=2.0.4,>=1.5.6 (from matplotlib)
  Downloading pyparsing-2.2.0-py2.py3-none-any.whl (56kB)
    100% |#####| 57kB 1.2MB/s
Requirement already satisfied (use --upgrade to upgrade): six>=1.5 in c:\python34\lib\site-packages (from python-dateutil>matplotlib)
Installing collected packages: cyclar, pyparsing
Successfully installed cyclar-0.10.0 pyparsing-2.2.0
You are using pip version 7.1.2, however version 9.0.1 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.

C:\Python34\Scripts>install nltk
'install' is not recognized as an internal or external command,
operable program or batch file.

C:\Python34\Scripts>pip install pyyaml nltk
Collecting pyyaml
  Downloading PyYAML-3.12.tar.gz (253kB)
    100% |#####| 253kB 506kB/s
Collecting nltk
  Downloading nltk-3.2.5.tar.gz (1.2MB)
    100% |#####| 1.2MB 136kB/s
Requirement already satisfied (use --upgrade to upgrade): six in c:\python34\lib\site-packages (from nltk)
Installing collected packages: pyyaml, nltk
  Running setup.py install for pyyaml
  Running setup.py install for nltk
Successfully installed nltk-3.2.5 pyyaml-3.12
You are using pip version 7.1.2, however version 9.0.1 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.

C:\Python34\Scripts>
```

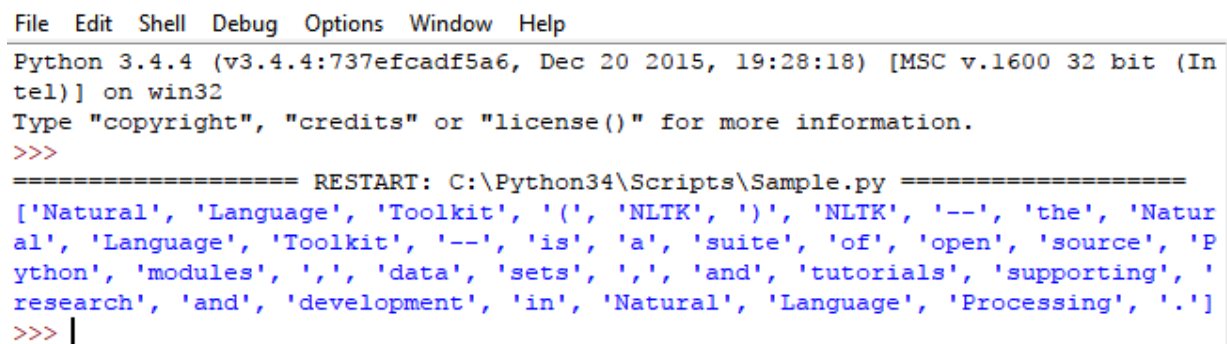
Step 10: Create a sample python script as described below. Please note that you need to download all the components for nltk.download. You can also go to the python prompt and complete nltk.download.

Sample Script

```
import nltk
nltk.download
from nltk.tokenize import sent_tokenize, word_tokenize
```

```
sample = "Natural Language Toolkit (NLTK) NLTK -- the Natural Language Toolkit -- is a
suite of open source Python modules, data sets, and tutorials supporting research and
development in Natural Language Processing."
print(word_tokenize(sample))
```

Output



The screenshot shows a Python 3.4.4 command prompt window. The title bar includes 'File', 'Edit', 'Shell', 'Debug', 'Options', 'Window', and 'Help'. The prompt text indicates the Python version and architecture: 'Python 3.4.4 (v3.4.4:737efcadf5a6, Dec 20 2015, 19:28:18) [MSC v.1600 32 bit (Intel)] on win32'. It also shows the standard copyright notice. The user has entered three empty lines ('>>>'). The prompt then displays the output of the script, which is a list of tokens from the NLTK download message, including 'Natural', 'Language', 'Toolkit', '(', 'NLTK', ')', 'NLTK', '--', 'the', 'Natural', 'Language', 'Toolkit', '--', 'is', 'a', 'suite', 'of', 'open', 'source', 'Python', 'modules', ',', 'data', 'sets', ',', 'and', 'tutorials', 'supporting', 'research', 'and', 'development', 'in', 'Natural', 'Language', 'Processing', '.'.

```
File Edit Shell Debug Options Window Help
Python 3.4.4 (v3.4.4:737efcadf5a6, Dec 20 2015, 19:28:18) [MSC v.1600 32 bit (Intel)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
>>>
===== RESTART: C:\Python34\Scripts\Sample.py =====
['Natural', 'Language', 'Toolkit', '(', 'NLTK', ')', 'NLTK', '--', 'the', 'Natural', 'Language', 'Toolkit', '--', 'is', 'a', 'suite', 'of', 'open', 'source', 'Python', 'modules', ',', 'data', 'sets', ',', 'and', 'tutorials', 'supporting', 'research', 'and', 'development', 'in', 'Natural', 'Language', 'Processing', '.']
>>> |
```