NLTK Installation

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum.

Thanks to a hands-on guide introducing programming fundamentals alongside topics in computational linguistics, plus comprehensive API documentation, NLTK is suitable for linguists, engineers, students, educators, researchers, and industry users alike. NLTK is available for Windows, Mac OS X, and Linux. Best of all, NLTK is a free, open-source, community-driven project.

The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical Natural Language Processing (NLP) for English written in the Python programming language.

Pre-Requisites

- A system running on Windows/Ubuntu APP/Ubuntu OS
- A user account with sudo/administration privileges
- Access to a terminal window/command-line

Before continuing with this tutorial, make sure you are logged in as root or a user with sudo/administration privileges.

In this tutorial, we will show you how to install NLTK module on Windows and Ubuntu.

1. Install NLTK module on Windows
2. Install NLTK module on Ubuntu APP (Windows 10) or Ubuntu OS
3. Install NLTK module on PyCharm

If you are working on the Windows system, please follow step 1, step 2, and step 3, but if you are working on Ubuntu OS, you follow only step 2 and step 3.

1. Install NLTK module on Windows

Windows systems typically do not have NLTK module build-in. Before installing the NLTK module, make sure you have installed Python 3+ version and upgraded Pip. Let us look at how to install NLTK module on Windows:

- Open a Command Terminal from Windows system and verify Python & Pip as follow:

```
$ python --version

$ python -m pip --version
```

```
C:\Program Files>
C:\Program Files>python --version
Python 3.7.7

C:\Program Files>python -m pip --version
pip 20.0.2 from C:\Users\somak\AppData\Roaming\Python\Python37\site-packages\pip (python 3.7)

C:\Program Files>
```

- To verify the NLTK module, go to Python Console and type as below:

```
>> import nltk

>> import numpy
```

```
Type "help", "copyright", "credits" or "license" for more information.
>>>
>>>
>>> import nltk
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
ModuleNotFoundError: No module named 'nltk'
>>>
>>> import numpy
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
ModuleNotFoundError: No module named 'numpy'
>>>
```

- If you do not have the NLTK module, then running the above command will give no module error as above. Now, to install the NLTK module from Command Terminal type the below command and also, we need to download NLTK data.

```
$ python -m pip install --user -U nltk

$ python -m pip install --user -U numpy

$ python -c "import nltk; nltk.download('all')"
```

```
C:\Program Files>
C:\Program Files>python -m pip install --user -U nltk
Requirement already up-to-date: nltk in c:\users\somak\appdata\roaming\python\python37\site-pac
Requirement already satisfied, skipping upgrade: click in c:\program files\python37\lib\site-pa
Requirement already satisfied, skipping upgrade: joblib in c:\users\somak\appdata\roaming\pytho
Requirement already satisfied, skipping upgrade: regex in c:\users\somak\appdata\roaming\python
Requirement already satisfied, skipping upgrade: tqdm in c:\users\somak\appdata\roaming\python\
WARNING: You are using pip version 20.0.2; however, version 20.1.1 is available.
You should consider upgrading via the 'C:\Program Files\Python37\python.exe -m pip install --up
```

```
C:\Program Files>python -m pip install --user -U numpy
Requirement already up-to-date: numpy in c:\program files\python37\lib\site-packages (1.18.5)
WARNING: You are using pip version 20.0.2; however, version 20.1.1 is available.
You should consider upgrading via the 'C:\Program Files\Python37\python.exe -m pip install --upg

C:\Program Files>python -c "import nltk; nltk.download('all')"
[nltk_data] Downloading collection 'all'
```

```
C:\Users\somak>python -c "import nltk; nltk.download('all')"
[nltk_data] Downloading collection 'all'
[nltk_data]    |
[nltk_data]    | Downloading package abc to
[nltk_data]    |     C:\Users\somak\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\abc.zip.
[nltk_data]    | Downloading package alpino to
[nltk_data]    |     C:\Users\somak\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\alpino.zip.
[nltk_data]    | Downloading package biocreative_ppi to
[nltk_data]    |     C:\Users\somak\AppData\Roaming\nltk_data...
[nltk_data]    |   Unzipping corpora\biocreative_ppi.zip.
[nltk_data]    | Downloading package brown to
[nltk_data]    |     C:\Users\somak\AppData\Roaming\nltk_data...
```

- Verify NLTK module as below:

```
C:\Users\somak>
C:\Users\somak>python
Python 3.8.3 (tags/v3.8.3:6f8c832, May 13 2020, 22:20:19) [MSC v.1925 32 bit (
Type "help", "copyright", "credits" or "license" for more information.
>>>
>>>
>>> import numpy
>>>
>>>
>>>
>>>
>>> import nltk
>>>
>>>
>>> exit()
```

2. **Install NLTK module on Ubuntu APP (Windows 10) or Ubuntu OS**

There are several ways you can install NLTK module on the Ubuntu system, and you can use any of the below process on Ubuntu APP or Ubuntu OS because all are the same.

1. Install NLTK module using Command Terminal
2. Install NLTK module using Shell Scripts

1. Install NLTK module using Command Terminal

To install the NLTK package with Command Terminal, open Command Terminal from the Ubuntu OS, or Ubuntu APP from Windows.

- First, go to Python console and verify it as below:

```
>> import nltk

>> import numpy
```

```
somak@LAPTOP-2QHNB62O:~$
somak@LAPTOP-2QHNB62O:~$ python3
Python 3.6.9 (default, Apr 18 2020, 01:56:04)
[GCC 8.4.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>>
>>>
>>> import nltk
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
ModuleNotFoundError: No module named 'nltk'
>>>
>>> import numpy
>>>
```

- Later, come out from the Python Console and run the below command to install NLTK module:

```
$ pip3 install --user -U nltk

$ pip3 install --user -U numpy

$ python3 -c "import nltk; nltk.download('all')"
```

```
somak@LAPTOP-2QHNB62O:~$
somak@LAPTOP-2QHNB62O:~$ pip3 install --user -U nltk
Collecting nltk
  Downloading https://files.pythonhosted.org/packages/92/75/ce35194d8e3022203cca0d2f896dbb88689f9b3fce8e9f9cff94291
    100% |████████████████████████████████| 1.4MB 814kB/s
Collecting click (from nltk)
  Using cached https://files.pythonhosted.org/packages/d2/3d/fa76db83bf75c4f8d338c2fd15c8d33fdd7ad23a9b5e57eb6c5de2
Collecting joblib (from nltk)
  Downloading https://files.pythonhosted.org/packages/b8/a6/d1a816b89aa1e9e96bcb298eb1ee1854f21662ebc6d55ffa3d7b3b5
    100% |████████████████████████████████| 307kB 2.0MB/s
Collecting regex (from nltk)
  Downloading https://files.pythonhosted.org/packages/f2/86/461fa59347051f378e661fcaf0a6d793450c779f4a29f338ae37c52
    100% |████████████████████████████████| 665kB 1.3MB/s
Collecting tqdm (from nltk)
```

```
somak@LAPTOP-2QHNB620:~$
somak@LAPTOP-2QHNB620:~$
somak@LAPTOP-2QHNB620:~$ pip3 install --user -U numpy
Collecting numpy
  Using cached https://files.pythonhosted.org/packages/b3/a9/b1bc4c935ed063766bce7d3e8
Installing collected packages: numpy
Successfully installed numpy-1.18.5
somak@LAPTOP-2QHNB620:~$
```

```
somak@LAPTOP-2QHNB620:~$
somak@LAPTOP-2QHNB620:~$
somak@LAPTOP-2QHNB620:~$
somak@LAPTOP-2QHNB620:~$ python3-c "import nltk; nltk.download('all')"
[nltk_data] Downloading collection 'all'
[nltk_data]    |
[nltk_data]    | Downloading package abc to /home/somak/nltk_data...
[nltk_data]    |   Unzipping corpora/abc.zip.
[nltk_data]    | Downloading package alpino to
[nltk_data]    |     /home/somak/nltk_data...
[nltk_data]    |   Unzipping corpora/alpino.zip.
[nltk_data]    | Downloading package biocreative_ppi to
[nltk_data]    |     /home/somak/nltk_data...
[nltk_data]    |   Unzipping corpora/biocreative_ppi.zip.
```

- Verify NLTK again as below:

```
somak@LAPTOP-2QHNB620:~$ python3
Python 3.6.9 (default, Apr 18 2020, 01:56:04)
[GCC 8.4.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>>
>>>
>>> import nltk
>>>
>>> import numpy
>>>
>>> exit()
```

2. Install NLTK module using Shell Script

First, we need to download the Robotic-Greeter folder from the Robotic-Greeter-GitHub link.

You can download it in two ways:

1. Clone it with Command Terminal
2. Download it as a Zip file

Inside of the Robotic-Greeter folder, we have the shell (Unix) script, which you need to run, and this script will automatically install NLTK on your computer.

1. **Using Clone method**

   Go to Ubuntu APP from Windows 10 or Command Terminal from Ubuntu OS and run the following command:
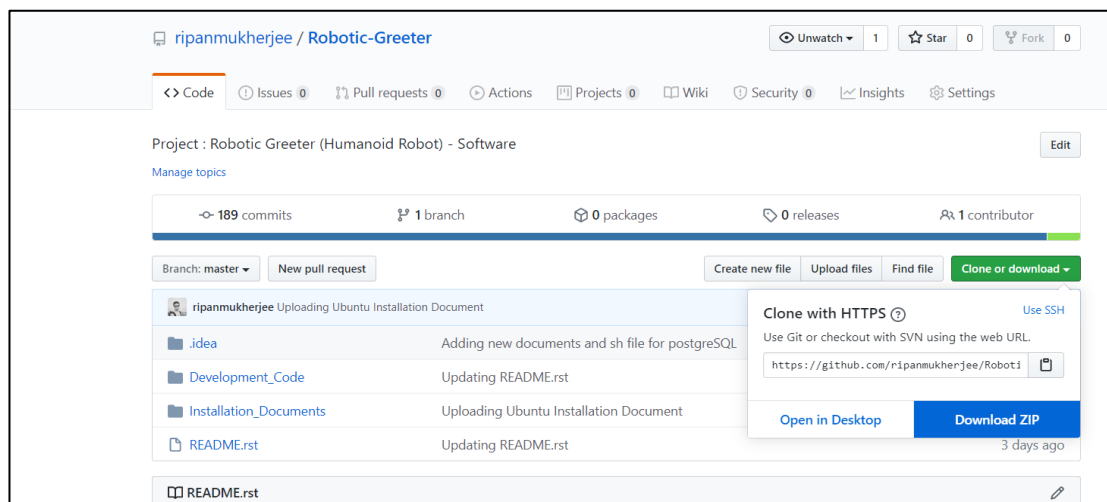
   ```
   $ git clone https://github.com/ripanmukherjee/Robotic-Greeter.git
   ```


   This command will automatically download the Robotic-Greeter folder on your computer.

2. **Download as Zip**

   Also, you can directly download the Zip file and Unzip it. Then it would be best if you put it in the proper directory or your project directory.

   Once the download is complete, please go to the following directory:

   ```
   $ cd Robotic-Greeter/Installation_Documents/NLTK_Installation
   ```

In this folder, you will get **NLTK_Installation.sh** script. To list the directories and files in this folder, run "ls -lrt" and later change the executable permission for the file with "chmod."

```
$ ls -lrt

$ chmod +x *.sh
```

```
somak@LAPTOP-2QHNB620:~$
somak@LAPTOP-2QHNB620:~$ cd Rob*
somak@LAPTOP-2QHNB620:~/Robotic-Greeter$ cd In*
somak@LAPTOP-2QHNB620:~/Robotic-Greeter/Installation_Documents$ cd NLTK*
somak@LAPTOP-2QHNB620:~/Robotic-Greeter/Installation_Documents/NLTK_Installation$
somak@LAPTOP-2QHNB620:~/Robotic-Greeter/Installation_Documents/NLTK_Installation$
somak@LAPTOP-2QHNB620:~/Robotic-Greeter/Installation_Documents/NLTK_Installation$ ls -lrt
total 4
-rwxrwxrwx 1 somak somak  326 Jun  6 20:23 NLTK_Installation.sh
-rw-rw-rw- 1 somak somak 1760 Jun  6 20:23 README.rst
somak@LAPTOP-2QHNB620:~/Robotic-Greeter/Installation_Documents/NLTK_Installation$
somak@LAPTOP-2QHNB620:~/Robotic-Greeter/Installation_Documents/NLTK_Installation$ chmod +x *.sh
somak@LAPTOP-2QHNB620:~/Robotic-Greeter/Installation_Documents/NLTK_Installation$
somak@LAPTOP-2QHNB620:~/Robotic-Greeter/Installation_Documents/NLTK_Installation$
```

After that, run the scripts as follow:

```
$ sh NLTK_Installation.sh
```

```
somak@LAPTOP-2QHNB620:~/Robotic-Greeter/Installation_Documents/NLTK_Installation$
somak@LAPTOP-2QHNB620:~/Robotic-Greeter/Installation_Documents/NLTK_Installation$ sh NLTK_Installation.sh
Starting Installing NLTK!!
Collecting nltk
  Downloading nltk-3.5.zip (1.4 MB)
     |                                    | 1.4 MB 1.1 MB/s
Requirement already satisfied, skipping upgrade: click in /usr/lib/python3/dist-packages (from nltk) (7.0)
Collecting joblib
  Downloading joblib-0.15.1-py3-none-any.whl (298 kB)
     |                                    | 298 kB 2.1 MB/s
Collecting regex
  Downloading regex-2020.6.8-cp38-cp38-manylinux2010_x86_64.whl (673 kB)
```
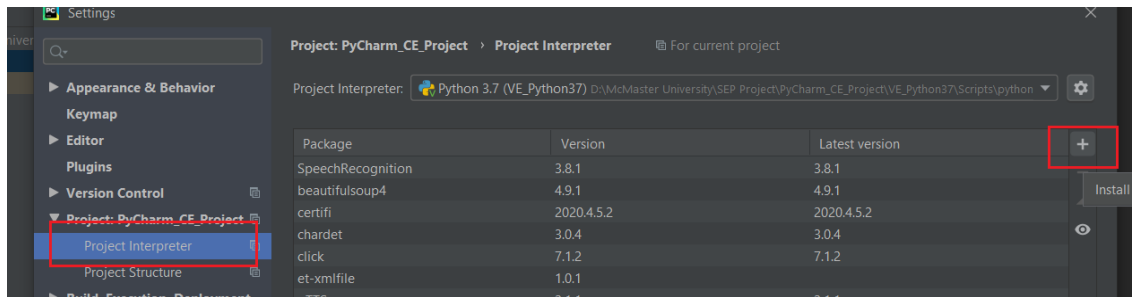
Later, you can verify NLTK again as below:

```
somak@LAPTOP-2QHNB620:~$ python3
Python 3.6.9 (default, Apr 18 2020, 01:56:04)
[GCC 8.4.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>>
>>>
>>> import nltk
>>>
>>> import numpy
>>>
>>> exit()
```
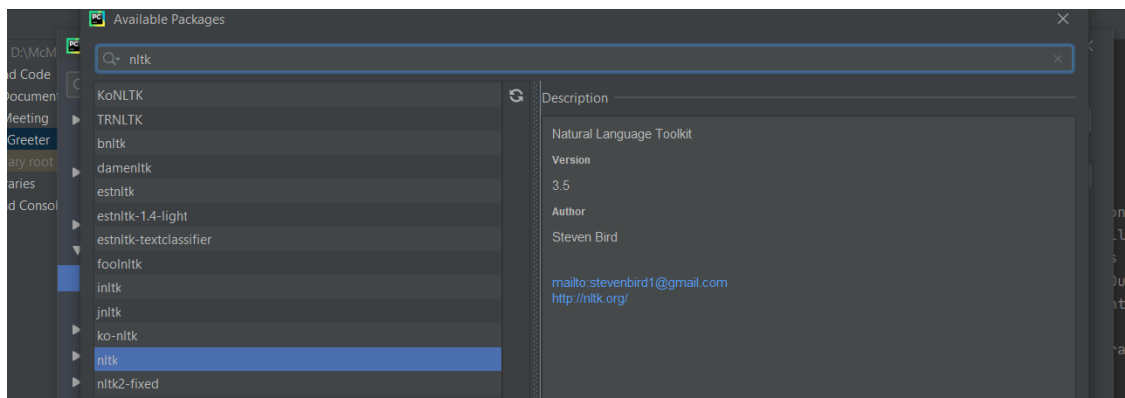
3. Install NLTK on PyCharm

If you wish to run or execute the code from PyCharm CE, and PyCharm CE gives error regarding NLTK module, you can also install it from PyCharm Packages as below:
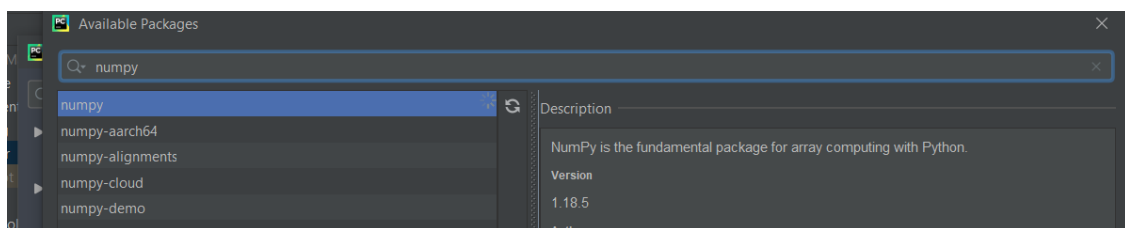
- Go to Settings and click on Project Interpreter.

- In the Project Interpreter, click on "+" to add packages.

- In the Available Packages tab, you can search for "nltk" and install it.

- In the Available Packages tab, you can search for "numpy" and install it.

For more details related to NLKT, please visit the [NLTK](#) Website.