* My Python Path to install lxml:

>C:\Users\priyanka\AppData\Local\Programs\Python\Python36-32\Scripts\pip install lxml

* Check the version of Python and lxml

>python

>>> from lxml import etree

>>> etree.LXML\_VERSION (3, 0, -198, 0)

* Install NLTK

Pip install nltk

Python

Import nltk

nltk.download(‘all’)

* **OS Module:** The OS module in Python provides a way of using operating system dependent functionality. The functions that the OS module provides allows you to interface with the underlying operating system that Python is running on – be that Windows, Mac or Linux. E,g, Return a list of the entries in the directory given by path - os.listdir(path)

Source: <http://www.pythonforbeginners.com/os/pythons-os-module>

* **isalpha()** - In Python, **isalpha()** is a built-in method used for string handling. The isalpha() methods returns “True” if all characters in the string are alphabets, Otherwise, It returns “False”.
* **cleaner()** - to clean the tags e.g. <style>, <meta> etc