Operating System: Windows 10

Python:2.7.10

IDE: Sublime Text 3

Packages: BeautifulSoup, requests, PrettyTable

For Windows machine:-

----------------------------

Check whether python is installed by using the following command in command prompt:-

python

If not installed, link for installing Python:-

https://www.python.org/download/releases/2.7.1

Check whether BeautifulSoup is installed by typing into the command prompt:

pyhton -c "import bs4"

If not, install it using the below command:

pip install beautifulsoup4

Check whether requests is installed by typing into the command prompt:

pyhton -c "import requests"

If not, install it using the below command:

pip install requests

Install nltk using the following commands in cmd:-

pip install nltk

nltk.download()

install prettytable using the following command in cmd:

pip install prettytable

After the packages have been installed, run the program from this folder:

Task 1:

=======

Run the following program:

Task1\_gettokens.py

The output files will be created in the same folder

Task 2:

======

Unigrams:

---------------

Run the following program on the output of Task1:

Task2\_unigrams\_program

The TermFrequency table will be created in the same folder as

unigram\_index1.txt

The DocumentFrequency table will be created in the same folder as

unigram\_doc\_freq.txt

Bigrams:

---------------

Run the following program on the output of Task1:

Task2\_bigrams\_program

The TermFrequency table will be created in the same folder as

bigram\_tf.txt

The DocumentFrequency table will be created in the same folder as

bigram\_doc\_freq.txt

Trigrams:

---------------

Run the following program on the output of Task1:

Task2\_trigrams\_program

The TermFrequency table will be created in the same folder as

trigram\_tf.txt

The DocumentFrequency table will be created in the same folder as

trigram\_doc\_freq.txt

Design Choices:

==============