Assignment#8

Collect the screenshots and original files for the following activities based on NLTK.

Activity# 1

 -Word Frequencies in the given text.

- No use of NLTK yet
- Function to create a dictionary of word frequencies of the text.
- - key is the word
- - value is the frequency (count) of the word in the text

Activity#2

Web Scraping

- Read the text from the web.
- Tokenize the text.
- Display the result in binary form.

Activity#3

Tokenize using the 're' module (Regular Expressions)

- Convert the binary text to string.
- Tokenize and display the first 100 elements.

Activity#4

Tokenize using NLTK

- Clean the HTML text using beautiful soup
- Tokenize and display the elements.

Activity#5

Combine the word frequency program to find the frequencies of the word on a Web page.

- Using NLTK to produce and draw frequencies.

Activity#6

Additional activities

- Part of speech (POS) tagging.
- Identify named entities (Named entity recognition)
- Display a parse tree.