Student Name: Satwik Hosamani

SUID: 378000248

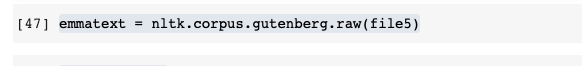
LAB TITLE: Installing nltk

Book Title: Poems by William Blake

Graphical user interface, application

Description automatically generated with medium confidence

1. To convert the data into string so that it can be used for processing use nltk.corpus.gutenberg.raw(),for convert into string we use the raw() function



1. Nltk.word\_tokenize() is used to acquire the tokens. After having the string to break it into words, this is the next move that is done using the word\_tokenize() function.

Graphical user interface, text, application

Description automatically generated

1. By using w.lower() to lowercase the tokens,we get the words
2. FreqDist is used for frequency distribution

The FreqDist is a feature in the library of nltk that takes words and outputs words and their frequency.

1. With most common(30), get the 30 top frequency words and print the word, frequency pairs.

The function most-common() returns the most common terms and their count[3].

The top 30 frequent words are:

A picture containing table

Description automatically generated

Observations:

The filler words which are required to make a sense in the sentence cover the major area of the top frequent words.

Lessons learned:

1. To make the data ready for processing using the raw() functions
2. To retrieve words from the string using the word\_tokenize() function
3. To use FreqDist to construct a frequency distribution of terms and to search the most common words in the distribution using the most\_common() function.