# CODE:

Text and Speech Analysis-**CCS369**

**Observation (EXP-3)**

import nltk

from nltk.corpus import gutenberg

# Step 2: Download the gutenberg corpus if not already downloaded nltk.download('gutenberg')

# Step 3: Get a list of file IDs in the gutenberg corpus file\_ids = gutenberg.fileids()

# Step 4: Print the first 5 file IDs

print("First 5 file IDs in Gutenberg corpus:") print(file\_ids[:5])

# Step 5: Get the raw text of the first book in the Gutenberg corpus first\_book\_raw\_text = gutenberg.raw(file\_ids[0])

# Step 6: Print the first 500 characters of the raw text print("\nFirst 500 characters of the first book:") print(first\_book\_raw\_text[:500])

# Output:

First 500 characters of the first book:

[Emma by Jane Austen 1816] VOLUME I

CHAPTER I

Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy disposition, seemed to unite some of the best blessings

of existence; and had lived nearly twenty-one years in the world with very little to distress or vex her.

She was the youngest of the two daughters of a most affectionate, indulgent father; and had, in consequence of her sister's marriage, been mistress of his house from a very early period. Her mother had died t

[nltk\_data] Downloading package gutenberg to /root/nltk\_data... [nltk\_data] Package gutenberg is already up-to-date!