**Write a program for the Information Retrieval System using appropriate NLP tools (such as NLTK, Open NLP, …) a. Text tokenization b. Count word frequency c. Remove stop words d. POS tagging**

**Name:** Yogita Sunil Khalate

**Roll No:** 3335 **Class:** TEIT

**CODE:**

import nltk

nltk.download('punkt')

from nltk.corpus import stopwords

nltk.download('stopwords')

from nltk.tokenize import word\_tokenize

from nltk.probability import FreqDist

from nltk.tag import pos\_tag

nltk.download('averaged\_perceptron\_tagger')

text = input("Enter Your Text: ")

print("Text: " +text)

words = word\_tokenize(text)

print(" Tokenized Words: ")

print(words)

words = [word.lower() for word in words]

fdist = FreqDist(words)

print(" Word Frequency: ")

for word, freq in fdist.items():

    print(f"{word}: {freq}")

stop\_words = set(stopwords.words('english'))

filtered\_words = [words for word in words if word.casefold() not in stop\_words]

print("Filtered Words: ")

print(filtered\_words)

pos\_tags = pos\_tag(words)

print("POS Tags: ")

print(pos\_tags)

**OUTPUT:**

[nltk\_data] Downloading package punkt to /root/nltk\_data...

[nltk\_data] Unzipping tokenizers/punkt.zip.

[nltk\_data] Downloading package stopwords to /root/nltk\_data...

[nltk\_data] Unzipping corpora/stopwords.zip.

[nltk\_data] Downloading package averaged\_perceptron\_tagger to

[nltk\_data] /root/nltk\_data...

[nltk\_data] Unzipping taggers/averaged\_perceptron\_tagger.zip.

Enter Your Text: Hi this is yogita khalate

Text: Hi this is yogita khalate

Tokenized Words:

['Hi', 'this', 'is', 'yogita', 'khalate']

Word Frequency:

hi: 1

this: 1

is: 1

yogita: 1

khalate: 1

Filtered Words:

[['hi', 'this', 'is', 'yogita', 'khalate'], ['hi', 'this', 'is', 'yogita', 'khalate'], ['hi', 'this', 'is', 'yogita', 'khalate']]

POS Tags:

[('hi', 'NN'), ('this', 'DT'), ('is', 'VBZ'), ('yogita', 'JJ'), ('khalate', 'NN')]