

```
#Problem Statement: Write a program for the In
formation Retrieval System
#using appropriate NLP tools (such as NLTK, Op
en NLP, ...) a. Text tokenization
#b. Count word frequency c. Remove stop words
d. POS tagging
#Name:Ghanishtha Anil Rane
#Class:TEIT
#Roll no:3363
```

```
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 w
ord.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: I am Ghanishtha Rane. I am 20 years old. i am from Bvcoew

Text: I am Ghanishtha Rane. I am 20 years old. i am from Bvcoew

Tokenized Words:

['I', 'am', 'Ghanishtha', 'Rane', '.', 'I', 'am', '20', 'years', 'old', '.', 'i', 'am', 'from', 'Bvcoew']

Word Frequency:

i: 3

am: 3

ghanishtha: 1

rane: 1

.: 2

20: 1

years: 1

old: 1

from: 1

bvcoew: 1

Filtered Words:

[['i', 'am', 'ghanishtha', 'rane', '.', 'i', 'am', '20', 'years', 'old', '.', 'i', 'am', 'from', 'bvcoew'], ['i', 'am', 'ghanishtha', 'rane', '.', 'i', 'am', '20', 'years', 'old', '.', 'i', 'am', 'from', 'bvcoew'], ['i', 'am', 'ghanishtha', 'rane', '.', 'i', 'am', '20', 'years', 'old', '.', 'i', 'am', 'from', 'bvcoew'], ['i', 'am', 'ghanishtha', 'rane', '.', 'i', 'am', '20', 'years', 'old', '.', 'i', 'am', 'from', 'bvcoew'], ['i', 'am', 'ghanishtha', 'rane', '.', 'i', 'am', '20', 'years', 'old', '.', 'i', 'am', 'from', 'bvcoew'], ['i', 'am', 'ghanishtha', 'rane', '.', 'i', 'am', '20', 'years', 'old', '.', 'i', 'am', 'from', 'bvcoew'], ['i', 'am', 'ghanishtha', 'rane', '.', 'i', 'am', '20', 'years', 'old', '.', 'i', 'am', 'from', 'bvcoew'], ['i', 'am', 'ghanishtha', 'rane', '.', 'i', 'am', '20', 'years', 'old', '.', 'i', 'am', 'from', 'bvcoew'], ['i', 'am', 'ghanishtha', 'rane', '.', 'i', 'am', '20', 'years', 'old', '.', 'i', 'am', 'from', 'bvcoew']]

POS Tags:

[('i', 'NN'), ('am', 'VBP'), ('ghanishtha', 'JJ'), ('rane', 'NN'), ('.', '.'), ('i', 'VB'), ('am', 'VBP'), ('20', 'CD'), ('years', 'NNS'), ('old', 'JJ'), ('.', '.'), ('i', 'VB'), ('am', 'VBP'), ('from', 'IN'), ('bvcoew', 'NN')]