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')]