

Name: Aniket Sarjerao Sable

Roll No: 54

=====

AI Practical 04

Problem Statement:

Write a program for the Information Retrieval System using
appropriate NLP tools (such as NLTK, OpenNLP, etc.) and perform
the following operations:

a. Text Tokenization

b. Count Word Frequency

c. Remove Stop Words

d. POS Tagging

=====

Importing Required Libraries

import nltk

from nltk import word_tokenize, FreqDist, pos_tag

from nltk.corpus import stopwords

Downloading Required NLTK Resources

nltk.download('punkt')

nltk.download('stopwords')

nltk.download('averaged_perceptron_tagger')

NLP Operations

a. Function for Tokenization

def **text_tokenization**(text):

return word_tokenize(text)

b. Function to Count Word Frequency

def **count_word_frequency**(tokens):

return FreqDist(tokens)

c. Function to Remove Stop Words

def **remove_stop_words**(tokens):

stop_words = set(stopwords.words('english'))

return [word for word in tokens if word.lower() not in stop_words]

d. Function for Part-of-Speech Tagging

def **pos_tagging**(tokens):

return pos_tag(tokens)

Example Text for Demonstration

text = "Natural Language Processing is a fascinating field of Artificial Intelligence."

User Input for Operation Choice

```
operation = input("Select operation (a. Tokenization, b. Word Frequency, c. Remove Stop Words, d. POS Tagging): ").lower()
```

```
# -----
```

```
# Decision Making and Output Based on User Input
```

```
# -----
```

```
if operation == 'a':
```

```
    tokens = text_tokenization(text)
```

```
    print("Tokens:", tokens)
```

```
elif operation == 'b':
```

```
    tokens = text_tokenization(text)
```

```
    word_frequency = count_word_frequency(tokens)
```

```
    print("Word Frequency:")
```

```
    for word, freq in word_frequency.items():
```

```
        print(f"{word}: {freq}")
```

```
elif operation == 'c':
```

```
    tokens = text_tokenization(text)
```

```
    filtered_tokens = remove_stop_words(tokens)
```

```
    print("Tokens without Stop Words:", filtered_tokens)
```

```
elif operation == 'd':
```

```
    tokens = text_tokenization(text)
```

```
    pos_tags = pos_tagging(tokens)
```

```
    print("POS Tags:")
```

```
    for word, tag in pos_tags:
```

```
        print(f"{word} → {tag}")
```

```
else:
```

```
    print("Invalid operation. Please select a, b, c, or d.")
```

```
# -----
```

```
# Sample Output:
```

```
# Select operation (a. Tokenization, b. Word Frequency, c. Remove Stop Words, d. POS Tagging): d
```

```
# POS Tags:
```

```
# Natural → JJ
```

```
# Language → NNP
```

```
# Processing → NNP
```

```
# is → VBZ
```

```
# a → DT
```

```
# fascinating → JJ
```

```
# field → NN
```

```
# of → IN
```

```
# Artificial → JJ
```

```
# Intelligence → NNP
```

```
# . → .
```

```
# -----
```