# Name: Shantanu Rohile

# Roll No: 53

*# ==============================================*

# 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 # . → .

*#*