**Lab 04**

**AI102L**

**Task1: Text Processing & Word Frequency Analyzer with Stopwords Removal**

**Objective:**

Write a Python program that processes a paragraph of text, removes stopwords, counts word frequencies, and performs various operations using strings, lists, dictionaries, loops, and higher-order functions like map() and filter(). The program should also include an extra challenge to filter words starting with vowels using lambda functions.

**Requirements:**

1. **Input:**
   * Accept a paragraph of text from the user.
2. **Text Processing:( User defined function)**
   * Remove all punctuation from the text.
   * Convert the text to lowercase.
   * Split the text into individual words.
   * Use a predefined list of stopwords (provided below) to filter out common words that do not contribute significantly to the meaning of the text.
3. **Word Frequency Counting:( User defined function)**
   * Count the frequency of each unique word in the processed text.
   * Store the word frequencies in a dictionary.
4. **Display Top 5 Most Frequent Words:( User defined function)**
   * Sort the dictionary by frequency in descending order.(use lambda ,reverse=True)
   * Display the top 5 most frequent words along with their counts.

All the above three tasks should be a separate user-defined function. The below tasks will be solved in the main() function, add stopword here in the main function and use it in the preprocessing function. The above functions must be called from the main function to ensure modularity in our program and give the program a specific app structure.

1. **Use map() to Convert Words to Uppercase:**
   * Apply the map() function to convert all words in the processed list to uppercase.
2. **Use filter() to Extract Words Appearing More Than Twice:**
   * Apply the filter() function to extract words that appear more than twice in the frequency dictionary.
3. **Extra Challenge: Filter Words Starting with a Vowel:**
   * Use a lambda function with filter() to extract words that start with a vowel (a, e, i, o, u).
4. **Output:**
   * Display the following:
     + Top 5 most frequent words.
     + List of words converted to uppercase.
     + Words appearing more than twice.
     + Words starting with a vowel.

stopwords = [

"the", "a", "an", "is", "are", "it", "that", "this", "these", "those",

"i", "you", "he", "she", "we", "they", "me", "him", "her", "us", "them",

"my", "your", "his", "her", "our", "their", "mine", "yours", "hers", "ours", "theirs",

"and", "but", "or", "nor", "for", "so", "yet", "as", "while", "since", "because",

"in", "on", "at", "by", "with", "about", "against", "through", "during", "before", "after",

"to", "from", "of", "at", "into", "throughout", "among", "between", "over", "under",

"have", "has", "had", "do", "does", "did", "will", "would", "can", "could", "should", "may", "might", "must",

"be", "being", "been", "am", "is", "are", "was", "were",

"not", "no", "nor", "neither", "never", "always", "sometimes", "often", "usually", "generally",

"more", "less", "many", "few", "some", "all", "most", "every", "each", "both", "several",

"one", "two", "three", "four", "five", "six", "seven", "eight", "nine", "ten",

"said", "says", "saying", "told", "tells", "telling",

"go", "goes", "going", "went", "gone",

"get", "gets", "getting", "got", "gotten",

"make", "makes", "making", "made",

"know", "knows", "knowing", "knew", "known",

"think", "thinks", "thinking", "thought",

"take", "takes", "taking", "took", "taken",

"see", "sees", "seeing", "saw", "seen",

"come", "comes", "coming", "came", "come",

"want", "wants", "wanting", "wanted",

"look", "looks", "looking", "looked",

"use", "uses", "using", "used",

"find", "finds", "finding", "found",

"tell", "tells", "telling", "told",

"ask", "asks", "asking", "asked",

"i'm", "you're", "he's", "she's", "it's", "we're", "they're",

"i've", "you've", "we've", "they've",

"i'd", "you'd", "he'd", "she'd", "we'd", "they'd",

"i'll", "you'll", "he'll", "she'll", "we'll", "they'll"

]

**Sample Input**

Take any paragraph as input

Enter a paragraph of text: Hello world! Hello everyone. This is a test. Hello world, this is only a test. If this is a test, then let's test it!

**Output:**

Top 5 most frequent words:

hello: 3

test: 3

world: 2

everyone: 1

only: 1

Words in uppercase: ['HELLO', 'WORLD', 'HELLO', 'EVERYONE', 'TEST', 'HELLO', 'WORLD', 'TEST', 'ONLY', 'TEST', 'TEST']

Words appearing more than twice: ['hello', 'test']

Words starting with a vowel: ['everyone', 'only', 'it']

**Task 2: Using the lambda function perform the following task: take an input, squares it, add 5, then return the answer**

**Task1 Psuedo**

import string def process\_text(text, stopwords): # Remove punctuation and convert to lowercase text = text.translate(str.replace('', '', string.punctuation)).lower() # Split into words

# Remove stopwords  
  
return words

def count\_word\_frequency(words): # Count frequency of each word

return frequency

def display\_top\_words(frequency, n=5): # Sort by frequency in descending order and display top n words

def main(): # Define a list of stopwords manually stopwords = ["the", "a", "an", "is", "are", "it", "that", "this", "these", "those", "i", "you", "he", "she", "we", "they", "me", "him", "her", "us", "them", "my", "your", "his", "her", "our", "their", "mine", "yours", "hers", "ours", "theirs", "and", "but", "or", "on","nor", "for", "so", "yet", "as", "while", "since", "because", "i'll", "you'll", "he'll", "she'll", "we'll", "they'll" # Added common contractions ]

# Accept input from the user  
text = input("Enter a paragraph of text: ")  
  
# Process the text, including stop word removal(call function)  
  
  
# Count word frequency(call function)  
  
  
# Display top 5 most frequent words(call function)  
  
  
# Use map() to convert all words to uppercase  
  
  
  
# Use filter() to extract words that appear more than twice  
  
  
# Extra Challenge: Filter words that start with a vowel using lambda

main()