**Kulvir Singh**

**19BCE2074**

**Experiment 2**

*Question 3*

***Problem Statement :***

Write a program (using nltk toolkit in python environment) to tokenize

a) Sentence

b) A paragraph

***Procedure :***

Install the nltk toolkit in python environment using the pip install command. Import nltk to the code file. Download the class stopwords and punkt from nltk toolkit. Import stropwords from nltk.corpus and word\_tokenize from nltk.tokenize. Open a file using open method and store its contents in a variable using read method. Create a variable which stores the stopwords that are fetched from nltk. Use the tokenize method to create tokens of the document/file read. Loop through the tokens and filter out the stopwords. Display the tokens and filtered paragraph.

***Code :***

!pip install nltk

import nltk

nltk.download('stopwords')

nltk.download('punkt')

from nltk.corpus import stopwords

from nltk.tokenize import word\_tokenize

f = open('demo.txt','r')

paragraph = f.read()

stop\_words = set(stopwords.words('english'))

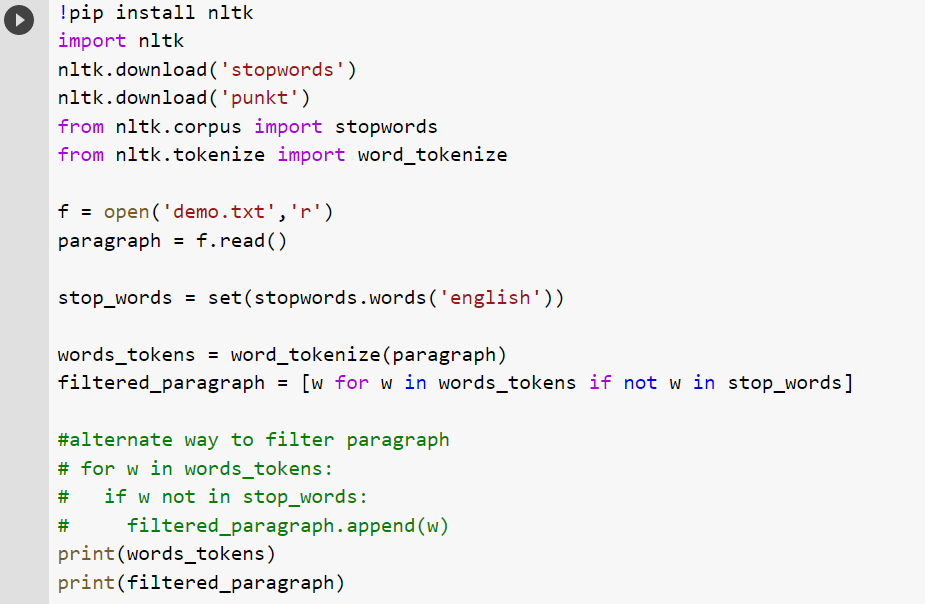
words\_tokens = word\_tokenize(paragraph)

filtered\_paragraph = [w for w in words\_tokens if not w in stop\_words]

print(words\_tokens)

print(filtered\_paragraph)

***Code Screenshot:***



***Output Screenshots :***

