Name: Manav Mehta

Roll No.: 41445   **Batch:** R4   **Class:** BE-IV

Course: Information Retrieval

Assignment 3

i. Code -

```python
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize

text = """
C is a widely-used, general-purpose programming language originally developed by Dennis
Ritchie at Bell Labs in 1972. It is a foundational language that has influenced many others,
including C++, Java, and Python. Known for its simplicity, efficiency, and control over system
resources, C is particularly favored in systems programming, operating systems, and
embedded systems. It provides low-level memory manipulation, making it both powerful and
flexible. The language is procedural, meaning programs are organized into functions and
data is passed through arguments. C's portability makes it ideal for cross-platform
applications, and its performance optimization capabilities make it suitable for resource-
constrained environments. Despite its long history, C remains relevant today due to its
minimalistic design, giving programmers fine-grained control over hardware and memory,
allowing efficient code execution.

The language also introduces a comprehensive standard library, which aids in managing
tasks like input/output and mathematical computations. C's popularity in the development of
compilers and operating systems is notable. Programs like Unix, Linux, and Windows
kernels are written in C, making it indispensable in the tech industry. The language's
structure is compact, allowing clear and maintainable code. However, it requires caution due
to the absence of some modern safety features, such as automatic memory management,
which leads to risks like buffer overflows and memory leaks. Nonetheless, C's combination
of versatility and control ensures it remains a critical tool for both software and hardware
development.
"""
text = text.lower()

words = text.split()
stop_words = set(stopwords.words("english"))
filtered_words = [word for word in words if word.lower() not in stop_words]
cleaned_text = " ".join(filtered_words)

print("Original Text:", text)
print("---" * 20)
print("Text after Stop Word Removal:", cleaned_text)
```

ii. Output -

Text after Stop Word Removal: c widely-used, general-purpose programming language originally developed dennis ritchie bell labs 1972. foundational language influenced many others, including c++, java, python. known simplicity, efficiency, control system resources, c particularly favored systems programming, operating systems, embedded systems. provides low-level memory manipulation, making powerful flexible. language procedural, meaning programs organized functions data passed arguments. c's portability makes ideal cross-platform applications, performance optimization capabilities make suitable resource-constrained environments. despite long history, c remains relevant today due minimalistic design, giving programmers fine-grained control hardware memory, allowing efficient code execution. language also introduces comprehensive standard library, aids managing tasks like input/output mathematical computations. c's popularity development compilers operating systems notable. programs like unix, linux, windows kernels written c, making indispensable tech industry. language's structure compact, allowing clear maintainable code. however, requires caution due absence modern safety features, automatic memory management, leads risks like buffer overflows memory leaks. nonetheless, c's combination versatility control ensures remains critical tool software hardware development.

PS C:\Users\neils\OneDrive\Desktop\PICT Documents and Stuff\PICT CourseWork BE 2024-25>