Name: Manav Mehta

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

Course: Information Retrieval

Assignment 1

i. Code -

```
from sklearn.feature_extraction.text import TfidfVectorizer
import numpy as np

doc1 = [
    """The gentle rustling of leaves in the morning breeze carried a sense of calm over the
meadow.
Birds flitted between the trees, their songs creating a melodic backdrop to the stillness of the
scene.
As the sun crept higher into the sky, its golden rays illuminated the dew-kissed grass, casting
a radiant glow over the landscape.
The beauty of the moment was fleeting yet profound, a reminder of nature's quiet,
unwavering presence."""
]

doc2 = [
    """In the heart of the bustling metropolis, neon signs flickered against the night sky,
illuminating the crowded streets below.
The steady hum of traffic was punctuated by the blaring of horns and the chatter of
pedestrians rushing from one place to another.
Skyscrapers loomed overhead, their glass windows reflecting the vibrant energy of the city.
This was a world that never slept, always in motion, driven by ambition and a relentless pace
that left little room for pause."""
]

doc3 = [
    """The soft light of dawn filtered through the canopy, casting dappled shadows on the
forest floor.
A stream murmured quietly nearby, its clear water reflecting the hues of the early morning
sky.
Birds sang their gentle chorus, and the fragrance of wildflowers filled the air with a delicate
sweetness.
Every element of the forest seemed to exist in perfect harmony, creating a sense of peace
that flowed through the stillness like a whispered secret of the earth."""
]

corpus = [doc1[0], doc2[0], doc3[0]]

vectorizer = TfidfVectorizer()
vectorizer.fit(corpus)
x1 = vectorizer.transform(doc1).todense()[0]
x2 = vectorizer.transform(doc2).todense()[0]
```

```
x3 = vectorizer.transform(doc3).todense()[0]

def cosine_similarity(p1, p2):
    return np.dot(p1, p2.T) / (np.linalg.norm(p1, 2) * np.linalg.norm(p2, 2))

print("Cosine similarity between doc1 and doc2 : ", cosine_similarity(x1, x2)[0][0])
print("Cosine similarity between doc1 and doc3 : ", cosine_similarity(x1, x3)[0][0])
```

ii. Output -

```
PROBLEMS    OUTPUT    DEBUG CONSOLE    TERMINAL    PORTS    POSTMAN CONSOLE

PS C:\Users\neils\OneDrive\Desktop\PICT Documents and Stuff\PICT_CourseWork_BE_2024-25> python -u "c:\Users\neils\OneDrive\Desktop\PICT Documents and Stuff\PI
CT_CourseWork_BE_2024-25\PICT_CourseWork_BE_2024-25\LP-IV\IR\Assignment1.py"
Cosine similarity between doc1 and doc2 :  [[0.48011105]]
Cosine similarity between doc1 and doc3 :  [[0.56188431]]
PS C:\Users\neils\OneDrive\Desktop\PICT Documents and Stuff\PICT_CourseWork_BE_2024-25>
```