

Name: Manav Mehta

Roll No.: 41445

Batch: R4

Class: BE-IV

Course: Information Retrieval

Assignment 2

i. Code -

```
import numpy as np
from nltk.corpus import stopwords

cachedStopWords = stopwords.words("english")

def create_links(text):
    text = text.lower()
    text = " ".join([word for word in text.split() if word not in cachedStopWords])
    sentences = text.split(".")
    sentence_words = [set(s.split()) for s in sentences]
    links = {}
    for i, si in enumerate(sentence_words):
        links[i] = []
        for j, sj in enumerate(sentence_words):
            if len(si.intersection(sj)) > 0:
                links[i].append(j)
    for i in range(len(sentence_words)):
        if i in links[i]:
            links[i].remove(i)
    return links, len(sentence_words), sentences

def create_adjacency_matrix(links, num_docs):
    adj_matrix = np.zeros((num_docs, num_docs))
    for doc, out_links in links.items():
        if out_links: # Avoid division by zero if no outgoing links
            for link in out_links:
                adj_matrix[link][doc] = 1 / len(out_links)
    return adj_matrix

def page_rank(links, num_docs, d=0.85, max_iters=100, eps=1e-6):
    ranks = np.ones(num_docs) / num_docs
    adj_matrix = create_adjacency_matrix(links, num_docs)
    teleport = np.ones(num_docs) / num_docs
    for i in range(max_iters):
        new_ranks = d * adj_matrix @ ranks + (1 - d) * teleport
        if np.linalg.norm(new_ranks - ranks, ord=1) < eps:
            break
    ranks = new_ranks
    return ranks
```

```
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.

"""

```
links, num_docs, sentences = create_links(text)
```

```
ranks = page_rank(links, num_docs)
```

```
print("Score for each sentence: ")
```

```
for rank, sentence in zip(ranks, sentences):
```

```
    print(rank, " : ", sentence)
```

```
print("Score for each sentence sorted by rank (descending): ")
```

```
rank_sorted_sentences = list(sorted(zip(ranks, sentences), key=lambda x: x[0],  
reverse=True))
```

```
for rank, sentence in rank_sorted_sentences:
```

```
    print(rank, " : ", sentence)
```

```
print("Summarized document: \n\n")
```

```
k = int(len(rank_sorted_sentences) * 0.7)
```

```
print(" ".join([ s[1] for s in rank_sorted_sentences[0:k]]))
```

```
print("\n\n")
```

ii. Output -

```
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\PICT Coursework_BE_2024-25\PICT Coursework_BE_2024-25\LP-IV\IR\Assignment2.py"
Score for each sentence:
0.0934996335755438 : c widely-used, general-purpose programming language originally developed dennis ritchie bell labs 1972
0.058395809696538584 : foundational language influenced many others, including c++, java, python
0.08037830790031476 : known simplicity, efficiency, control system resources, c particularly favored systems programming, operating systems, embedded systems
0.04413286742384233 : provides low-level memory manipulation, making powerful flexible
0.07549192294905227 : language procedural, meaning programs organized functions data passed arguments
0.0471104935149445 : c's portability makes ideal cross-platform applications, performance optimization capabilities make suitable resource-constrained environments
0.10219337075192399 : despite long history, c remains relevant today due minimalistic design, giving programmers fine-grained control hardware memory, allo
```