Text Summarization (Extractive Method)

Text summarization is a process of creating concise versions of the original text while retaining key information. Humans are naturally good summarizers for we have the ability to understand the overall meaning of a text just by reading it. But how can machines do the same ? Well this is what we are going to discuss in this article.

There are numerous applications of the automatic text summarization including following:

1. Reading lengthy customer reviews and converting them into smaller and meaningful versions to be used to take necessary actions.
2. Converting news articles into a short summary. Mobile app **inshorts** is an example of this.
3. Creating concise summary reports from business meeting notes.

There are two types of text summarization methods, namely:

1. Extractive Text Summarization
2. Abstractive Text Summarization

The core idea behind this method is to find the similarities among all the sentences and return the sentences having maximum similarity scores. We use **Cosine Similarity** as the similarity matrix and **TextRank** algorithm to rank the sentences based on their importance.

Approach :

1. Import necessary packages like NLTK, networkx, numpy, regex.
2. Create a function which is used to read the text and convert it into sentences.
3. Basic text preprocessing to remove all the special characters.
4. Next we will create vectors from the sentences and calculate cosine similarity between these vectors.
5. Next we create a similarity matrix of dimension n x n to store the similarity values.
6. Finally the main function to call all the above functions in the pipeline.