**Text Summarization**

It is basically creating a summary of a long text given i.e extracting the core ideas of a document /essay/paragraph or other text forms. Here the summarization is done in an Extractive way i.e take up a document and extract important sentences from it as it is i.e without changing any words. It is taken as selecting some sentences on the basis of their importance in the document.

The algorithm that has been used here is **TextRank** (which finds its roots associated with Google’s PageRank used for ranking webpages.).

The PageRank of websites W, X, Y, and Z:

PageRank(W)=(1-d) + d \* (PageRank(X)/L(X)+ Pagerank(Y)/L(Y)+ PageRank(Z)/L(Z))

Where L(A)=no. of other pages that webpage A links to.

d= constant (prevent the PageRank to be zero mostly used value is 0.85).

Similarly, TextRank of sentences goes like this :

TextRank(A)=(1–d)+ d\*(TextRank(B)\*M[B,A]+TextRank(C)\*M[A,C])

Where M [B, A] = The spatial distance of the sentence A and B(similarity matrix)

The system accepts the text and converts it into tokens. Then it cleans the text and removes stopwords using nltk corpus. Then it converts the sentence into the vector using the Word2Vec algorithm. Later each sentence is padded to make sure of equal size. Then the Similarity matrix is filled by calculating the spatial distance between sentence sentences. Finally, the TextRank of every sentence is calculated Which is initialized as 1. At last, pick the sentence that has high priority according to calculated TextRanks.

**Libraries:** NumPy, ntlk , re, genism and scipy

-M.Mithin Kumar