ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

TEXT SUMMARIZATION USING EXTRACTIVE TECHNIQUES

Md. Ahmed[1], K. Naga Bhavani[2], K. Harsha Vamsi[3], S. N. Radha Sowmya[4], K. DivyaSri[5],

[1]Asst Professor, [2,3,4,5]Students

Dept of Computer Science And Engineering

Sri Vasavi Institute Of Engineering & Technology, Pedana, A.P, India

ABSTRACT:

Text Summarizer refers to the technique of shortening long pieces of text. The intention is to create a coherent and fluent summary having only the main points outlined in the document. The need for text summarization is "Today, our world is parachuted by the gathering and dissemination of huge amounts of data. With such a big amount of data circulating in the digital space, there is need to develop Natural Language Processing algorithms that can automatically shorten longer texts and deliver accurate summaries that can fluently pass the intended messages."

There are mainly two types of how to summarize text in Natural Language Processing they are Extraction-based summarization and Abstractive-based summarization. The extractive text summarization technique involves pulling key phrases from the source document and combining them to make a summary. The abstraction technique entails paraphrasing and shortening parts of the source document. We use Extraction-based summarization model. This model takes a input that encapsulates some paragraphs and returns a text summary that represents the key information or message in the input text. That text summary reduces reading time, accelerates the process of researching for information, and increases the amount of information that can fit in an area.

Keywords : Text Summarizer, Extraction, Abstraction

INTRODUCTION:

A summary is a text that is produced from one or more texts, that conveys important information in the original text, and it is of a shorter form. Text summarization refers to the technique of shortening long pieces of text. Text summarization finds the most informative sentences in a document. Automatic text summarization is a common problem in machine learning and natural language processing (NLP).

DOI : 10.36893.JK.2022.V12I01N01.809-813

ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

LITERATURE SURVEY:

Baxendale [1] has done his research at IBM on Extractive summarization. He extracted important sentence by using the position of text. The author has tested 200 paragraphs towards his goal to find that in 85% of the sentences which author has taken first topic which is main topic sentence and the last sentence came 7%. The most accurate sentence would be selected from these two sentences.

Edmundson [2] has done research on extracted summarization in this he extracted important sentence by using two features position and word frequency importance were taken from the previous works. The author has added two they are: presence of cue words, and the skeleton of the document.

Luhn [3] has done his research on the extractive summarization. In his research he extracted important sentence by calculating word frequency and phrase frequency that gives the useful measure of its significance.

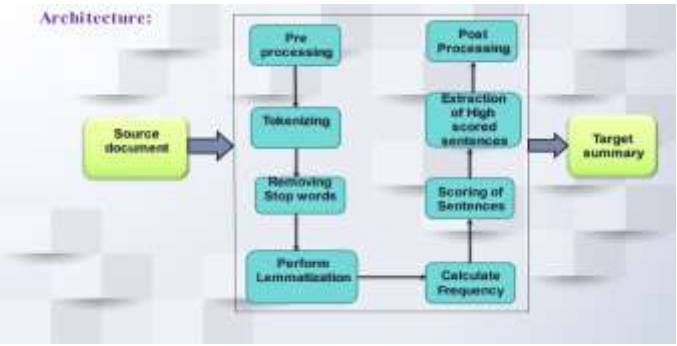
PROPOSED METHOD:

In present technology, there is huge quantity of information is generating on the internet day by day. It requires lot of time and effort for analyzing such amount of data. Previously human effort is required more for analyzing the data which takes lot of time. For this a better mechanism has to be provided for extracting useful information quickly and effectively.

Text summarization is one of the methods for identifying the important meaningful information in a document and compressing them into a shorter version preserving its overall meanings. It reduces the time required for reading whole document and also it reduces space problem that is needed for storing large amount of data.

This project describes a system for the summarization of the given text input. For this we are using Extractive text summarization, which means pulling key phrases from source document. Here we are using two extractive based summarization techniques.

SYSTEM ARCHITECTURE:



DOI : 10.36893.JK.2022.V12I01N01.809-813

ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

IMPLEMENTATION:

We are using Extractive text summarization which means pulling key phrases from source document. Here we are using two extractive based summarization techniques.

1) Sentence Ranking
2) Text Ranking

1) Sentence Ranking:

Sentences are ranked by assigning weights and they are ranked based on their weights.Highly ranked sentences are extracted from the input document so it produces high-quality summary of the input text.

Algorithm:

Input: A text format of the data is taken as input.

Output: An appropriate summarized output text is generated.

Reading the given text and given text is tokenized.

1. The stop-words are removed from the sentences.
2. Perform lemmatization for each token.
3. Frequency of individual token is calculated.
4. Weighted frequency of token is calculated by dividing frequency with maximum one.
5. Weights of each sentence is calculated by substituting weighted frequency of token in sentence.
6. Finally, summarizer will extract the weighted frequency sentences whose value is greater than or equal to average of sum of sentences in order to find summary of text.

2) Text Rank:

Text Rank is a general purpose, graph based ranking algorithm for NLP.

Algorithm:

Input: A text format of the data is taken as input.

Output: An appropriate summarized output text is generated which is shorter.

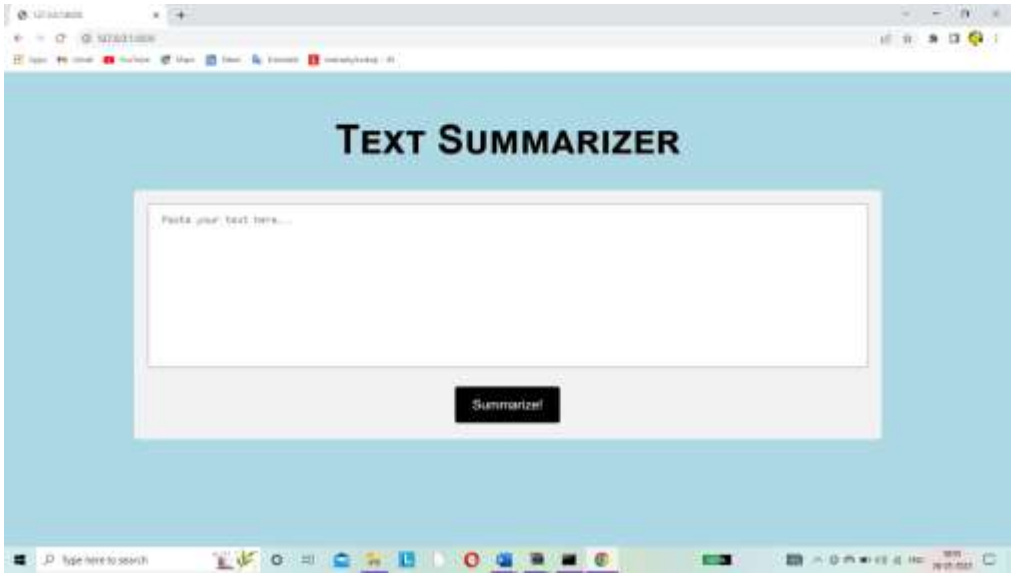
DOI : 10.36893.JK.2022.V12I01N01.809-813

ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022

1. Reading the given text and it is tokenized into sentences.
2. Now, We will find vector representation (TF-IDF) for each and every sentence.
3. The similarity between sentence vectors are calculated and stored in a matrix.
4. For sentence rank algorithm, Similarity matrix is then converted into a graph, with sentences as vertices and score
5. similarity as edges.
6. Finally, we got a certain number of ranked sentences whose rank is greater than or equal to the average of all the sentences from the final summary.

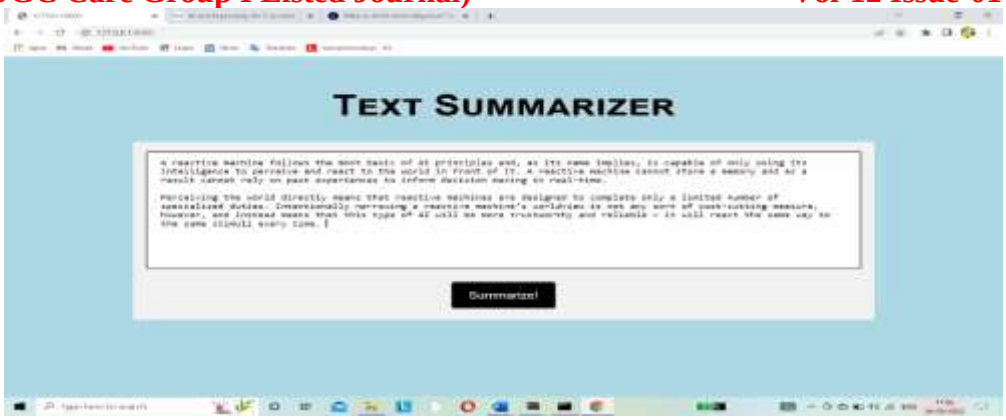
IMPLEMENTATION RESULTS:

The below figure displays the command prompt screen with some commands



DOI : 10.36893.JK.2022.V12I01N01.809-813

ISSN: 2278-4632
Vol-12 Issue-01 No.01: 2022



The below figure displays that new page which displays summary of the source text

CONCLUSION:

Text summarization is a complex task which contains many sub-tasks in it. Every sub task has an ability to get good quality summaries. The important part in extractive text summarization is identifying necessary sentences from the given sentences. In this project we proposed extractive based text summarization by using sentence ranking and textrank. The sentences which are extracted from input given by User Interface are produced as a summarized text and it is displayed in new page.

REFERENCES:

[1] Baxendale, P.(1958). "Machine-made index for technical literature" –an experiment. IBM Journal of Research developement354-361

[2] Edmundson, H. P. (1969). New methods in automatic extracting. Journal of the ACM, 16(2):264–285.

[3] Luhn, H (1958). "The automatic creation of literature abstracts". IBM Journal of Research Development, 2(2):159-165.

[4] Yulia Ledeneva, Alexander Gelbukh, and René Arnulfo García-Hernández, "Terms Derived from Frequent Sequences for Extractive Text Summarization", CICLing2008, LNCS 4919, pp. 593–604, (2008).

DOI : 10.36893.JK.2022.V12I01N01.809-813