IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Text Summarization Using Machine Learning

[1]Vaishnavi T. Jagade, [2]Saniya R. Jadhav, [3]Tulshi S. Dhavale, [4]Sneha .N Gogawale, [5] S.D. Sapate

[1]Student, [2]Student, [3]Student, [4]Student, [5]Lecturer

[1 2 3 4 5] Department of Computer Technology,

[1 2 3 4 5] Bharati Vidyapeeth's Jawaharlal Nehru Institute of Technology (Poly), Pune, Maharashtra, India

Abstract: Text summarization tools play a pivotal role in the era of information overload, aiding users in extracting the most pertinent information from voluminous texts. This abstract provides a succinct overview of text summarization tools, highlighting their significance, techniques employed, and potential applications .Text summarization tools employ various techniques, such as Text Preprocessing-(Tokenization, Lowercasing), Text Representation sequence to sequence models, Evaluation metrics, Hyper parameter Tuning ,extractive and abstractive summarization, to condense lengthy documents into shorter, coherent versions. Extractive summarization selects salient sentences or phrases from the original text, while abstractive summarization generates concise summaries by interpreting and rephrasing content .These tools rely on natural language processing (NLP), machine learning algorithms, and deep learning models to achieve these tasks. We also use tokenization ML algorithm to get quality of summary. It also have additional feature of uploading a pdf or file and getting its summary.

I. INTRODUCTION

Welcome to the fascinating world of text summarization! In this project, we will explore the art of distilling information from lengthy texts into concise summaries. Our journey will involve understanding various techniques, algorithms, and tools that play a crucial role in extracting the essence of written content. Get ready to dive into the realm of natural language processing and uncover the key elements that make text summarization a vital component in today's information rich landscape. This innovative approach not only saves time but also enhances information retention, catering to the fast-paced demands of our information-driven society .In today's digital age, the abundance of textual data poses a significant challenge for individuals and organizations seeking to extract meaningful insights efficiently. Amidst this sea of information, the ability to condense large volumes of text into concise and coherent summaries has become indispensable. Text summarization, a branch of natural language processing (NLP), offers a solution to this challenge by automating the process of distilling key information from lengthy documents.Traditional methods of text summarization often rely on predefined rules or heuristics, which may not adequately capture the nuances and complexities of language. However, with recent advancements in machine learning and deep learning techniques, there has been a paradigm shift towards data-driven approaches for text summarization. These methods leverage large corpora of text paired with neural network architectures to learn how to generate summaries that preserve the salient content of the original documents.

II. LITERATURE SURVEY

This literature review reveals the detailed work that has been carried out till date on the content of solar shadowing

.

1.Text Summarization Using Text Frequency Ranking Sentence Prediction-2020

In this paper, the authors explored a novel approach to text summarization. They proposed a method that utilizes text frequency to rank and predict sentences for inclusion in a summary. This approach is likely designed to streamline the summarization process by identifying and prioritizing the most relevant and informative sentences in a text document. Focuses on text summarization. Utilizes text frequency for sentence ranking.. Includes sentence prediction. Algorithm refinement for accuracy and efficiency.Application diversification (news, research, social media).Multilingual support. User customization options. Standardized evaluation metrics.

2. A Survey of Automatic Text Summarization: Progress, Process and Challenges-2021

This survey paper provides a comprehensive overview of the state of the art in automatic text summarization. It discusses the progress made in this field, outlines the underlying processes and methodologies, and highlights the key challenges that

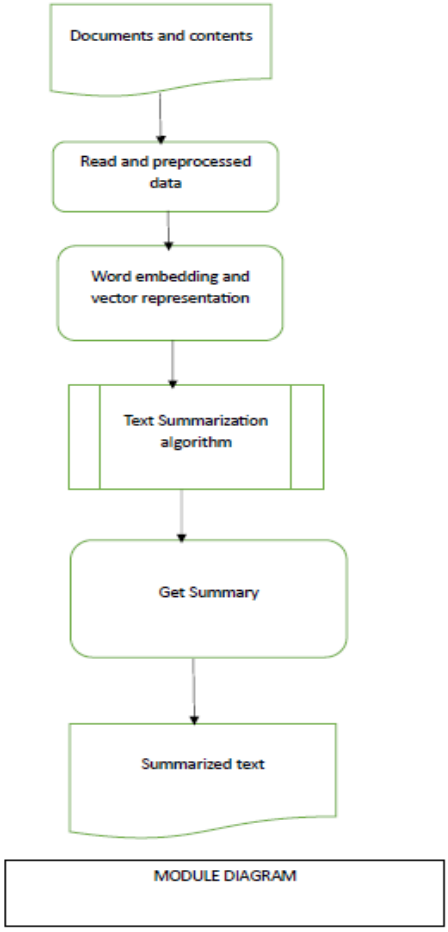
researchers face in developing effective text summarization systems .Comprehensive survey of automatic text summarization. Discusses progress, processes, and challenges .Incorporates the latest research up to November 2021. Periodic updates to the survey. Exploration of deep learning in summarization. Expansion into multimodal summarization .Development of standardized evaluation metrics.

III. METHODOLOGY

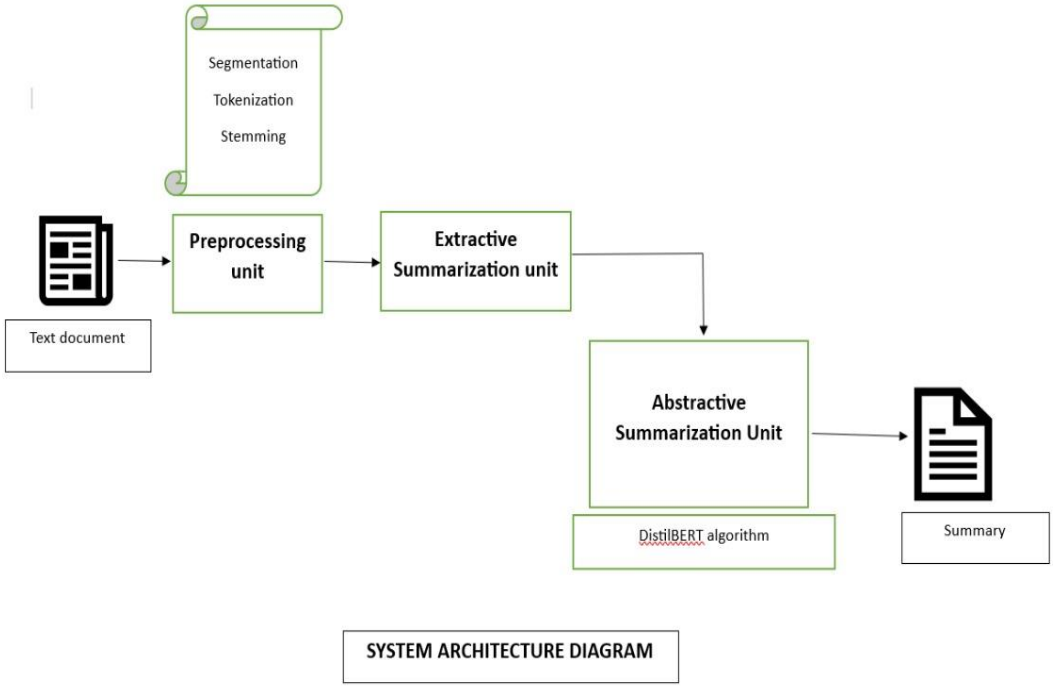
3.1 Component used:

- Processor: Intel(R) Core(TM) i3-6006U CPU @ 2.00GHz
- Installed Memory(RAM): 4.00 GB
- System Type: 64-Bit Operating System, x64 Based Processor
- Visual studio code
- PYTHON
- JavaScript
- CSS

3.2 Module Diagram:



3.2 Architecture Diagram:

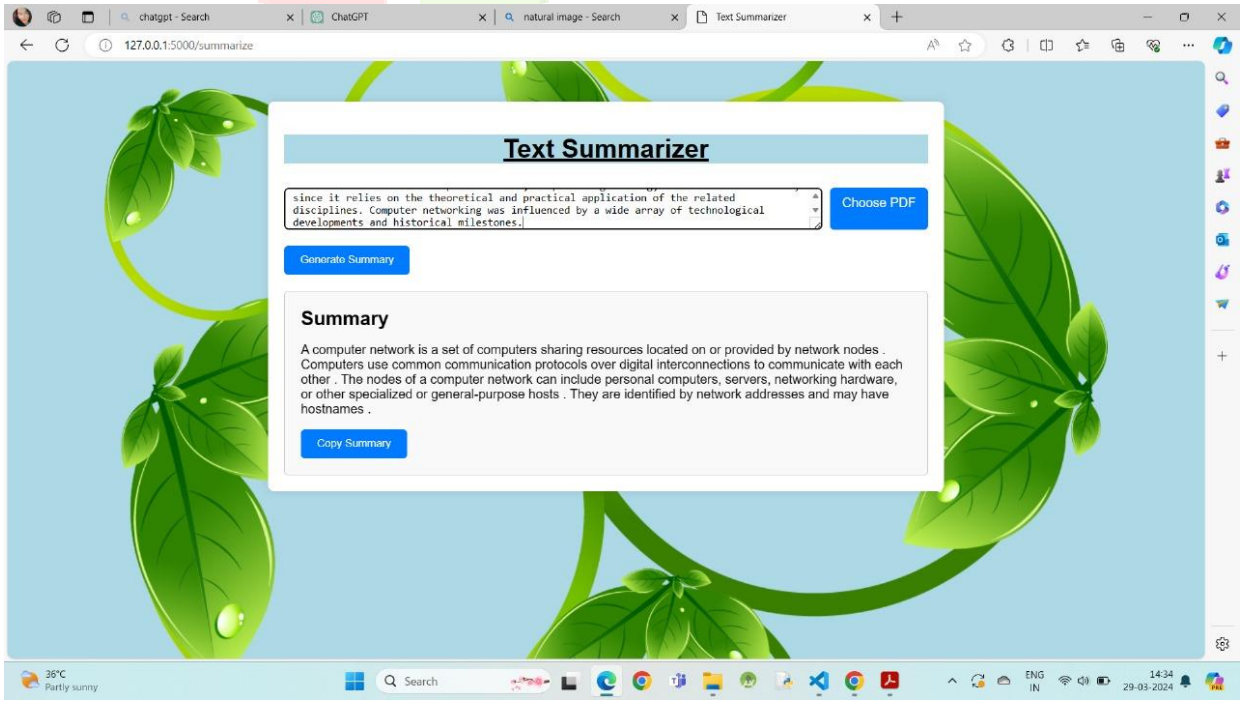


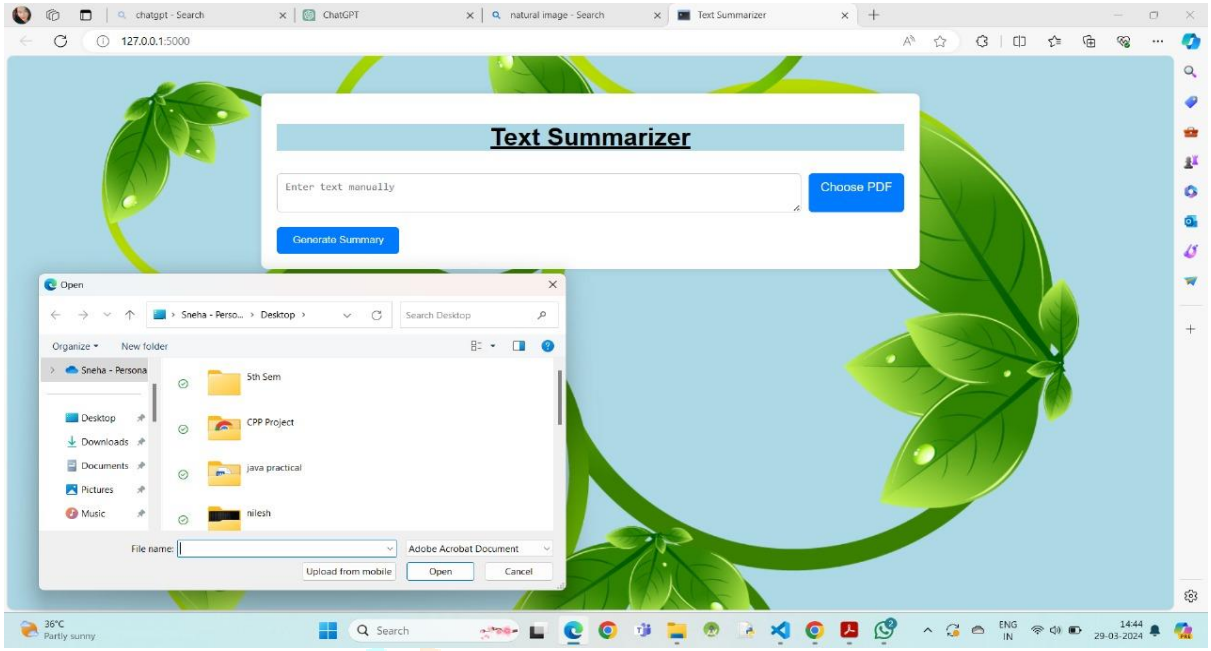
SYSTEM ARCHITECTURE DIAGRAM

3.3 Working:

A text summarization project starts with collecting textual data. This data undergoes preprocessing to clean it up and remove unnecessary elements like formatting and stop words. Relevant features are then extracted from the text, such as word frequencies or key phrases. The summarization algorithm is applied, which can either extract important sentences directly from the text (extractive summarization) or generate new sentences that capture the essence of the original text (abstractive summarization). The quality of the summaries is evaluated on evaluation results and feedback, the summarization algorithm may undergo refinement. Once optimized, the summarization model is deployed for practical use, considering factors like scalability, efficiency, and ethical implications throughout the process.

IV.IMPLEMENTATION





V. Future Work

As summarization technologies become more advanced, it will be crucial to address ethical concerns related to privacy, bias, and the potential misuse of summarized content. Future research should focus on developing safeguards and guidelines to ensure responsible use of summarization technologies.

VI. Conclusion

In conclusion, machine learning-based text summarization offers a promising approach to condense large volumes of text into concise summaries. By leveraging algorithms such as neural networks and natural language processing techniques, these systems can effectively extract key information and generate summaries that capture the essence of the original text. While challenges like maintaining coherence and handling diverse text types remain, ongoing research and advancements continue to enhance the capabilities and accuracy of text summarization models, making them valuable tools for various applications in information retrieval, content analysis, and beyond.

VII. Acknowledgment

We express our profound gratitude to all those who have contributed to the successful completion of our sun tracking solar panel project.
We would like to extend our sincere thanks to our Professor S.D Sapate whose expert guidance & insightful feedback helps us to complete this research.

References

1. Text Summarization Using Text Frequency Ranking Sentence Prediction-2020
2. A Survey of Automatic Text Summarization: Progress, Process and Challenges-2021
3. Comparison of Sentiment Analysis on AutoSummarized Text & Original Text using various Summarization Techniques-2021
4. Abstractive vs Extractive Text Summarization (Output based approach) - A Comparative Study-2020
5. Extracted Summary Based Recommendation System for Indian Legal Documents-2020
6. News Text Summarization Based on Multi-Feature and Fuzzy Logic-2020
7. Tool to Extract and Summarize Methodologies of Research Articles for Visually Impaired Researchers2019
8. Toward Summary Extraction Method for Functional Topic-2017
9. Text Summarization and Dimensionality Reduction using Learning Approach-2020
10. Extractive Automatic Text Summarization using SpaCy in Python and NLP-2021