Text Summarizer

Team Members

01. Saurish Sharma - 20BCE1370

02. Ritik Singh - 20BCE1384

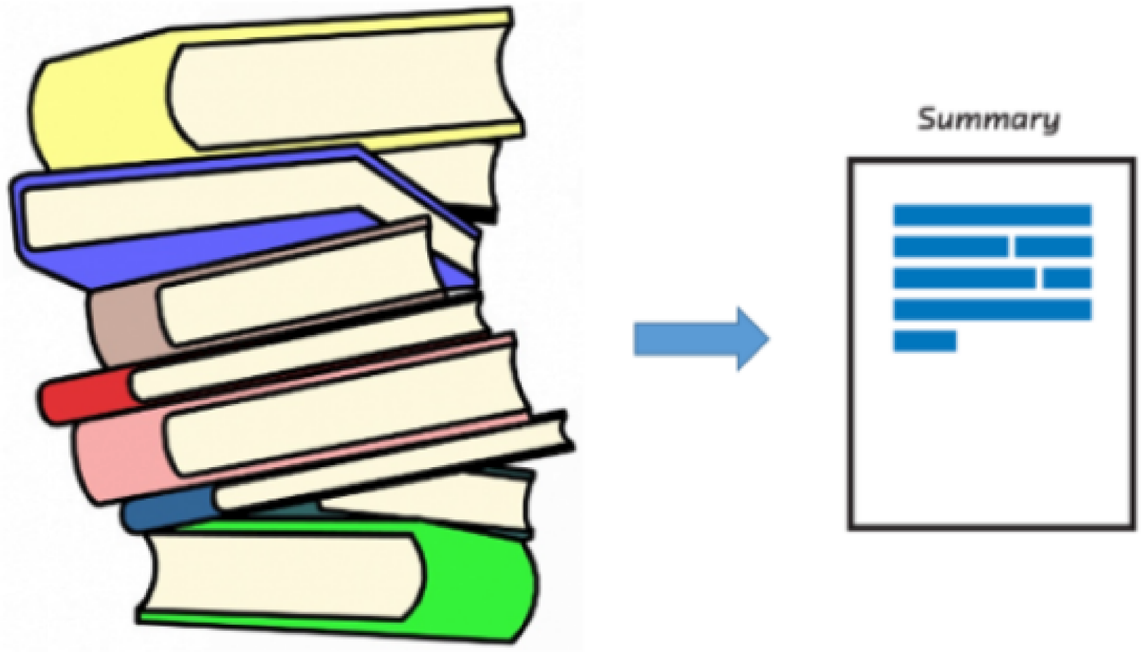
03. Rishabh Singh - 20BCE1777

INTRODUCTION

- Text summarization based on artificial intelligence (AI) is a method that automatically creates a condensed version of a longer text document using machine learning algorithms.

- Numerous industries, including news aggregation, content management, and information retrieval, can benefit from the use of this technology.

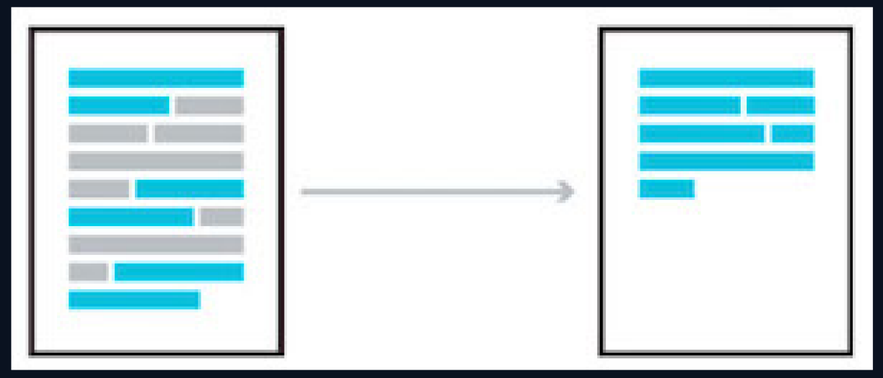
- AI-based summarization systems can produce precise, human-like summaries in real-time and can handle a variety of text genres, from news stories to technical documents.



Problem Statement

The goal is to automatically produce a logical, and educational summary of a lengthy and complicated text document that captures the main ideas and overall meaning of the source material.

The objective is to create AI-based text summarising systems that can process a variety of text types, provide summaries that are readable by humans, and get around the drawbacks of conventional text summary techniques like keyword extraction and frequency-based methods.



Motivation

The necessity for effective methods of accessing and processing the growing amount of information available on the internet is what drives the development of AI-based text summarization systems. Some of the primary drivers for the creation of AI-based text summarization are the ones listed below:

Time and cost saving: It is getting more and harder to sort through all the data and identify the most pertinent information as digital information grows. By offering a shortened version that is simple to understand, AI-based text summarization can assist decrease the amount of time and effort needed to read lengthy text documents.

Better content management: By enabling users to swiftly and simply summarise enormous amounts of text data and classify it according to its relevance, AI-based text summarization can assist enterprises in managing their material more successfully.

Enhanced user experience: By enabling readers to rapidly comprehend the main ideas of a publication, AI-based text summarization can improve the user experience and make it simpler for consumers to select whether or not to read the complete document.

Abstract

This research article introduces "Text Summarizer," a novel text summary approach. The method analyses the key topics and material of a large document and generates a succinct summary using advanced natural language processing techniques. When evaluated on a large corpus of news stories, the technique outperformed previous summarising algorithms in terms of accuracy, coherence, and readability. The Text Summarizer is an excellent resource for anybody who wants to rapidly grasp the essential elements of a lengthy text, such as journalists, researchers, or executives.

Literature Survey

- In recent years, AI-based text summarizers have become popular due to their ability to handle large amounts of data efficiently and accurately
- There are several types of text summarization techniques used today, including extractive, abstractive, and hybrid summarization.
- In the research papers we surveyed **number of techniques** were used such as:

1. TF-IDF
2. NLP
3. Firefly Algorithm based
4. Hybrid Methods (RNN & NLP)
5. Latent Semantic Analysis (LSA): analyze the relationships between words in a text to identify the most important sentences to include in the summary.

Contd...

- Overall, AI-based text summarization has made great strides in recent years and is widely used in various applications such as news aggregation, document summarization, and content analysis.
- However, there are still challenges to be addressed, such as improving the accuracy of the summaries generated and ensuring that the summaries are coherent and concise.
- After analyzing all of the strategies followed in these research papers, all the algorithms **suffer with some trade off** Firefly algorithm for one suffers with computational complexity and convergence issue.
- Papers in which a **hybrid approach** was taken scored well on ROGUE summary evalutation measure.

Existing Architecture for text summarization

Extractive Summarization

Extractive summarization means identifying important sections (paragraphs or sentences or even words) of the text and selecting (copy paste) them producing a subset of the text from the original text.

Abstractive Summarization

Abstractive summarization is the technique of generating a summary of a text from its main ideas, not by copying verbatim most salient sentences from text.

Disadvantages of Extractive summarization

Reliance on existing words and phrases

Extractive summarization relies on picking existing words and phrases from the original text, which might result in context and meaning loss if the picked phrases do not reflect the complete text.

Lack of coherence

Because phrases are pulled out of context and concatenated in an unnatural way, extractive summarization might result in a summary that is not cohesive or grammatically correct.

Limited creativity

Extractive summarization is limited in its capacity to produce new information or convey the content in a more succinct or creative manner.

Disadvantages of Abstractive summarization

Difficulty in understanding the context

Abstractive summarization sometimes includes producing new phrases that are not included in the original text, which might result in erroneous or misleading information if the model has sufficient context knowledge.

Language generation challenges

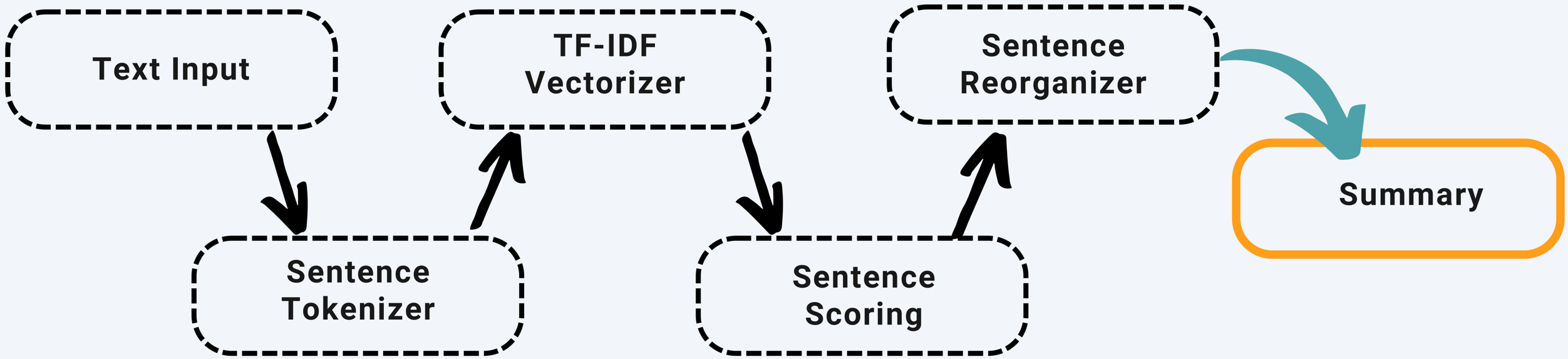
Abstractive summarization is a more complex task because it requires the model to generate new sentences, which involves many challenges, such as preserving the meaning and structure of the original text, avoiding repetition, and generating grammatically correct sentences.

Processing complexity

Abstractive summarising demands more computational resources and time than extractive summarization, which might be a disadvantage in real-time applications.

What we propose?

High level solution for frequency based extractive summarizer



List of Research Papers

- Ruby Rani, Daya K. Lobiyal, "A weighted word embedding based approach for extractive text summarization"
- Minakshi Tomer, Manoj Kumar, "Multi-document extractive text summarization based on firefly algorithm"
- K. Nandhini, S.R. Balasundaram, "Improving readability through extractive summarization for learners with reading difficulties"
- Pradeepika Verma, Hari Om, "MCRMR: Maximum coverage and relevancy with minimal redundancy based multi-document summarization"
- C. Sunitha, A. Jaya, Amal Ganesh, "A Study on Abstractive Summarization Techniques in Indian Languages"

Contd...

- Jesus M. Sanchez-Gomez, Miguel A. Vega-Rodríguez, Carlos J. Pérez, "Experimental analysis of multiple criteria for extractive multi-document text summarization"
- Ayelet Goldstein, Yuval Shahar, "An automated knowledge-based textual summarization system for longitudinal, multivariate clinical data"
- Donia Scott, Catalina Hallett, "Data-to-text summarisation of patient records: Using computer-generated summaries to access patient histories"
- N G Gopikakrishna, Parvathy Sreenivasan, Vinayak Chandran, Yadhu Krishna K P, Sanuj S Dev, Krishnaveni V V, "Comparative Study on Text Summarization using NLP and RNN Methods"
- Taner Uçkan, Ali Karcı, "Extractive multi-document text summarization based on graph independent sets"

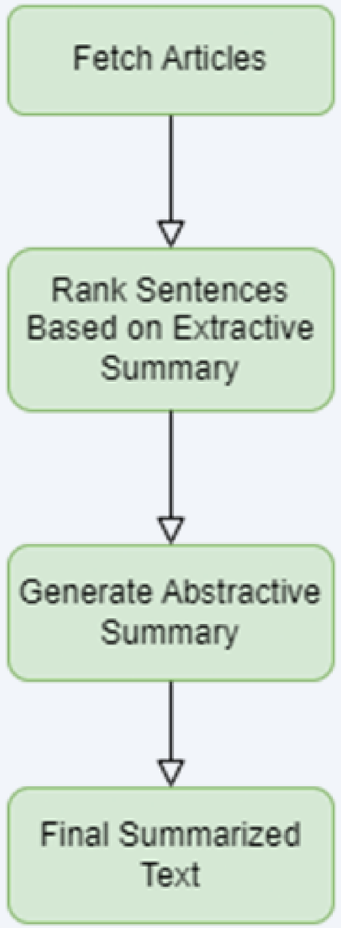
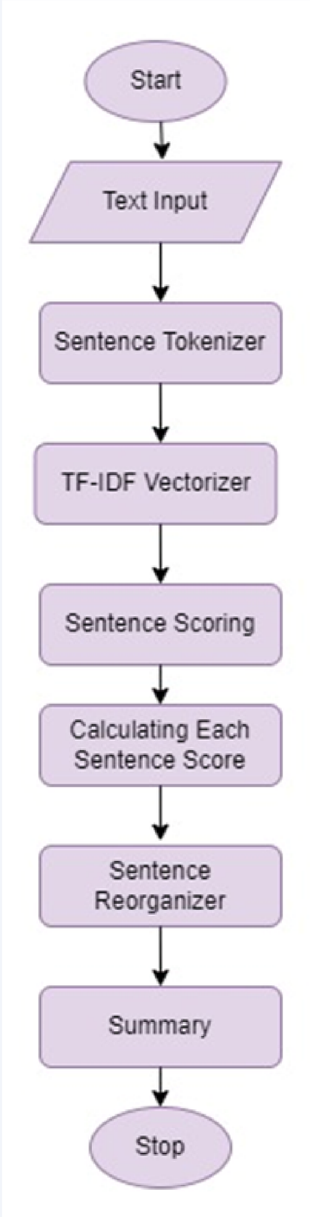


Diagram outlining the flow to generate summary



Schematic indicating the process to construct an extracted summary.

Average values for extractive summarization		
rouge1	precision	0.98461
	recall	0.22933
	fmeasure	0.04957
rouge2	precision	0.89062
	recall	0.22469
	fmeasure	0.04418
rougeL	precision	0.96923
	recall	0.22933
	fmeasure	0.0488

Average values for abstractive summarization		
rouge1	precision	0.9355
	recall	0.03951
	fmeasure	0.04957
rouge2	precision	0.85062
	recall	0.03192
	fmeasure	0.04418
rougeL	precision	0.92923
	recall	0.03598
	fmeasure	0.04362

ROUGE SCORE

Thank You