

# NLP Techniques for Text Summarization

Harshith A, Anjaneya Vinay D, Gurudev O, Mohit V

## Introduction

In Natural Language Processing (NLP), text summarizing is a method used to consolidate long text documents by extracting the most important information and presenting it in a brief summary. This is achieved through advanced algorithms and models that can understand, interpret, and generate human-like text. It represents a significant stride in the field of artificial intelligence, offering a way to efficiently process and condense large volumes of text data, making information more accessible and digestible for users.

**Challenges:** Summarization algorithms must overcome various challenges like capturing context, dealing with ambiguities, and maintaining the coherence and fluency of the text.

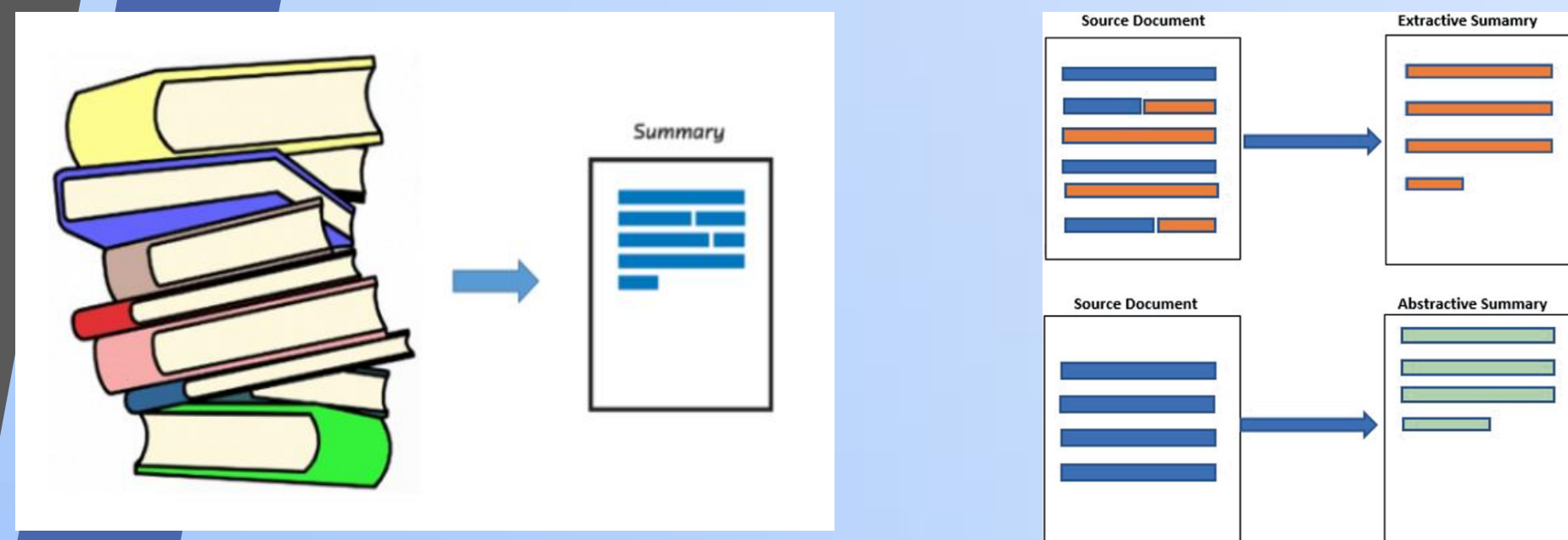


Figure 1. Text summarization

## Applications

Text summarization has a wide range of applications across various fields. Some of the key areas where text summarization is particularly useful:

1. News Aggregation and Media
2. Academic Research
3. Business Intelligence
4. Legal Document Analysis
5. Customer Feedback Analysis
6. Healthcare Documentation
7. Search Engines
8. Email Management
9. Social Media Monitoring

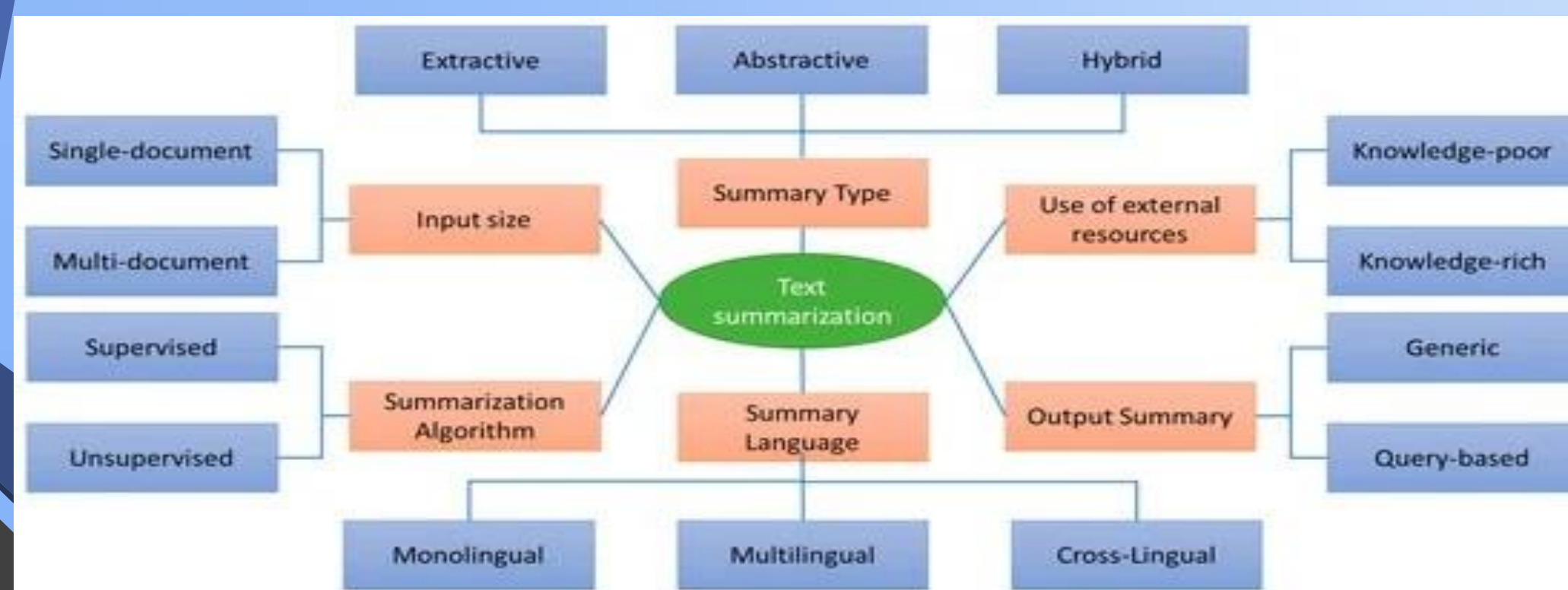


Figure 2. Types of Text summarization.

## Extractive Summarization

- Extractive summarization is a method in Natural Language Processing (NLP) where the goal is to condense a text document by extracting key sentences or phrases directly from the original content and combining them to form a coherent summary.
- This method is often favored for its straightforwardness and accuracy in preserving the exact wording and factual details from the source material.

**Process:** It typically involves ranking segments of the source text (such as sentences) based on their relevance and importance, often using statistical methods, machine learning algorithms, or a combination of both.

**Challenges:** One of the limitations is that extractive summaries might lack fluency and coherence since the extracted sentences may not always connect well.

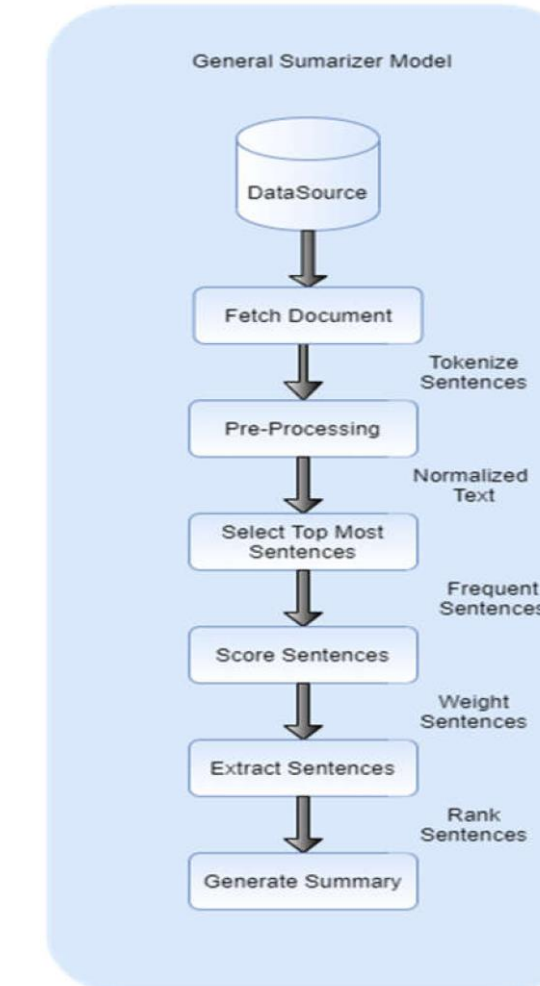


Figure 3. Extractive Summarization workflow.

## Methods of Extractive Summarization

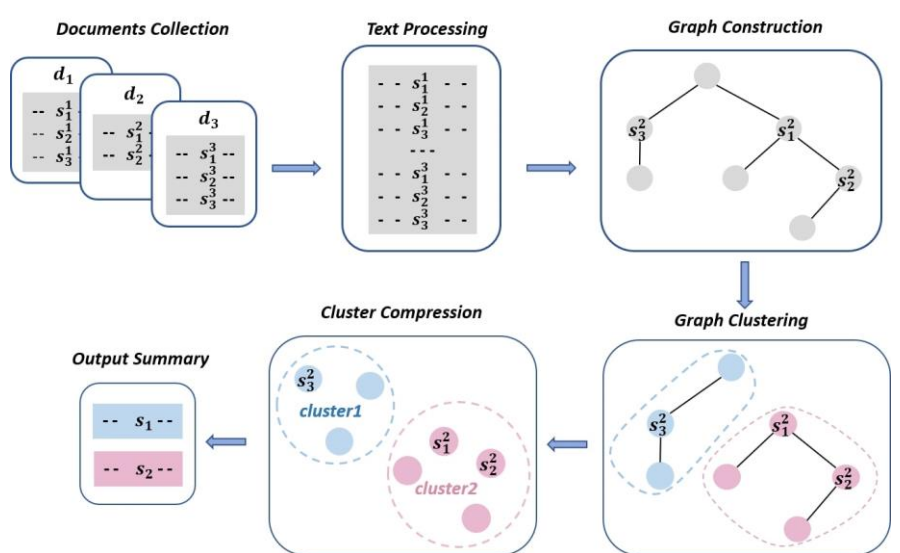
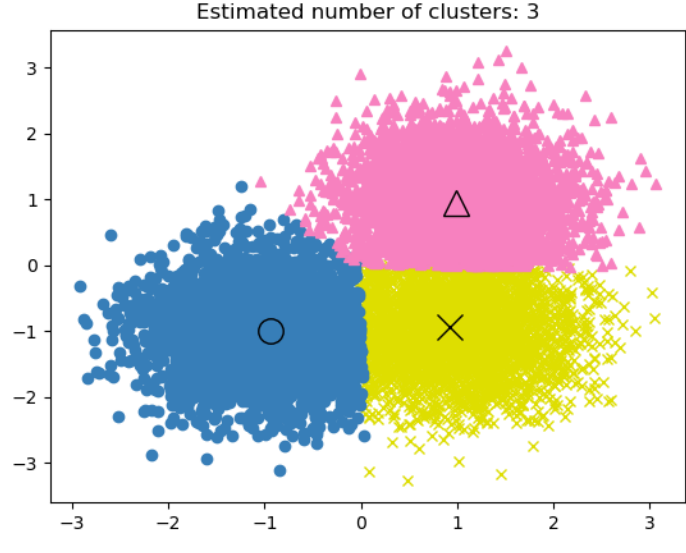
1. Heuristic-Based Methods
  - Frequency-Driven Approaches (Term Frequency-Inverse Document Frequency)
  - Position-Based Techniques
2. Graph-Based Methods
3. Machine Learning Approaches
4. Topic-Based Methods (Clustering Techniques)

Figure 5. Clustering method.

## Abstractive Summarization

Abstractive summarization aims to understand the main ideas of the source text and generate a new, shorter text that conveys the same information, often rephrasing or paraphrasing the original content.

**Process:** This type of summarization typically involves advanced NLP techniques, including semantic understanding and natural language generation, often utilizing deep neural networks.



Figure 6. Abstractive Summarization.

## Methods of Abstractive Summarization

1. Neural Network-Based Approaches ( RNN )
2. Transformer-Based Models ( BERT, GPT, T5 )

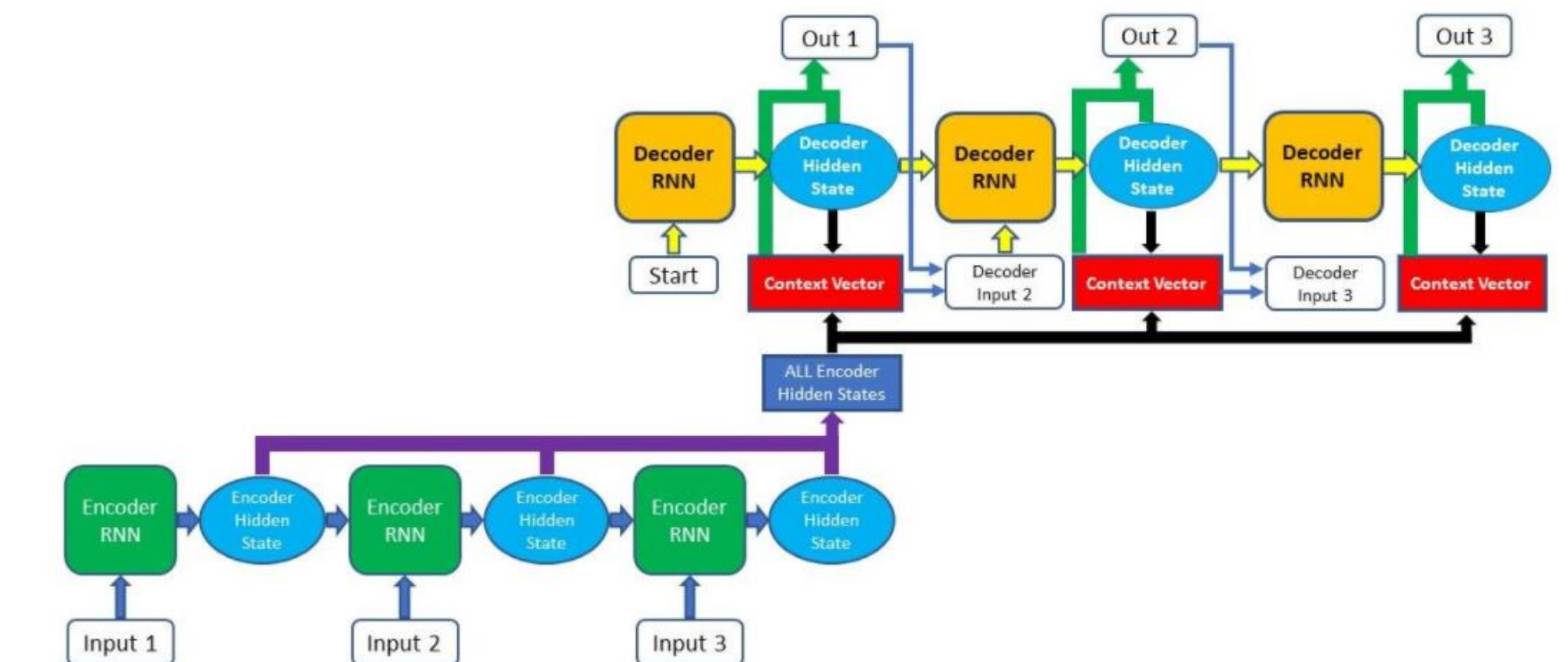


Figure 7. Transformer Based Model.

3. Attention Mechanisms

## Implementation and Results

### Extractive summarization:

- We performed an extractive summarization of a story called "Gift of Magi". Here, the scoring mechanism weighs words based on their frequency in the main text relative to their frequency in both text and title, while also considering the presence of title words in each sentence.
- We got the ROUGE & BLEU scores of approximately 40% and 29% respectively. The performance of this system can be improved by utilizing word embeddings and leveraging pre-trained language models like BERT.

### Abstractive Summarization:

- We have also implemented abstractive summarization using encoder-decoder architecture including self-attention mechanisms on inshorts news dataset containing news headlines and articles.
- We got the ROUGE scores of approximately 25%. Results can be improved by changing the hyperparameters (increasing epochs and layers) i.e. by performing more training which would require more computing power.

## Conclusion

- In conclusion, text summarization in Natural Language Processing (NLP) stands as a crucial field of artificial intelligence, offering a way to efficiently process and condense large volumes of text data. The field has seen remarkable advancements, especially with the advent of deep learning and transformer-based models that have significantly improved abstractive summarization techniques. These developments promise to enhance the efficiency and applicability of NLP summarization tools across various domains.
- However, challenges such as context preservation, idiomatic expression handling, and the balance between brevity and comprehensiveness continue to be areas for future research. As these technologies evolve, they are expected to become more integrated into everyday applications, revolutionizing the way we access and comprehend large volumes of text.

## References

1. Surveying the Landscape of Text Summarization with Deep Learning: A Comprehensive Review - Guanghua Wang, Weili Wu
2. Text Summarization: A Technical Overview and Research Perspectives Korrapati Sindhu, Karthick Seshadri
3. Review of automatic text summarization techniques & methods Adhika Pramita Widyassari a b, Supriadi Rustad a, Guruh Fajar Shidik a, Edi Noersasongko a, Abdul Syukur a, Affandy Affandy a, De Rosal Ignatius Moses Setiadi
4. An abstractive text summarization technique using transformer model with self-attention mechanism Sandeep Kumar & Arun Solanki
5. Extractive Text Summarization Using Sentence Ranking By J.N. Madhuri, R. Ganesh Kumar