Text Summarization

Ritesh Jha, Sanjay Sharma, Ashish Kumar

Under Supervison of
Dr. Tanu Gupta

School of Engineering and Technology

K.R. MANGALAM UNIVERSITY
THE COMPLETE WORLD OF EDUCATION

Introduction

Text summarization is a crucial technique in Natural Language Processing (NLP) that automatically condenses large volumes of text into shorter, more digestible summaries. There are two main types: extractive summarization, which selects key sentences directly from the original text, and abstractive summarization, which generates new sentences to capture the main ideas. Our project focuses on developing a model using [Fine-tuning] Fine-tuning a model typically refers to taking a pre-trained machine learning model and help to create accurate and informative summaries. The goal is to improve information retrieval and make content more accessible across various industries.

Objective

1. **Improve Summary Accuracy**: Ensure the summaries accurately capture the main ideas of the original text.

2. **Handle Different Texts**: Make sure the model can work with various types of text, like articles or reports.

3. **Make Summaries Easy to Understand**: Focus on creating summaries that are clear and easy for users to read.

4. **User-friendly Interface**:Design a simple and easy-to-use interface for users to input text and get summaries quickly.

Problem Statement

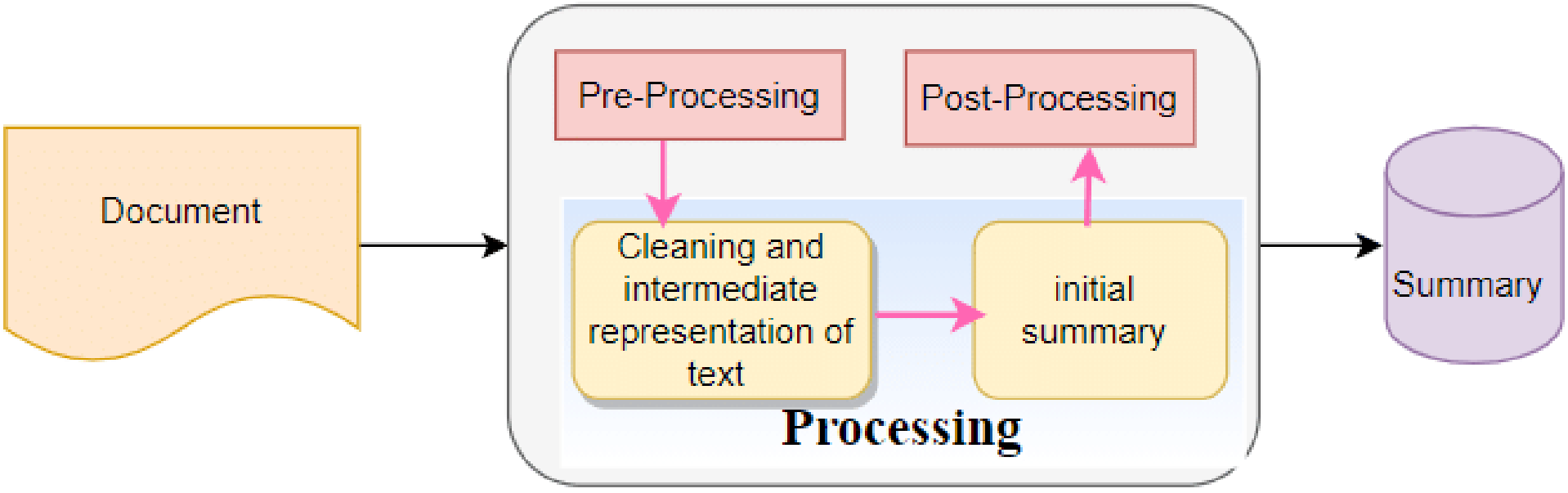
The system should preserve the original meaning and include the most important points.

Developing such a system is challenging because it needs to understand natural language and choose or rewrite content intelligently.

Creating summaries manually is time-consuming, especially when dealing with large numbers of documents.

Users often want a quick summary to save time and focus only on the important po

Methology


Document
Pre-Processing
Post-Processing
Cleaning and intermediate representation of text
initial summary
Processing
Summary

Expected Outcomes

•**Concise Summaries**:
•The system should be able to reduce long documents (e.g., articles, research papers, news reports) into a shorter, more digestible form without losing important information.

Improved Readability:
•The output summary should be easy to read and should flow naturally. It should not sound like a collection of random sentences but rather a coherent summary.

Time Efficiency:
•Users can save time by reading summaries instead of lengthy documents, helping them process information more quickly.

Accuracy and Relevance:
•The summary should be highly relevant, avoiding irrelevant or redundant information. It should focus only on the most important content.

Applications

News & Media:

•**Automatic News Summaries**: With a large volume of news articles published daily, automated summarization can provide readers with concise summaries of breaking news stories or articles, helping them stay informed quickly.

Healthcare:

•**Medical Research Papers**: Summarization helps researchers and healthcare professionals stay up-to-date with the latest research findings by condensing long academic papers into key takeaways.

Education:

•**Study Materials**: Students can use text summarization tools to create condensed versions of textbooks, research papers, or lecture notes, helping them review essential information.

Challenges and Limitations

Loss of Important Information:
•**Challenge**: Especially in extractive summarization, important but subtle information can be missed if the algorithm doesn't rank sentences effectively or if it doesn't understand nuanced meaning.

Difficulty with Long and Complex Documents:
•**Challenge**: Long and highly detailed documents, such as scientific papers or legal texts, present a challenge because the system must identify and extract only the most relevant content without overwhelming the reader.

Evaluating Quality of Summaries:
•**Challenge**: While there are automated metrics (like ROUGE) to evaluate summaries, they don't always correlate well with human judgment. Evaluating the quality of summaries in terms of coherence, readability, and informativeness remains a challenge.

Conclusion

Text summarization plays a crucial role in today's information-heavy world, where vast amounts of text are generated daily across industries. The ability to quickly extract key points and distill long documents into concise summaries has wide-reaching benefits, including time-saving, improved decision-making, and enhanced content accessibility. As technologies like machine learning and deep learning evolve, both **extractive** and **abstractive** summarization methods are becoming more sophisticated, providing systems that can generate accurate, relevant, and coherent summaries.

THANK YOU