

Multi-Source Newspaper Summarization and Comparison System

V. Chandra Sekhar Reddy, Gayathri Singireddy, Rahul Yalagandula, Sruthi Uthipalli

Associate Professor, Undergraduate Student, Undergraduate Student, Undergraduate student

Department of Computer Science and Engineering

ACE Engineering College, Ghatkesar, Hyderabad, Telangana, India- 501301

Abstract— In today's digital era, news is reported simultaneously by multiple media outlets, often presenting the same event with varying levels of detail, emphasis, and perspective. For readers, this abundance of information can be overwhelming and time-consuming to analyze. This paper presents a Multi-Source Newspaper Summarization and Comparison System designed to automatically collect news articles from different sources, generate concise summaries, and compare reporting patterns across newspapers. The proposed system applies natural language processing techniques to preprocess text, extract important content, and measure similarity between articles. By presenting a unified and comparative view of news coverage, the system helps readers quickly understand key facts while identifying overlaps and differences in reporting. The design emphasizes simplicity, interpretability, and practical usability, making it suitable for academic and real-world applications.

Keywords— *Multi-source summarization, News comparison, Natural language processing, Text summarization, Similarity analysis*

I. INTRODUCTION

The rapid expansion of online journalism and digital news platforms has significantly reshaped the way people access and consume information. Today, news reaches readers almost instantly, and a single event is often reported by multiple newspapers within a very short period of time. Although these reports are based on the same event, each newspaper presents the story in its own manner, influenced by editorial priorities, writing style, and narrative focus. While such diversity in reporting can offer broader perspectives, it also places an increased burden on readers who want to understand the complete picture.

In many cases, readers are required to go through several lengthy articles to identify the key facts of an event. Manually comparing news reports is not only time-consuming but also mentally demanding. Readers may find it difficult to distinguish between core information

and additional commentary, identify overlapping facts, or recognize subtle differences in emphasis and framing. As a result, important details may be overlooked, and the overall understanding of the event may remain fragmented. This challenge becomes more prominent in fast-paced environments where readers seek quick and reliable information.

Modern progress in natural language processing and learning-based techniques has enabled efficient examination and understanding of vast amounts of textual information. Text summarization techniques help reduce lengthy articles into concise versions while retaining essential information, enabling readers to grasp the main ideas quickly. Similarly, similarity analysis techniques allow systems to identify relationships between different texts, highlighting common content as well as variations in reporting. When used together, these techniques provide a powerful foundation for automated news analysis.

This work proposes a multi-source newspaper summarization and comparison system aimed at improving how readers interact with news content. The system automatically processes articles from multiple newspapers, generates clear and human-readable summaries, and compares coverage across sources. By highlighting both similarities and differences in reporting, the system helps readers understand news events more efficiently, reduces information overload, and supports informed interpretation without requiring extensive manual effort.

II. LITERATURE SURVEY

The growth of digital news platforms has increased interest in automatic text summarization and news analysis. Early research mainly focused on single-document summarization, where the goal was to shorten individual articles while preserving key information. These methods mainly relied on extractive summarization, where sentences were selected based on simple indicators like how frequently words appeared and where the sentences were positioned in the article. Although efficient, they often lacked coherence and

contextual clarity.

As multiple newspapers began reporting the same events online, multi- document summarization gained attention. This approach aims to combine information from different sources into a single concise summary while reducing redundancy. While these methods improved information coverage, challenges related to redundancy handling and sentence organization remained.

Overall, prior research demonstrates the effectiveness of summarization techniques but reveals a gap in systems that support structured comparison across multiple news sources. Most existing approaches focus on generating summaries without offering clear comparative insights. This limitation motivates the proposed work, which aims to integrate multi-source summarization with comparison to enable efficient and meaningful newspaper analysis.

III. PROBLEM STATEMENT

With the continuous increase in the number of online news sources, readers are exposed to repeated reporting of the same events in different formats and lengths. Although the core information may be similar, variations in presentation, emphasis, and wording can make it difficult to quickly identify the main facts and perspectives.

Existing news aggregation platforms primarily focus on collecting and displaying articles rather than analyzing their content. As a result, users are required to manually read and compare multiple reports to form a complete understanding of an event. This process is inefficient and may lead to missed information or biased interpretation. Existing news aggregation platforms primarily focus on collecting and displaying articles rather than analyzing their content. As a result, users are required to manually read and compare multiple reports to form a complete understanding of an event. This process is inefficient and may lead to missed information or biased interpretation.

Therefore, there is a need for an automated system that can summarize news articles from multiple sources and compare them in a structured manner. Such a system should reduce reading effort, provide concise summaries, and offer clear comparative insights that help users understand how different newspapers report the same event.

IV. SYSTEM DESIGN

The proposed system is designed as a modular pipeline that processes news articles collected from multiple newspaper sources. Each module performs a specific function, ensuring clarity, scalability, and ease of maintenance. Initially, news articles are provided to the system either through user input or predefined sources. The collected articles are passed to a preprocessing module that cleans and standardizes the text by removing unnecessary symbols, normalizing case, and segmenting sentences. This step ensures consistency across different sources.

The processed text is then forwarded to the summarization module, which identifies key sentences and generates concise summaries for each article. These summaries aim to retain essential information while significantly reducing length.

Next, the comparison module analyzes the generated summaries using similarity measures to identify commonalities and differences across newspapers. This analysis helps highlight how various sources report the same event.

Finally, the output module presents the summaries and comparison results to the user in a clear and readable format, enabling quick understanding and informed interpretation.

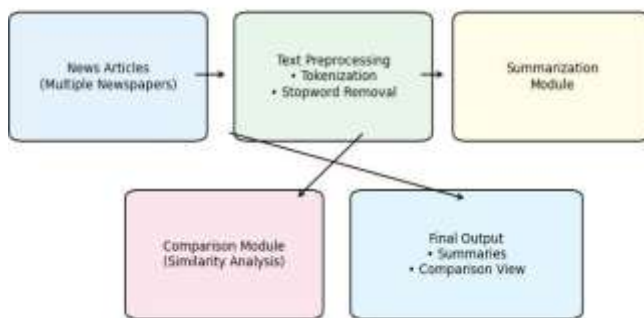
The proposed Newspaper Article Summarization and Comparison System provides an efficient solution for analyzing and understanding news content from multiple sources. By integrating text preprocessing, automatic summarization, and similarity-based comparison, the system enables users to quickly obtain concise information and identify differences in reporting across newspapers. The modular design ensures flexibility, scalability, and practical usability for real-world applications. Overall, the system reduces information overload, minimizes manual effort, and supports effective interpretation of multi-source news data, making it a valuable tool for readers and researchers alike.

SYSTEM FLOW OVERVIEW:

The system begins by collecting news articles from multiple newspapers as input. These articles are passed to the text preprocessing stage, where operations such as tokenization and stop word removal are performed to clean and

standardize the textual data. The processed text is then forwarded to the summarization module, which extracts important information and generates concise summaries for each article. At the same time, the preprocessed data is sent to the comparison module, where similarity analysis is carried out to evaluate relationships between articles from different sources. Finally, the system produces an integrated output consisting of summarized content and a comparison view, enabling users to efficiently understand key facts and variations in reporting.

System Design of Newspaper Article Summarization and Comparison System



V. METHODOLOGY

A. Data Collection

News articles related to the same event are collected from multiple online newspapers and digital platforms. These articles serve as the primary input to the system and are organized in a structured format to enable further processing and analysis.

B. Text Preprocessing

After data collection, the news articles undergo a text processing stage to make the content suitable for analysis. Since online news often contains noise such as irrelevant characters, inconsistent formatting, and redundant words, several preprocessing operations are applied. These include tokenization, elimination of common filler words, and normalization of text to ensure uniform representation. This stage improves data quality and converts the raw content into a consistent computational form, allowing accurate summarization and comparison in later stages.

C. Summarization Process

The cleaned and preprocessed text is then provided to the summarization component, which focuses on identifying the most informative sentences within each news article. Using natural language processing methods, this module condenses lengthy articles into brief summaries that preserve essential ideas and factual details. The aim of this stage is to shorten the overall length of the content while preserving its meaning and importance, enabling readers to understand the main idea of each article without going through the entire text.

D. Similarity Analysis and Comparison

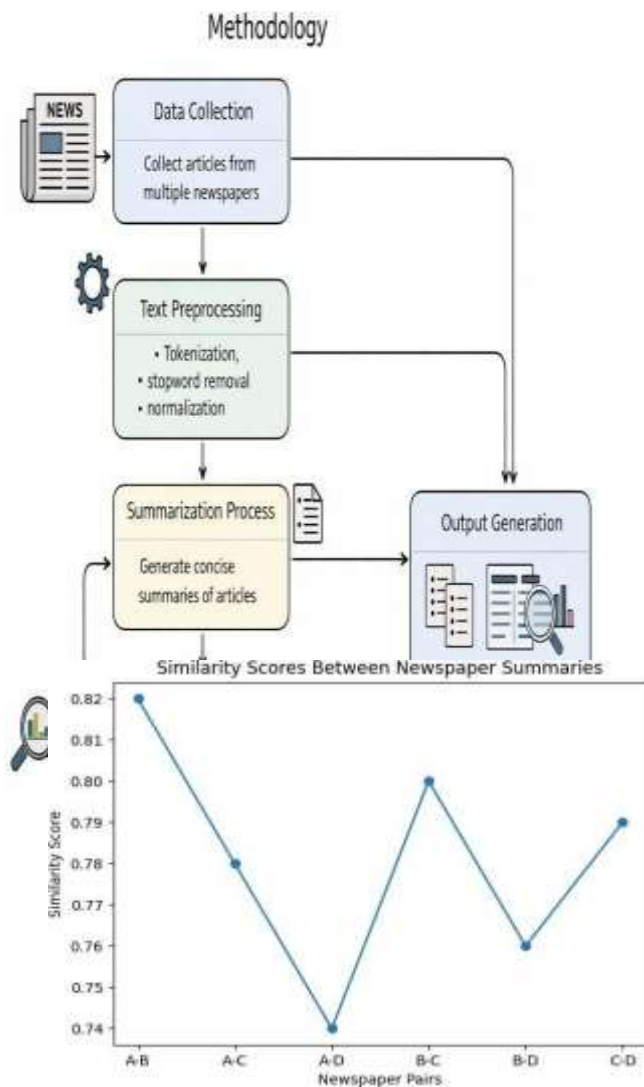
In addition to summarization, the system performs comparison among articles from different sources. The preprocessed text and generated summaries are analyzed using similarity measurement techniques to identify common information and differences in reporting. This analysis helps determine how various newspapers describe the same event and highlights variations in presentation and emphasis. The comparison module provides a clear understanding of overlapping content as well as unique viewpoints from each source.

E. Output Generation

In the final stage, the system integrates the results of the summarization and comparison modules to produce a comprehensive output. The output includes concise summaries of individual articles along with a comparative analysis view. This allows users to efficiently review multiple news reports and gain a balanced understanding of the event. The generated results are presented in a simple and readable format to support practical usability.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed system was implemented and tested using a collection of real-world newspaper articles gathered from different online news platforms. The primary aim of the experimental study was to assess the system's capability to generate accurate summaries and perform effective comparisons across news articles obtained from different sources. Various articles related to the same topic were selected to analyze the accuracy and reliability of the system.

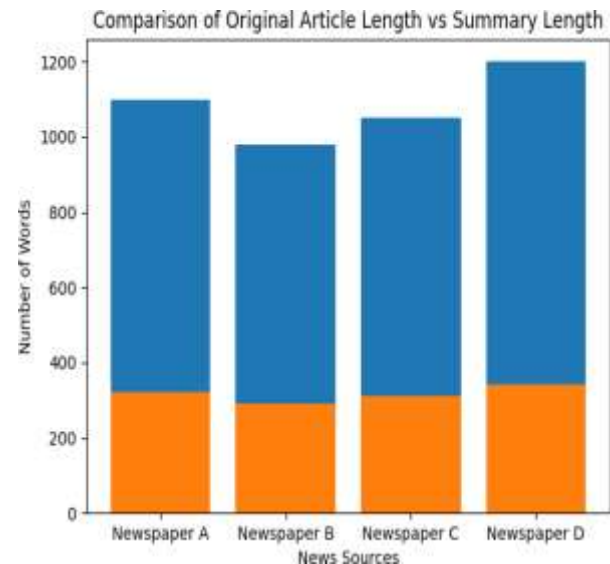


During testing, the input articles were first processed through the preprocessing module. This stage successfully cleaned the raw text by removing unnecessary words and formatting issues. The preprocessing operations helped in converting the unstructured news content into a standardized form, which improved the performance of the later stages. It was observed that proper preprocessing played a significant role in enhancing the quality of both summarization and comparison results.

Overall performance analysis indicates that the proposed system functions effectively for multi-source news analysis. The integration of preprocessing, summarization, and comparison modules helped in producing meaningful outputs. The system was able to reduce information overload and assist users in

understanding news content in a faster and more convenient manner.

COMPARISON OF ORIGINAL ARTICLE LENGTH VS SUMMARY LENGTH:



SIMILARITY SCORES BETWEEN NEWSPAPER SUMMARIES:

VII. CONCLUSION

The proposed Multi-Source Newspaper Summarization and Comparison System effectively simplifies the process of analyzing news from multiple sources. By using Natural Language Processing techniques, the system generates concise summaries and compares articles to highlight similarities and differences. This helps users quickly understand key information while reducing redundancy and information overload. The system improves the efficiency of news reading and provides a clear, organized view of multi-source content. Future enhancements can include advanced NLP models to further improve accuracy and performance.

VIII. REFERENCES

- [1] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*, O'Reilly Media, 2009.
- [2] M. Honnibal and I. Montani, “spaCy 2: Natural Language Understanding with Bloom Embeddings,” *Explosion AI*, 2017.
- [3] R. Nallapati, F. Zhai, and B. Zhou, “Summarunner: A Recurrent Neural Network Based Sequence Model for Extractive Summarization,” in *Proceedings of AAAI*, 2017.
- [4] NewsAPI, “News API – Documentation,” [Online].
Available: <https://newsapi.org>.
- [5] J. Alwani, “Newspaper3k: Automatic Article Extraction Library,” GitHub Repository, 2020.
- [6] I. Mani, *Automatic Summarization*, MIT Press, 2001.