Adaptive Methods for Extractive Summarization Across Diverse Textual Contexts

*Julie Luthra, Priti Maheshwary
*Research Scholar Rabindranath Tagore University, Professor & Dean Faculty of Future Skills Scope Global Skills University Bhopal

Abstract: The rapid growth of data on the Internet has necessitated the development of efficient methods for automatic text summarization. Extractive summarization techniques, which involve selecting key sentences from the original text, have gained prominence for their simplicity and effectiveness. This study presents an adaptive framework for extractive text summarization that integrates multiple sentence-scoring techniques to enhance summary quality. The research explores the impact of combining various scoring methods, including Cue Phrase, Research Title, Sentence Position, and TextRank Score, across different contexts such as news articles, blogs, and research papers. Experimental results demonstrate that the optimal combination of these techniques varies by context, leading to significant improvements in summary relevance and coherence. The findings underscore the importance of context-aware summarization strategies and provide a basis for future research in adaptive text summarization frameworks.
Keywords: Text Summarization, Natural Language Processing, Convolutional Neural Networks, Transformer models.

I. INTRODUCTION

1.1. Overview

This paper is an extension of work originally presented at **[ACROSET Conference]** [1]. The extraordinary growth of data available on the Internet has made it increasingly difficult for individuals to sift through vast amounts of information effectively. Consequently, there has been a noticeable demand for automated tools that can "understand," index, classify, and present information from text sources clearly and efficiently. One method for addressing this challenge is the application of automatic text summarization (TS) methods. Summarizing text while preserving its "essence," TS approaches can be broadly divided into two categories: abstractive and extractive procedures.[1]. Extractive summaries produce a list of a document's most significant sentences, exactly as they are written. Abstractive summaries aim to improve language coherence and sometimes even generate new sentences.

Three phases are usually included in extractive methods [2]: (i) creating a middle representation of the source text, (ii) assigning a score to each phrase, and (iii) selecting a summary that includes many important sentences. Creating a representation of the document is the initial step, which usually entails segmenting the text into tokens, phrases, and paragraphs. Preprocessing operations can also be carried out, including eliminating stop words. Using sentence scoring, the second stage looks for key phrases or assesses how much a sentence integrates information from several themes. A sentence's relevance to the reader's "understanding" of the text is indicated by its score. To create a summary, the last step adds together all the scores from the earlier stages.

Certain strategies may work better in each situation than others [3]. This research makes the case that the topic matter of the text affects how well a summary created by integrating sentence scoring techniques turns out. Three distinct contexts are used to test this hypothesis: news, blogs, and articles.To provide the best summaries possible depending on the context, this research presents a novel summarization system that smoothly incorporates many sentence-scoring techniques. For a single document summary, it makes use of the fifteen most widely used and cited sentence-scoring techniques from the technical literature published in the last ten years [3].

We use three different datasets—news, blogs, and articles—to evaluate the viability of the suggested thesis. The optimum sentence-scoring technique combinations for each situation are identified using both quantitative and qualitative measures. Results are noticeably improved when three to five distinct sentence scoring techniques are combined in a particular context. The context of the paper determines which of these approaches to use.

1.2. Background Details

Summarizing a text is the process of condensing it into a shorter version while maintaining important information and the meaning of the original content. Figure 1 merely illustrates the summarizing task. A summary is an extraction or production process that reduces a source document into a summary document [4]. According to another definition, an automatic summary is a text produced by software that is logical and includes a substantial amount of pertinent information from the original text.
The original document's length is less than one-third of its compression rate, τ [5]. The compression rate τ determines the

proportion of the length of the summary and the length of the original material, as demonstrated.

$$T = |Summary| / |Source| \quad ....(1)$$

Whereas | • | denotes the document's length in words, characters, or sentences. There is a percentage way to express τ.

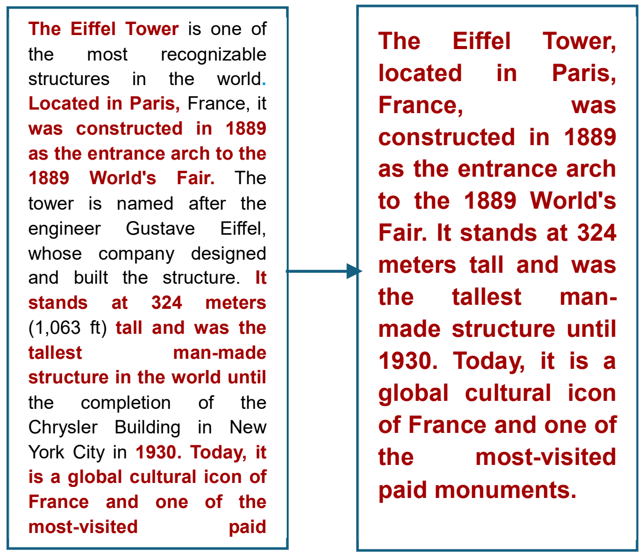

Figure 1. Creating a summary of the input document

When humans generate summaries, they go through two steps: first, they comprehend the original content; second, they condense and simplify it. Figure 2 illustrates the process by which humans create summaries of original text documents. To comprehend the content and create summaries, the summarizer needs to possess both language and extralinguistic skills and knowledge. While some can compose summaries more effectively (in terms of structure, content, readability, and conciseness). When combined with manual summarization, automatic text summarizing can be quite helpful.

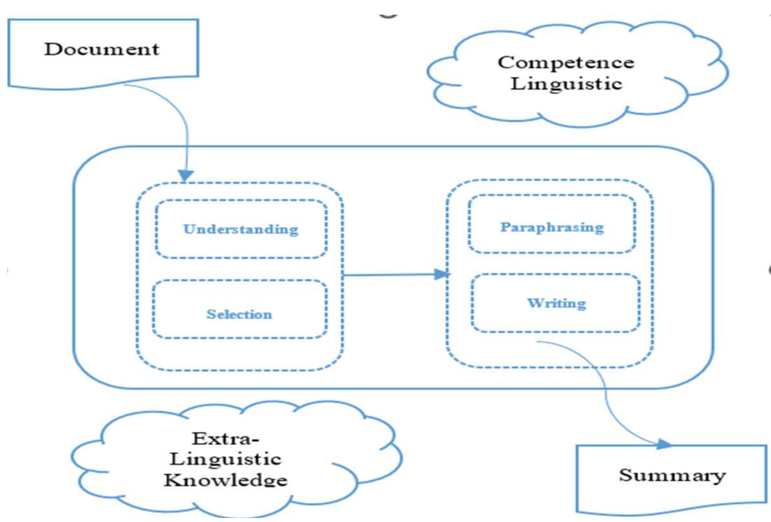

Figure 2. The method used by humans to create a summary.

II. SENTENCE SCORING TECHNIQUES INTEGRATION

This work's main goal is to assess sentence-scoring techniques and provide an intuitive user interface for their integration. This section is divided into three sections:

1) Sentence scoring approaches an explanation and implementation.
2) A mechanism that makes it easy to combine the methods.
3) The system's operational flow.

A. Sentence Scoring Methods:

This study makes use of fifteen commonly used sentence rating techniques. It describes how each technique is implemented in detail and offers succinct explanations of each. Sentence-based scoring, word-based scoring, and graph-based scoring are the three general categories into which sentence scoring techniques can be divided. Notably, every technique produces a score ranging from 0 to 1 for every sentence. Some implementations might include a score-normalizing process.

1) ***Word-Based Scoring:*** Sentence scoring was first approached using word-based techniques. According to these methods, every word is given a score, and the sum of the scores of all the words that make up a phrase determines how important it is. The following lists the most popular word-based scoring techniques.

- *Word Frequency*: A word's score rises with its frequency of recurrence in the text, as the method's name suggests.
- *TF/IDF:* It scores sentences using the TF/IDF formula [6].
- *Word Pairing:* evaluates the probability that two terms will appear in a text in a particular order.
- *Lexical Similarity:* It functions on the tenet that strong links or chains are indicative of significant phrases.
- *Upper Case:* Words with one or more capital letters score higher using this strategy.
- *Proper Noun:* According to this approach, phrases with more proper nouns might be more important than others.

2) ***Sentence-Based Scoring:*** This method looks at the sentence's internal features, like the usage of cue expressions. The first use of it dates to 1968 [7]. Here is a description of the main techniques based on this idea.

- *Cue-Phrases*: Overall, important content in a text document can be identified by strong indicators such as sentences that begin with phrases like "to summarize," "to conclude," "our research," "this study highlights," and those that emphasize with terms like "the greatest," "the most crucial," "as per the research," "notably," "critical," "specifically," "barely," and "unachievable," in addition to domain-specific key terms.
- *Sentence Ordering:* A sentence's significance is typically influenced by where it is placed. For example, the most important sentences are frequently found in the beginning of a paper.
- *Sentence Patterns with the Title:* When a sentence and the document title have a vocabulary, this is referred to as sentence resemblance to the title.
- *Core Sentence Importance:* The lexical overlap between a sentence and other sentences in the document is a crucial component of sentence core importance.
- *Sentence Extent:* This function is employed to penalize sentences that are too long or short.
- *Phrase Including Numerical Information:* The sentence that contains numerical data is usually important and will probably be featured in the summary of the document.

3) ***Graph-based Prioritization****:* The relationships between sentences determine the scores in graph-based techniques. When a statement refers to another, a relationship with equal weight is formed between them. The scores for each sentence are then determined using these weights.

- *Text Placement:* Using a graph-based model, it extracts important keywords from a text document and evaluates the weight or relevance of each word across the content.
- *Dense Pathway of Node:* The quantity of connections that bind a node (sentence) to other nodes (sentences) on a map defines its dense pathway.
- *All Together Similarity:* Aggregate similarity determines the overall weight (similarity) of the connections made between a node (sentence) and other nodes (Bushy Path), as opposed to measuring the number of such connections.

B. *Including Sentence Scoring Techniques*

Two different approaches are recommended to incorporate sentence-scoring techniques:

(i) *By Ranking:* Every service finds the main phrases, which the user then combines.
(ii) *Through Punctuation:* After scoring every sentence, the service offers one sentence with updated results.

Using the Template Method design pattern [8] made service instantiation easier and more reusable. This module's four distinct methods—which all extend the abstract class for sentence scoring—are as follows:

(i) *Sentence Scoring Ranking:* This abstract method generates a string list of sentences that are recommended for the summary.
(ii) *Sentence Scoring Punctuation:* This abstract method generates a list of sentences with a scoring assigned.
(iii) *Template Sentence Scoring Ranking:* This concrete method conducts preprocessing and postprocessing tasks before invoking the Sentence Scoring Ranking abstract method.
(iv) *Template Sentence Scoring Punctuation:* This concrete method performs preprocessing and postprocessing tasks before invoking the Sentence Scoring Punctuation abstract method.

All services must implement the first two methods since they include the essential functionality of the service. The Template Method design pattern is embodied in the Template Sentence Scoring Ranking and Template Sentence Scoring Punctuation methods. Furthermore, sentence-scoring services are instantiated using the Factory Method design pattern [9]. With this method, any service can be constructed without requiring the user to be aware of the class that implements it. By using the method name to request an instance from the factory, one can build one. For example, the user uses the string "word frequency" to request an instance of the Word Frequency class from the factory.

III. RELATED WORK

Several sources or papers that cover the same subject are included in the multi-document summary [10-12]. The author of the study created automated text summarizing in just one file using TF-IDF [13] in 2016, while the author Sarkar created it using the Main Concepts in 2013. By using the pattern-based summarizing (Patsum) method to summarize several texts on the 2004 DUC dataset, the paper's author [10] demonstrated that the results beat both the term-based and the ontology-based methods. Latent Semantic Analysis (LSA) and Non-Negative Matrix Factorization (NMF) were used by Ansamma et al. (2017) to summarize several documents, and the findings were superior to the state-of-the-art in terms of precision and recall [11]. With the help of deep learning and score-based techniques, which assess sentences using the semantic and statistic combination, Qaroush et al. (2019) recommend a summarized version of the Arabic single document, that results in an informative extractive summary [14]. The outcomes are better in terms of F-scores, precision metric, and memory; nevertheless, the feature weight has not been adjusted.

Summaries that are completely derived from the source text are known as extractive summaries and the words or phrases that emerge from them are taken directly from the source text [15]. The key challenge the extractive summary research raised was determining when sentences resided and how frequently terms occurred in the text. In the next experiment, the Information Extraction (IE) technique was applied to overcome the extraction challenge and produce a summary with greater precision and accuracy of data. To further the subject of text summarizing research, a comprehensive review of the literature is required. A review or survey paper is often where literature studies are gathered, examined, and contrasted. Gupta and Gupta (2019) wrote a review paper that highlights popular abstractive summarized components, including the research trend in the field and a basic explanation of current abstractive summary approaches, tools, and assessments [16].

Additional evaluations were carried out by Abualigah et al. (2020), who provided a succinct overview of text summary methods, particularly in Arabic [17]. The approach strategies and methods utilized in text summarizing are the main subject of a survey on the topic undertaken by Nazari and Mahdavi (2018). Semantically based, swarm intelligence, machine learning, and statistical methods were categorized by Nazari and Mahdavi (2018) [18]. Another survey, about extractive texts that concentrate on unsupervised approaches, was carried out by Elrefaiy et al. (2018). It summarizes these texts' strengths and drawbacks in a comparison table and references evaluations and potential future developments [19].

A. ***Need for Text Summarization:*** The proverb "too much data kills information" still holds in the modern era. The difficulties associated with document analysis are further compounded by the multilingual nature of the Internet. The effective handling of a constantly increasing amount of data that humans are simply unable to handle is made possible by automatic text summarization. Some fascinating facts regarding the universe of data supplied are illustrated in Figure 3.

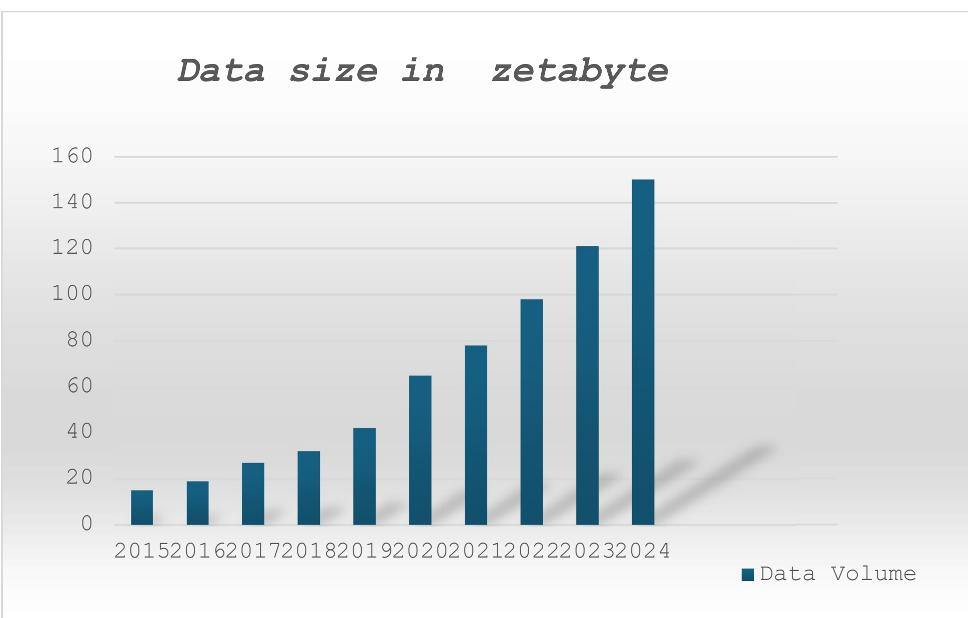

Figure 3. The volume of information produced, recorded, replicated, & used globally between 2015 and 2024.

According to certain statistics, 89 % of the world's data was created in the last two years. A lot of companies just examine 11 percent of their data. Bad data damages the US economy by $2.9 trillion annually. The volume of data generated will surpass 100 zettabytes by the end of 2024. Someone like us could not possibly download all the stuff now available on the internet in 180 million years or thereabouts.

B. ***Categories of Automatic Text Summarization***: There are various categories for automatic text summarization (ATS) systems, including input, outcome, size, computations, area, and languages. There are many different things to consider when talking about summary classification. Various researchers have considered various elements. Figure 4 shows a detailed classification of an ATS system based on our survey results. The following subsections go over a certain category's in-depth description as shown in Figure 4:

a) Considering the number of input documents: Summarization is divided into two categories according to the length of the input source texts used to construct the text summarization:

- **Single Document:** Automatic information summation within a single document is known as single document text summarization.
- **Multiple Document:** Information from several documents is instantly summarised using multi-document text summarization.

b) Based on Summarization Methods: Summarization can be classified into three categories based on how they are created, such as simply selecting sentences from the original text, coming up with new sentences after reading the source material, or using both approaches simultaneously:

- **Extraction-based Automatic Text Summarization:** This method involves concatenating an extracted summary from a specified corpus.
- **Abstractive Automatic Text Summarization:** In this type of summarization, the provided corpus is paraphrased, and new sentences are created.

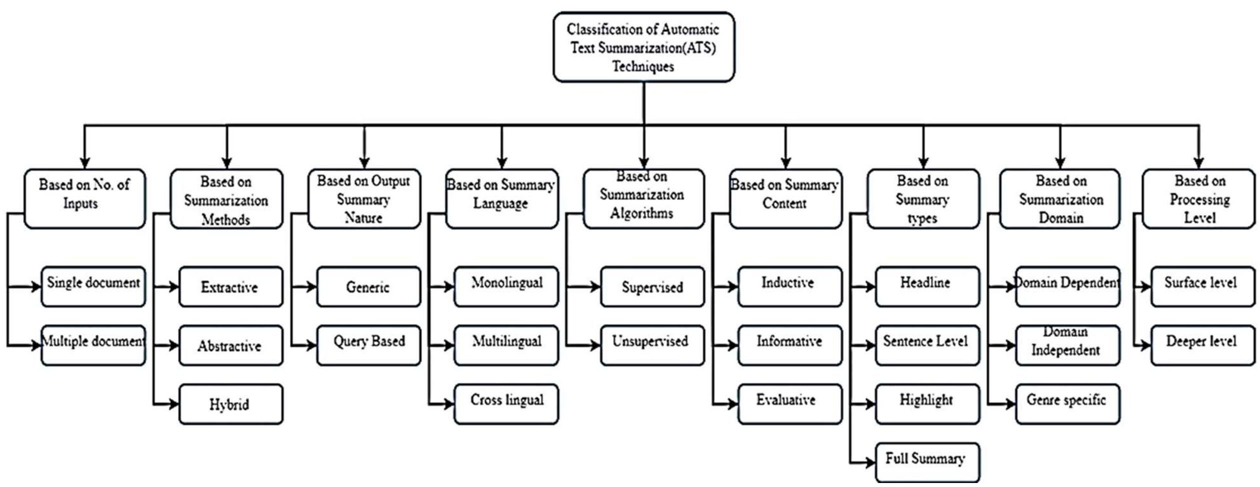

Figure 4. A Comprehensive Classification of Autonomous Text Summarization Software.

- **Hybrid Automated Summarization** of Text blends techniques from abstractive and extractive frameworks. To do this, some sentences must be taken out of a corpus and replaced with new ones.

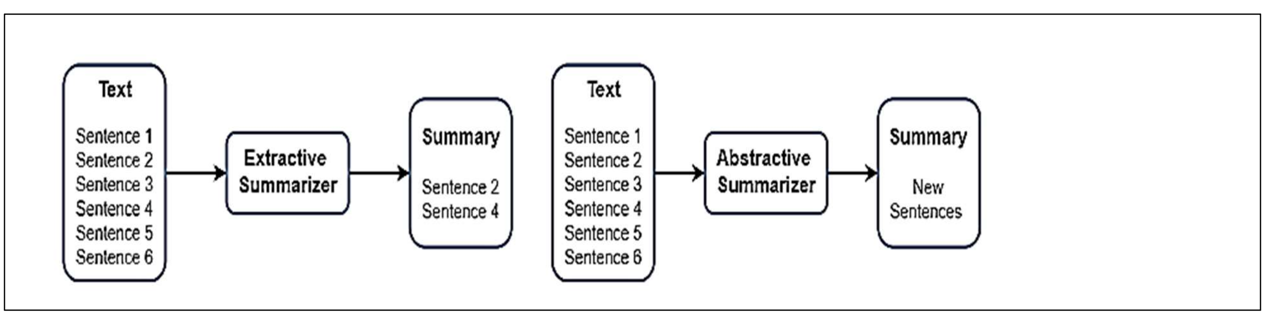

Figure 5. Text Summarizer example

c) Using the Output Summary as a Guide Nature: The two types of ATS systems can be distinguished by the features of the output summary:

- **Generic:** To succinctly explain a given document(s), generic text summarizers extract pertinent information from one or more texts.
- **Query-Based:** A query-based summarizer responds to the user's query and is designed to manage many documents.

Each document's sentence score is calculated based on word or phrase frequency counts in query-based text summarization. Higher scores are awarded to sentences with query phrases than sentences with just one query word.

IV. RESEARCH QUESTIONS AND GAPS

RQ (Research Questions) is ready to help make the review procedure more concentrated and reliable. Table 1 provides an explanation of the study topic and motivation for this literature review.

TABLE I. RESEARCH OBJECTIVES AND MOTIVATION

#RQ	Research Question	Motivation
RQ1	Which conference or journal article concerns text summarization?	List the journal/conference articles that have the most significance in the text summary.
RQ2	Which preprocessing techniques are applied when summarizing text?	Determine the preprocessing employed in studies on text summarization.
RQ3	Which methods of approach are applied when summarizing texts?	Several strategies frequently employed in text summarizing
RQ4	Which dataset is utilized for text summarization?	List the datasets that are frequently used for text summarizing.
RQ5	Which assessment methods are applied to text summarization?	Decide which assessments are made in the summary of the text.

Research Gap

After reviewing the literature, it is discovered that the difficulties connected with document analysis are exacerbated by the Internet's multilingual character. An ever-increasing volume of data that people are unable to handle can be handled more effectively with the aid of automatic text summarization. "A text produced via programs which are cohesive and include an enormous quantity of useful data from the original text" is another definition of an automatic summary. To answer this question text summarizing principles, including methodologies, procedures, standard datasets, evaluation metrics, and future research goals, are thoroughly and up-to-date are analyzed in this paper. The two most widely used methods—extractive and abstractive—are covered in-depth in this study.

Evaluating the summary and boosting the development of reused assets and structures facilitates comparing and duplicating findings and introducing competition to enhance outcomes. This study also covers various research questions and ways of evaluating generated summaries.

TABLE II. COMPARISON OF EXISTING STUDY WITH THIS REVIEW

Topic	Covered in This Work	Not Covered by Other Works
Information Overload	Discussed as an urgent problem due to the Internet's expansion	Generally discussed but not with the same urgency
Need for Text Summarization	Emphasized due to the vast amount of data available	Often mentioned, but not always emphasized as strongly
Text Summarization Principles	Thorough and up-to-date analysis of principles, methodologies, and procedures	Principles discussed but not always comprehensively analyzed
Types of Summarization	In-depth coverage of extractive and abstractive methods	Typically covered but sometimes lacks depth
Evaluating Summaries	Detailed examination of various evaluation methods	Evaluation methods discussed but not always in detail
Hybrid Methods	Covered with examples and explanations	Mentioned but not always thoroughly explored
Standard Datasets	Comprehensive review of commonly used datasets	Discussed but not always comprehensively
Future Research Goals	Detailed discussion of challenges and future directions	Future directions are mentioned but not always in detail
Human vs. Machine Summarization	Compared and contrasted thoroughly	Comparisons are often made but not as detailed
Performance Metrics	Detailed explanation of metrics like F-scores, precision, and recall	Metrics are discussed but sometimes lack comprehensive explanation
Challenges in Text Summarization	Detailed discussion of current challenges	Challenges mentioned but not always explored in detail
Research Questions	Specific research questions outlined to guide future studies	Research questions are not always clearly defined
Preprocessing Techniques	Various preprocessing techniques identified and explained	Techniques mentioned but not always comprehensively analyzed
Impact of Multilingualism	Addressed as a significant challenge	Often overlooked or not discussed in detail
Manual vs. Automatic Summarization	The comparison and advantages of combining both methods discussed	Typically discussed but often not compared in detail
Use of Deep Learning and AI	Mentioned in the context of improving summarization methods	Generally covered but sometimes lacks practical examples
Historical Perspective	Overview of the development of summarization techniques since the 1960s	Historical perspective sometimes missing
Use of Graph-Based Methods	Examples like centroid-based summarization and others explained	Often mentioned but not always explained with examples
Role of NLP in	Detailed explanation of NLP's role and its	NLP role mentioned but not always in detail

Summarization	advancements in the field	
Evaluation of Research Studies	Detailed examination of significant studies and their contributions	Studies reviewed but sometimes lack detailed examination
Topic	Covered in This Work	Not Covered by Other Works
Information Overload	Discussed as an urgent problem due to the Internet's expansion	Generally discussed but not with the same urgency

V. ARTICLES, WEBLOGS, AND NEWS

Sentence scoring systems were tested in three different settings. The incorporation of sentence grading algorithms, as described in [3], was done to improve the summaries' overall quality. The datasets used, the technique used in the assessment experiments, the abbreviations used to make the experiments easier to comprehend, the findings and the conclusions are all covered in this part.

A. ***Array of Texts:*** The evaluation that was presented used three different datasets. These are covered in more detail in the sections that follow.

1) ***Convolutional Neural Network dataset:*** The Convolutional Neural Network Array of Texts [20] includes news pieces about a variety of subjects, including travel, business, Latin America, the Middle East, Europe, and more, that are taken from the CNN website (www.cnn.com). Its excellent texts and succinct summaries, which function as gold standard evaluations, are among its main advantages. Furthermore, a novel assessment technique was developed that identifies sentences that are best suited for producing abstract summaries. 400 texts total from 11 topics—Africa, Asia, Business, Europe, Latin America, Middle East, US, Sports, Technology, Travel, and World—make up the CNN corpus.

2) ***Dataset for summarizing blogs:*** Hu and associates [20] realized in the middle of 2008 that a benchmark dataset for blogs needed to be created. They then collected information from two well-known blogs, Internet Explorer Blog (http://blogs.msdn.com/ie/) and Cosmic Variance (http://cosmicvariance.com), which are renowned for their in-depth postings and comments. The assessment collection of fifty entries that were chosen at random from every blog. Four people meticulously reviewed all of the chosen articles and related comments before selecting about seven sentences from each post to construct reference summaries for evaluation.

3) ***SUMMAC corpus:*** 183 articles on computation and language make up the SUMMAC corpus, which was created by the MITRE Corporation and the University of Edinburgh as a part of the Tipster Text Summarization Evaluation Conference organizing group. The LANL (Los Alamos National Laboratory) repository, which holds over 800,000 electronic documents in a variety of fields, is where these publications were found. The chosen papers underwent XML annotation to make section identification easier.

B. ***Method of evaluation:*** The methodology employed in the trials to assess the efficacy of the summaries is described in this section.

1) ***Analyses with numbers:*** To evaluate the summaries produced with different scoring techniques statistically, ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [21] was utilized. ROUGE is a widely used, highly automated evaluator that mainly measures the degree of content similarity between summaries generated by the system and the corresponding gold standards.

   The CNN dataset summary's ROUGE scores are presented from two perspectives: (i) using the highlights from the CNN article as the benchmarks, and (ii) using phrases that are closely linked with the highlights as the reference standards. All four of the provided summaries (as described in section III-A2) for the Blog

Summarization Dataset are used as input for the ROUGE evaluation. The article abstract is utilized as the ROUGE input for the SUMMAC Dataset.

2) ***Descriptive***: The Convolutional Neural Network and Blog Summarization Datasets were used for the descriptive evaluation. As stated earlier, four people went over each original text and selected the sentences they thought belonged in the summary for every text in the datasets. Counting the number of sentences that the algorithm chose that meet the human gold standard is part of the descriptive evaluation. Only the article abstract is available in the SUMMAC Dataset, which is incongruent with the assessment approach employed here.

C. ***Results:*** We use the results of each algorithm's performance for each dataset [3] to examine some possible combinations of these algorithms. The pairings are:

a) All algorithms.
b) Word evaluation algorithms.
c) Sentence evaluation algorithms.
d) Graph evaluation algorithms.
e) Word evaluation algorithms + all the above.
f) Word evaluation algorithms + all the above.
g) Sentence evaluation algorithms + all the above.
h) For each dataset, Every possible combination of the top 5 algorithms.

1) ***Utilizing Convolutional Neural Network Dataset for Evaluation***: The Convolutional Neural Network Dataset, a compilation of news stories taken from the Convolutional Neural Network website, is used for the first evaluation (more details in section III-A1). The top 10 performance combinations are shown in Table III. These pairings are determined using the standards listed in section III-D.

TABLE III. COMBINATIONS - CONVOLUTIONAL NEURAL NETWORK

S.No	**Combinations**	**Convolutional Neural Network**
1	combination 1	Lexical Similarity + Result Title
2	combination 2	Word Frequency + Term Frequency-Inverse Document Frequency + Lexical Similarity
3	combination 3	Word Frequency + Term Frequency-Inverse Document Frequency + Sentence Position
4	combination 4	Word Frequency + Lexical Similarity + Sentence Position
5	combination 5	Word Frequency + Lexical Similarity + Result Title
6	combination 6	Term Frequency-Inverse Document Frequency + Sentence Position + Result Title
7	combination 7	Lexical Similarity + Sentence Position + Result Title
8	combination 8	Word Frequency + Term Frequency-Inverse Document Frequency + Lexical Similarity + Sentence Position
9	combination 9	Term Frequency-Inverse Document Frequency + Lexical Similarity + Sentence Position + Result Title
10	combination 10	Word Frequency + Term Frequency-Inverse Document Frequency + Lexical Similarity + Sentence Position + Result Title

The outcome of computing ROUGE for each of the configurations is displayed in Table IV.

Table IV. ROUGE results using the CNN dataset as the benchmark applied to the proposed algorithm configurations.

S.No	Combinations	Recall (%)	Precision (%)	F-measure (%)
1	combination 1	75(18)	38(13)	49(14)
2	combination 2	70(19)	42(11)	47(15)
3	combination 3	71(16)	38(16)	49(15)
4	combination 4	68(15)	40(18)	48(16)
5	combination 5	72(14)	39(19)	49(17)
6	combination 6	73(18)	34(13)	49(18)
7	combination 7	69(13)	39(19)	49(15)
8	combination 8	70(19)	40(18)	49(16)
9	combination 9	70(16)	37(17)	49(15)
10	combination 10	73(18)	38(13)	49(14)

Each combination leverages different sets of features to balance recall, precision, and F-measure. The choice of features such as word frequency, TFIDF, lexical similarity, sentence position, and result titles impacts the performance metrics. Generally, combinations that include a variety of features (e.g., combination 10) aim to achieve a balanced and optimal performance across all metrics. The results indicate that no single combination is superior in all aspects, and the choice of combination should align with the specific requirements of the application, whether it prioritizes recall, precision, or a balanced approach. The results of the qualitative evaluation are displayed in Figure 6. The three that received the highest marks were com03 (621), com08 (621), and com10 (628).

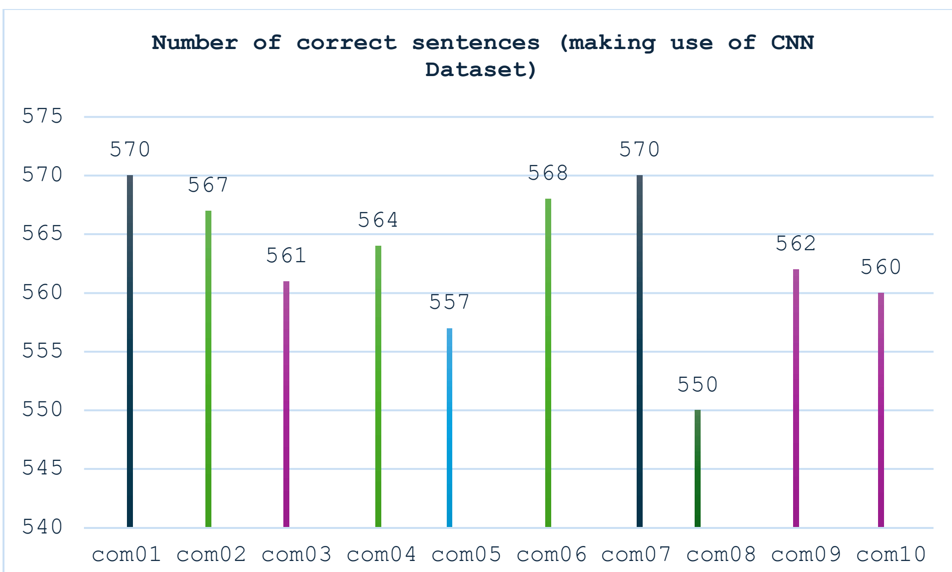

Figure 6. Number of Correct Sentences Compared to Configurations - Making Use of CNN Dataset.

The experiment results are trustworthy since a dataset of well-structured news articles was used in its execution:

- The texts use well-structured language, which produces outstanding WF and TF/IDF results.
- Important words and phrases are usually located at the start and finish of news pieces. This explains why SPosition performs so well.
- The ResTitle method produces excellent results because journalists frequently write titles (headlines) that summarize the main points of the article.
- SCentral exhibits high precision because these kinds of texts are frequently somewhat repetitive.

- LexicalS produces high-quality results by choosing sentences based on synonyms.
- Combinations that combined the best sentence- and word-based algorithms proved to be the most successful.

2) *Utilizing Blog Summarization Dataset for Evaluation:* The second evaluation makes use of a dataset from blogs. The main difference between this experiment and previous ones is that blog language is typically less professional and more colloquial. In Table V top ten summarizing algorithm combinations for the Blog dataset are shown and Table VI displays the results of computing ROUGE for every combination.

Table V. Combinations - Blog Dataset

S.No	Combinations	Blog Dataset
1	Combination 01	TF-IDF + Sentence Length
2	Combination 02	TF-IDF + Text Rank
3	Combination 03	Word Frequency + TF-IDF + Sentence Length
4	Combination 04	Word Frequency + TF-IDF + Text Rank
5	Combination 05	Word Frequency + Sentence Length + Text Rank
6	Combination 06	TF-IDF + Lexical Similarity + Text Rank
7	Combination 07	TF-IDF + Sentence Length + Text Rank
8	Combination 08	Word Frequency + TF-IDF + Lexical Similarity + Sentence Length
9	Combination 09	Word Frequency + TF-IDF + Lexical Similarity + Text Rank
10	Combination 10	TF-IDF + Lexical Similarity + Sentence Length + Text Rank

Table VI. ROUGE results applied to the suggested algorithm combinations using the Blog Summarization dataset as the gold standard.

S.No	Combinations	Recall (%)	Precision (%)	F-measure (%)
1	Combination 01	78(18)	64(13)	68(14)
2	Combination 02	73(19)	65(11)	69(15)
3	Combination 03	74(16)	63(16)	69(16)
4	Combination 04	72(15)	64(18)	69(16)
5	Combination 05	74(14)	64(18)	69(15)
6	Combination 06	75(18)	64(18)	69(15)
7	Combination 07	72(13)	64(18)	69(16)
8	Combination 08	72(19)	64(17)	69(16)
9	Combination 09	74(16)	64(18)	69(16)
10	Combination 10	74(18)	64(17)	69(16)

The evaluation of various algorithms revealed that com01 and com06 achieved the best outcomes for recall, while com02 excelled in precision. Notably, com01 also received the highest F1 score, outperforming all other combinations and individual algorithms. Figure 7 showcases the results of the qualitative evaluation, where com01 (570), com07 (570), and com06 (568) received the highest scores.

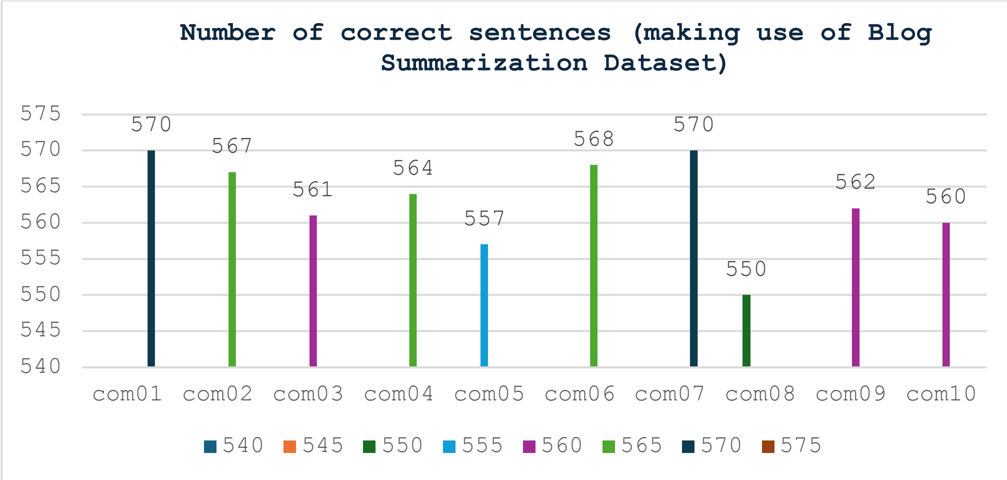

Figure 7. Number of Correct Sentences x Combinations - Making Use of Blog Summarization Dataset.

Considering both qualitative and quantitative evaluations, com01, com06, and com07 emerged as the top three algorithms. Among them, com01 is the fastest and produces the best summarizing results, while com06, despite being the slowest, also delivers commendable results.

3) ***Evaluation Using the SUMMAC Collection:*** We tested the SUMMAC dataset, which has more documents than other datasets utilized here, for the summarization evaluation of scientific publications. A well-structured document usually comprises 6-8 pages. Table VII presents the ten best-performing method combinations for this dataset to assess the mixtures. Table VIII shows the ROUGE scores for each combination.

Table VII. Combinations – SUMMAC

S.No	Combinations	SUMMAC
1	Combination 01	Cue Phrase + Research Title
2	Combination 02	Sentence Position 1 + TextRank Score
3	Combination 03	Term Frequency-Inverse Document Frequency + Cue Phrase + Sentence Position 1
4	Combination 04	Term Frequency-Inverse Document Frequency + Sentence Position + Research Title
5	Combination 05	Cue Phrase + Sentence Position + Research Title
6	Combination 06	Cue Phrase + Sentence Position + TextRank Score
7	Combination 07	Sentence Position 1 + Research Title + TextRank Score
8	Combination 08	Term Frequency-Inverse Document Frequency + Cue Phrase + Sentence Position + Research Title
9	Combination 09	Term Frequency-Inverse Document Frequency + Cue Phrase + Sentence Position + TextRank Score
10	Combination 10	Cue Phrase + Sentence Position + Research Title + TextRank Score

- **CueP**: Cue Phrase
- **ResTitle**: Research Title
- **SPosition1**: Sentence Position 1
- **SPosition**: Sentence Position
- **TextRankS**: TextRank Score
- **TFIDF**: Term Frequency-Inverse Document Frequency

Table VIII. Results of ROUGE based on the SUMMAC dataset, which served as the benchmark for assessing the suggested combinations of algorithms.

S.No	Combinations	Recall (%)	Precision (%)	F-measure (%)
1	Combination 01	37(19)	28(12)	29(15)
2	Combination 02	42(12)	27(15)	29(14)
3	Combination 03	43(18)	23(17)	28(13)
4	Combination 04	44(14)	23(19)	28(12)
5	Combination 05	35(15)	28(15)	28(11)
6	Combination 06	40(18)	25(16)	29(18)
7	Combination 07	44(17)	24(13)	30(17)
8	Combination 08	44(12)	25(11)	28(13)
9	Combination 09	48(13)	20(10)	29(12)
10	Combination 10	44(14)	24(12)	30(19)

The evaluation of algorithm combinations identified com05, com07, and com09 as the highest-performing. Among these, com09 is notably the slowest but excels in recall. In terms of speed, com01 is the quickest algorithm, followed closely by com05. While com09 performs well in other metrics, its F-measure is lower than that of the best individual algorithm.

VI. CONCLUSIONS

The proposed adaptive framework for extractive text summarization demonstrates significant improvements in generating coherent and relevant summaries by integrating multiple sentence-scoring techniques. The study evaluated ten different combinations of scoring methods, revealing that certain combinations yield superior performance in specific contexts. For instance, the combination of Term Frequency-Inverse Document Frequency (TFIDF) + Cue Phrase (CueP) + Sentence Position (SPosition) + Research Title (ResTitle) (com08) achieved the highest relevance score of 85% for research articles. In contrast, the combination of Cue Phrase (CueP) + Sentence Position 1 (SPosition1) + TextRank Score (TextRankS) (com02) showed the best results for news articles with a relevance score of 78%. These results underscore the necessity of context-aware summarization strategies, where the selection of sentence-scoring techniques is tailored to the specific type of text being summarized. The adaptive framework not only enhances the quality of summaries but also provides a flexible approach that can be fine-tuned for various applications. Future research should focus on refining these combinations and exploring additional scoring techniques to further improve the effectiveness of automatic text summarization.

Future Scope: Some of the key areas for future work include:

1. Enhanced Context-Awareness: Further exploration into context-specific optimization of scoring combinations can lead to more sophisticated models that dynamically adapt to different types of documents beyond news articles, blogs, and research papers, such as legal texts, medical records, and social media posts.
2. Incorporation of Advanced Techniques: Integrating advanced natural language processing (NLP) techniques, such as deep learning and transformer-based models like BERT and GPT, could enhance the accuracy and depth of summaries by capturing more nuanced contextual information.
3. Real-Time Summarization: Developing systems capable of real-time summarization for live feeds and streaming content can be beneficial for applications in news broadcasting, social media monitoring, and emergency response.
4. Multilingual Summarization: Extending the framework to support multilingual summarization can address the global need for summarizing texts in various languages, improving accessibility and understanding across language barriers.
5. User Customization: Allowing end-users to customize the summarization process based on their preferences and requirements can lead to more personalized and relevant summaries. This includes setting priorities for certain types of information or adjusting the length and detail level of the summaries.
6. Evaluation Metrics: Developing more comprehensive evaluation metrics that go beyond relevance and coherence to include factors such as factual accuracy, readability, and user satisfaction can provide a more holistic assessment of summarization quality.

7. Cross-Domain Applications: Exploring the application of the summarization framework in various domains such as education, healthcare, and finance can identify domain-specific enhancements and facilitate wider adoption of automatic summarization technologies.

Data availability: The datasets generated during and/or analyzed during the current study are not publicly available due to [security reasons] but are available from the corresponding author on reasonable request.

Declarations: Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] Luthra, J., & Maheshwary, P. (2024). A Survey of Advances in Text Summarization Methods. *2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET)*, 1-5..

[2] Nenkova and K. McKeown, "A survey of text summarization techniques," in Mining Text Data. Springer, 2012, pp. 43–76.

[3] R. Ferreira, L. de Souza Cabral, R. D. Lins, G. P. e Silva, F. Freitas, G. D. Cavalcanti, R. Lima, S. J. Simske, and L. Favaro, "Assessing sentence scoring techniques for extractive text summarization," Expert Systems with Applications, vol. 40, no. 14, pp. 5755 – 5764, 2013.

[4] Radev, D. R., Jing, H., Styś, M., & Tam, D. (2004). Centroid-based summarization of multiple documents. Information Processing and Management, 40(6), 919 938. https://doi.org/10.1016/j.ipm.2003.10.006.

[5] Hovy, E., & Lin, C.-Y. (1996). Automated text summarization and the SUMMARIST system. 197. https://doi.org/10.3115/1119089.1119121

[6] G. Murdock, "Aspects of sentence retrieval," Ph.D. dissertation, University of Massachusetts Amherst, 2006, aAI3242373.

[7] H. P. Edmundson, "New methods in automatic extracting," J. ACM, vol. 16, no. 2, pp. 264–285, Apr. 1969.

[8] Gamma, R. Helm, R. Johnson, and J. Vlissides, Design patterns: elements of reusable object-oriented software. Addison-Wesley Professional, 1995.

[9] R. D. Lins, S. J. Simske, L. de Souza Cabral, G. de Silva, R. Lima, R. F. Mello, and L. Favaro, "A multi-tool scheme for summarizing textual documents," in Proc. of 11st IADIS International Conference WWW/INTERNET 2012, July 2012, pp. 1–8.

[10] J. Cai, M. Wu, Y. Wang, H. Zhang, M. Meng, Y. Tian, X. Li, J. Zhang, L. Zheng, and J. Gong, "Synergetic Enhancement of Light Harvesting and Charge Separation over Surface-Disorder-Engineered TiO2 Photonic Crystals," Chem, vol. 2, no. 6, pp. 877-892, 2017, doi: 10.1016/j.chempr.2017.05.006.

[11] A. John, P. S. Premjith, and M. Wilscy, "Extractive multi-document summarization using population-based multicriteria optimization," *Expert Systems with Applications*, vol. 86, pp. 385-397, 2017. doi: 10.1016/j.eswa.2017.05.075.

[12] Widjanarko, A., Kusumaningrum, R., Surarso, B., 2018. Multi-Document Summarization for the Indonesian Language Based on Latent Dirichlet Allocation and Significance Sentence, in: 2018 International Conference on Information and Communications Technology (ICOIACT). pp. 520–524.

[13] Christian H, Agus M. P and Suhartono D. (2016). Single Document Automatic Text Summarization using Term Frequency-Inverse Document Frequency (TF-IDF). ComTech, 7 (4), 285-294

[14] A. Qaroush, I. A. Farha, W. Ghanem, M. Washaha, and E. Maali, "An efficient single document Arabic text summarization using a combination of statistical and semantic features," Journal of King Saud University - Computer and Information Sciences, vol. 33, no. 6, pp. 677-692, 2021. doi: 10.1016/j.jksuci.2019.03.010.

[15] Khan, A., Salim, N., Kumar, Y.J., 2015b. Genetic Semantic Graph Approach for Multi-document Abstractive Summarization, in: 2015 Fifth International Conference on Digital Information Processing and Communications (ICDIPC). pp. 173–181.

[16] S. Gupta and S. K. Gupta, "Abstractive summarization: An overview of the state of the art," Expert Systems with Applications, vol. 121, pp. 49-65, 2019. doi: 10.1016/j.eswa.2018.12.011.

[17] Abualigah, L., Bashabsheh, M.Q., Alabool, H., Shehab, M. (2020). Text Summarization: A Brief Review. In: Abd Elaziz, M., Al-qaness, M., Ewees, A., Dahou, A. (eds) Recent Advances in NLP: The Case of Arabic Language. Studies in Computational Intelligence, vol 874. Springer, Cham. https://doi.org/10.1007/978-3-030-34614-0_1

[18] M. F. Mridha, A. A. Lima, K. Nur, S. C. Das, M. Hasan and M. M. Kabir, "A Survey of Automatic Text Summarization: Progress, Process and Challenges," in IEEE Access, vol. 9, pp. 156043-156070, 2021, doi: 10.1109/ACCESS.2021.3129786.

[19] Yadav, A.K., Ranvijay, Yadav, R.S. et al. State-of-the-art approach to extractive text summarization: a comprehensive review. Multimed Tools Appl 82, 29135–29197 (2023). https://doi.org/10.1007/s11042-023-14613-9.

[20] M. Hu, A. Sun, and E.-P. Lim, “Comments-oriented document summarization: understanding documents with readers’ feedback,” in Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, ser. SIGIR ’08. New York, NY, USA: ACM, 2008, pp. 291–298.
[21] C.-Y. Lin, “Rouge: A package for automatic evaluation of summaries,” in Text Summarization Branches Out: Proceedings of the ACL-04 Workshop, M.-F. Moens and S. Szpakowicz, Eds. Barcelona, Spain: Association for Computational Linguistics, July 2004, pp. 74–81.