Fuzzy Systems and Soft Computing
ISSN : 1819-4362

A SYSTEMATIC STUDY ON TEXT SUMMARIZATION IN NATURAL LANGUAGE PROCESSING WITH RESPECT TO RECENT ADVANCES AND CHALLENGES

Mrs. J. Joice Research Scholar, PG and Research Department of Computer Science, Government Arts and Science College, Kangeyam.
Dr. C. Sathya Assistant Professor, PG and Research Department of Computer Science, Government Arts and Science College, Kangeyam.

Abstract
Text summarization in Natural Language Processing (NLP) has gained significant attention due to the exponential growth of digital information. NLP research involves various aspects such as collecting and publishing research data, utilizing machine learning and deep learning techniques, and addressing challenges like understanding words in context and cultural differences. As the volume of textual data continues to surge, automated summaries have become crucial for efficient information retrieval and consumption across diverse domains. NLP systems involve input and output structures of speech and written text, with two main components: Natural Language Understanding (NLU) and Natural Language Generation (NLG). This paper reviews the key concepts, techniques, and advancements in text summarization. Examine extractive and abstractive summarization approaches, discuss their underlying methodologies, and highlight recent advancements, including neural network-based models and pre-trained transformers. Additionally, explore the evaluation metrics and challenges associated with text summarization, concluding with potential future directions in this dynamic field.

Keywords: Text Summarization, NLP, Extractive, Abstractive, Neural Network

1. Introduction
Natural Language Processing (NLP) encompasses a broad range of studies focusing on the interaction between computers and human language, aiming to enable machines to comprehend, interpret, and generate human language effectively [1]. NLP research involves various aspects such as collecting and publishing research data, utilizing machine learning and deep learning techniques, and addressing challenges like understanding words in context and cultural differences [2]. Furthermore, NLP extends to supporting requirements engineering by applying NLP techniques to analyze linguistic aspects of requirements documents, detect language issues, and establish traceability links between requirements, showcasing the interdisciplinary nature of NLP research involving computer science, linguistics, logic, and psychology [3]. This comprehensive approach highlights the significance of NLP in transforming industries, enhancing human-machine interactions, and advancing computational modeling of language across various domains. NLP systems involve input and output structures of speech and written text, with two main components: Natural Language Understanding (NLU) and Natural Language Generation (NLG) as in fig. 1.1.NLP encompasses several key components essential for processing and understanding human language. These components include Natural Language Understanding (NLU), speech recognition, syntactic analysis, semantic analysis, pragmatic analysis, and speech synthesis.

1.1.Text Summarization in NLG
Natural language generation (NLG), on the other hand, focuses on creating human language. It takes structured data and converts it into natural-sounding text or speech that humans can easily understand. This involves tasks like Text Summarization which generating a concise summary of a longer text. Text Summarization (TS) is the process of distilling the most important information from a source text to produce a concise and coherent summary [4]. As the volume of textual data continues to surge, automated

summarization has become crucial for efficient information retrieval and consumption [5]. TS play a crucial role in distilling large volumes of textual data into concise and meaningful summaries, aiding in efficient comprehension and utilization of information [6].

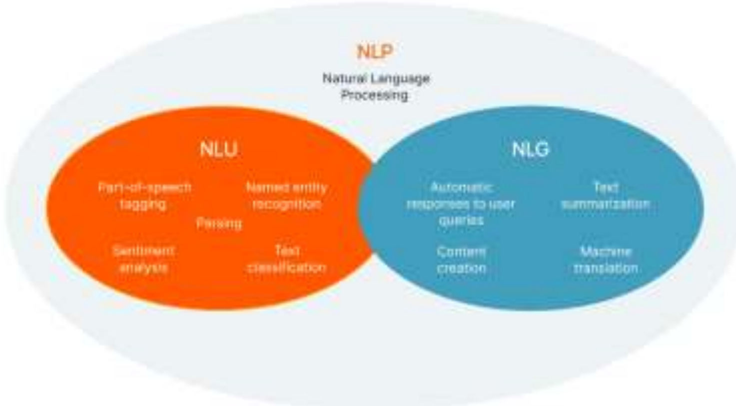

Figure: 1.1 NLP Main Components

Various approaches, such as extractive and abstractive summarization techniques, are employed to generate summaries that capture the essence of the original text while enhancing linguistic fluency and precision [7]. Semantic similarities are leveraged to extract key information from texts, focusing on shared concepts, themes, and contextual relationships to produce coherent and relevant summaries [8]. The use of advanced Natural Language Processing (NLP) techniques aids in analyzing and prioritizing key concepts, ensuring that the generated summaries reflect a deeper understanding of the text's semantics [9]. This process is crucial in an era of information overload, helping users quickly grasp key points from large documents.

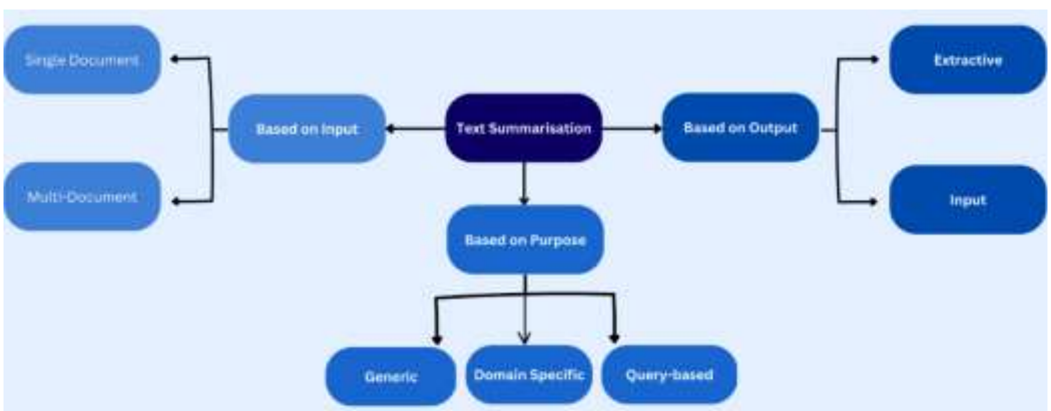

Figure: 1.2 Text Summarization

TS have wide-ranging applications, from summarizing news articles and scientific papers to creating brief overviews of lengthy legal documents. The field has evolved significantly, moving from simple extractive methods to sophisticated abstractive techniques powered by deep learning. Additionally, experiments utilizing structural properties of sentences, term expansion using WordNet, and a local thesaurus have shown promising results in selecting appropriate extractive summaries [10].Literature reviews have highlighted the importance of accuracy in text summarization, especially in scientific documents, leading to the exploration of different methods and techniques to enhance summarization processes [11]. TS offers several key benefits, including the ability to condense large volumes of text into shorter versions while retaining essential information, thus saving time and effort for readers [12] as in fig. 1.2. This process is crucial in various domains such as news articles, scientific papers, legal documents, and social media, where extracting important insights efficiently is paramount [13]. By utilizing both extractive and abstractive summarization techniques, text summarization can provide accurate and concise summaries that capture the essence of the original text without diluting its main theme, enhancing readability and comprehension for users [14]. Additionally, text summarization aids in tasks like financial research, search engine optimization, media monitoring, and question-answering bots, demonstrating its versatility and applicability across different fields [15]. Overall, text summarization plays a vital role in improving information retrieval, enhancing productivity, and facilitating quick understanding of complex textual data. The process is crucial for various applications such as news articles, scientific papers, legal documents, and social media, enabling effective analysis and decision-making. Overall, text summarization systems based on semantic similarities offer valuable tools for enhancing efficiency in information processing and decision-making across diverse domains. As we delve deeper into this study, we'll explore the various approaches, challenges, and future directions of text summarization, highlighting its importance in managing and disseminating information effectively in our digital age. This paper aims to provide an in-depth review of text summarization techniques, focusing on their evolution, current state, and future trends.

2. Approaches to Text Summarization

Text summarization is the process of condensing a text document into a briefer version while retaining its key information and meaning. It has a wide range of applications in various domains. Extractive and abstractive summarizations are the two main approaches (in fig. 2.1) to text summarization [16]. Another approach leverages the synergy of BERT for extractive summarization and GPT for abstractive summarization, resulting in a hybrid system that produces high-quality summaries across diverse domains [17].

Figure: 2.1 Approaches of Text Summarizations

2.1. Extractive Summarization

Extractive summarization involves selecting significant sentences, phrases, or sections from the original text and concatenating them to form a summary [18]. This method relies on identifying key components of the text without generating new content. Extractive summarization indeed entails selecting important sentences or phrases from the source text to create a summary [19] as in fig. 2.2. Various methods have been proposed to enhance the effectiveness of extractive summarization, such as employing linguistic features and machine learning techniques like maximum likelihood estimation and maximum entropy for generating summaries in low-resource languages like Hindi [20]. Additionally, the use of transformer models and deep learning techniques has been explored to extract meaningful information from clinical texts, aiming to improve information retrieval in medical literature [21].

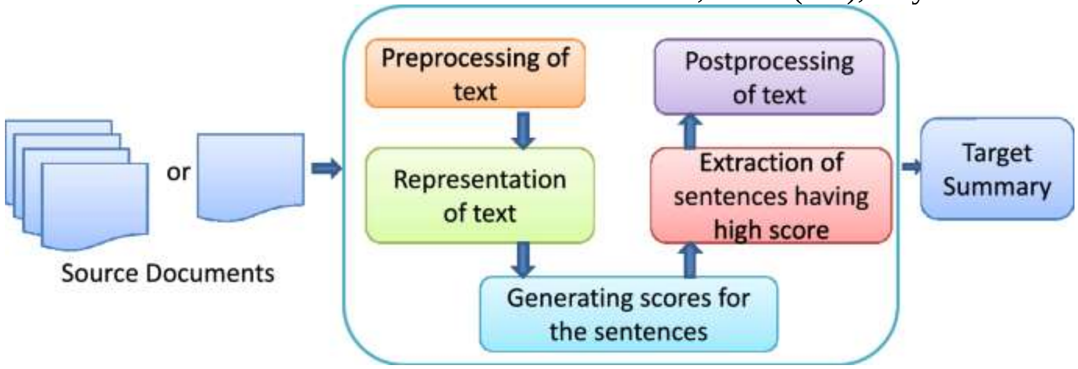

Figure: 2.2. Extractive Text Summarization

Furthermore, studies have shown that extractive summarization is commonly utilized in legal document summarization due to its ability to retain critical aspects and produce well-structured summaries that cover all legal elements, with models like C4.5 being identified as effective for this purpose [22]. Extractive summarization plays a crucial role in condensing lengthy texts while preserving essential information.

2.1.1. Statistical Methods

Statistical methods play a crucial role in extractive summarization, where the goal is to condense lengthy texts by selecting important sentences or phrases. Various approaches have been proposed, such as weight assignment of keywords in local and global files, sentence ranking algorithms [23], and the use of Elementary Discourse Units (EDUs) for extractive unit selection [24]. Furthermore, recent advancements in the field have introduced new statistical approaches that leverage hidden clustering structures within the text to improve the accuracy of extractive summarization, surpassing both extractive and abstractive methods in terms of ROUGE-2 metric by 10% [25]. Statistical methods are integral in extractive summarization, aiding in the selection of critical information for creating concise and informative summaries. These statistical techniques, when applied in educational settings, can help in summarizing educational materials effectively and efficiently, aiding students and educators in comprehending and retaining key information from various sources.

- **Term Frequency-Inverse Document Frequency (TF-IDF)**: TF-IDF (Term Frequency-Inverse Document Frequency) plays a crucial role in extractive summarization by ranking the importance of words in a document based on their frequency and uniqueness across a set of documents [24]. The application of TF-IDF in extractive summarization enhances the efficiency of summarization techniques, streamlining the process of condensing large volumes of text into coherent and informative summaries [25].
- **LexRank and TextRank**: TextRank and LexRank algorithms aid in extractive summarization by automatically identifying key sentences in a document to create concise summaries while preserving essential information through different approaches [26]. TextRank is a graph-based algorithm that assigns importance scores to sentences based on their relationships within the text, allowing for the extraction of key sentences [27]. On the other hand, LexRank employs a similar approach but incorporates the concept of eigenvector centrality to determine sentence importance [28].
- **Cluster Based Methods**: Cluster-based methods can indeed be utilized to identify key sentences in multi-document summarization tasks. By merging Cluster-Based and Graph-Based methods key information can be extracted while minimizing redundancy [29]. Within each cluster, the TextRank algorithm ranks sentences by importance and representativeness, contributing to the identification of key sentences [30]. Additionally, in the context of Document Summarization (DS), a novel pretraining objective is introduced that selects the ROUGE-based centroid of each document

cluster as a summary proxy, showcasing the effectiveness of cluster-based approaches in summarization tasks [31].

2.2. Abstractive Summarization

Abstractive summarization involves generating new sentences that convey the core ideas of the original text as in fig. 2.3. This approach mimics human-like summarization by paraphrasing and rephrasing content [32]. Abstractive text summarization in natural language processing and information retrieval offers various applications such as condensing news articles, scientific papers, legal documents, and social media content into concise summaries while preserving key information and generating new sentences to capture the essence of the original text [33].

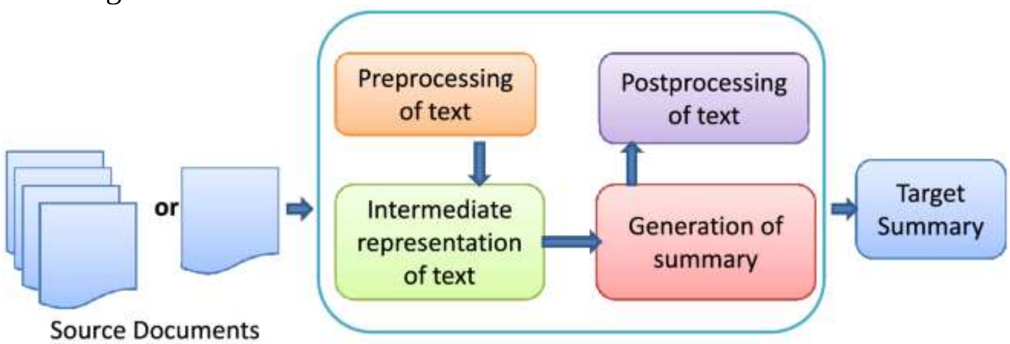

Figure: 2.3. Abstractive Summarization

2.2.1. Rule-Based Methods

Rule-based abstractive summarization involves creating summaries by following predefined linguistic rules and patterns. While traditional methods rely on rule-based strategies, recent advancements explore innovative approaches like Discriminative Adversarial Search (DAS) to alleviate exposure bias without external metrics [34]. By combining syntactic text simplification, concept frequency scoring, and semantic transformation rules to improve summary informativeness while maintaining linguistic quality, outperforming existing abstractive methods [35]. Rule-based approaches evolutes in to more sophisticated techniques that enhance the effectiveness and quality of abstractive summarization in various domains, including legal texts and general document summarization.

- **Template-based**: Template-based summarization involves structuring summaries following predefined formats, as seen in the creation of gold-standard opinion summaries from tweets [36]. Use predefined templates to generate summaries.
- **Ontologies and Semantic Networks**: On the other hand, ontology and semantic network-based methods focus on leveraging semantic relationships between terms to generate concise summaries, as demonstrated in the graph-based abstractive summarization model using SciBERT and the graph transformer network (GTN) for scientific articles [37]. Utilize structured knowledge representations to generate summaries.
- **Sequence-to-Sequence (Seq2Seq) Models**: Seq2Seq models have gained popularity for their ability to generate concise summaries by mapping input sequences to output sequences [38]. Comprise encoder-decoder architecture for summary generation. These models integrate neural-based techniques, such as LSTM units and attention mechanisms, with knowledge-based methods like Word Sense Disambiguation (WSD) to enhance content generalization and coherence [39].

3. Advances in Neural Network-Based Summarization

The advent of deep learning and neural networks has revolutionized text summarization. Neural networks significantly enhance text summarization techniques by leveraging advanced architectures and methodologies that improve accuracy, coherence, and efficiency. Innovative approaches like the Modified

Generative Adversarial Network (MGAN) utilize a three-phase process that combines extractive and generative techniques, achieving notable accuracy improvements over traditional models [40]. Furthermore, the introduction of gated attention mechanisms in graph neural networks enhances the extraction of key information while minimizing redundancy, leading to more readable summaries [41]. The application of Spiking Neural Networks (SNN) also demonstrates advantages in robustness and power efficiency compared to conventional deep learning models [42]. Pre-trained transformer models, such as BERT, GPT, and T5, have set new benchmarks in summarization tasks.

3.1 Transformer Models

Transformer models, known for their attention mechanism, have shown significant advancements in various natural language processing tasks, including text summarization [43]. Leverage self-attention mechanisms for parallel processing and improved performance. Moreover, innovations such as incorporating keyword information into the Transformer framework have shown to enhance the relevance and coherence of generated summaries, addressing common issues like detachment from the main focus [41]. The pre-trained models BERT, GPT, and T5 exhibit distinct approaches to summarization, reflecting their unique architectures and design philosophies.

- **BERT (Bidirectional Encoder Representations from Transformers)**: Pre-trained on large corpora, fine-tuned for summarization. BERT, with its bidirectional encoder architecture, excels in tasks requiring deep contextual understanding, making it particularly effective for extractive summarization where key information is identified from the text [44].
- **GPT (Generative Pre-trained Transformer)**: Autoregressive model generating coherent summaries. GPT, a generative model, is optimized for text generation, allowing it to produce coherent summaries that may not strictly adhere to the original text, thus excelling in abstractive summarization [44].
- **T5 (Text-To-Text Transfer Transformer)**: Converts all NLP tasks, including summarization, into a text-to-text format. T5 adopts a text-to-text framework, simplifying various NLP tasks, including summarization, by treating them uniformly as text generation problems, which enhances its versatility across different summarization tasks [44].

4. Evaluation Metrics

The evaluation of text summarization models is complex, with various metrics demonstrating differing effectiveness. Evaluating the quality of summaries is crucial for comparing different summarization methods. ROUGE, BLEU, and METEOR are significant metrics in NLP evaluation, each serving distinct roles in assessing translation quality.

- **ROUGE (Recall-Oriented Understudy for Gisting Evaluation)**: ROUGE, primarily used for summarization, evaluates the overlap of n-grams, making it useful for assessing the quality of generated text in various contexts [45]. ROUGE has been the benchmark for summarization evaluation for nearly two decades, facilitating comparability across thousands of studies [51]. Measures overlap of n-grams, word sequences, and word pairs between the generated summary and reference summary. ROUGE Metrics further categorize into sub blocks for deeper evaluation. ROUGE-N, Measures the overlap of n-grams between the system and reference summaries [53]. For example, ROUGE-1 evaluates unigrams, and ROUGE-2 evaluates bigrams. ROUGE-L, Measures the longest common subsequence (LCS) between the system-generated and reference summaries [52]. ROUGE-W, A weighted version of ROUGE-L that assigns different weights to matches based on their positions [52]. ROUGE-S, Measures the overlap of skip-bigrams, where the bigrams may have arbitrary gaps between words [53]. For extractive summarization, ROUGE-N (as in Equ. (4.1)), and ROUGE-L (as in Equ. (4.2)) are the most appropriate and widely used metrics because they effectively measure both content coverage (through n-gram overlap) and structural similarity (through LCS).

$$ROUGE-N=\frac{S\in References\ ngramn_n\in SCount_{match}(ngramn_n)}{S\in References\ ngramn_n\in SCount(ngramn_n)}\quad \text{--- Equ. (4.1)}$$

Where, n is the length of the n-gram (e.g., 1 for ROUGE-1, 2 for ROUGE2, $Count_{match}$ $(ngramn_n)$ is the number of n-grams in both the generated and reference summaries, $Count(ngramn_n)$ is the total number of n-grams in the reference summary.

$$ROUGE-L=\frac{LCS\ (X,Y)}{Length\ of\ Reference}\quad \text{--- Equ. (4.2)}$$

Where, LCS(X,Y) is the length of the longest common subsequence between the system summary X and the reference summary Y.

- **BLEU (Bilingual Evaluation Understudy)**: BLEU was originally designed for specific tasks, and its transferability to other NLP applications is limited, which can skew performance evaluations [54]. Evaluates the precision of n-grams. BLEU, which measures lexical overlap between machine-generated translations and human references, is widely used but has limitations in capturing nuanced errors, such as entity deviations [46]. It is a metric used to evaluate the quality of machine-generated text, particularly in tasks like machine translation. BLEU compares the machine-generated text (candidate) with one or more human-written reference texts by measuring the overlap of words and phrases (n-grams). The BLEU score is computed using a combination of n-gram precision and a brevity penalty (BP) as in Equ. (4.3).

$$BLEU=BP.exp\ {}_{n=1}^{N}w_n\log p_n\quad \text{--- Equ. (4.3)}$$

Where, BP is the brevity penalty, c is the length of the candidate translation, r is the length of the reference translation, p_n is the **modified precision** for n-grams of length n.

- **METEOR (Metric for Evaluation of Translation with Explicit ORdering)**: Considers synonymy, stemming, and word order. METEOR, on the other hand, incorporates synonymy and stemming, providing a more flexible evaluation that can better aligns with human judgment [45]. METEOR is a more linguistically informed and recall-sensitive metric than BLEU (as Equ. (4.4)). Its incorporation of semantic matching and word order penalties makes it especially valuable for tasks where meaning and fluency matter, such as summarization, translation, and text generation [55]. For extractive summarization, METEOR can be useful when evaluating the quality of the extracted sentences in terms of both content and coherence.

$$METEOR=F\times(1-Penalty)\quad \text{--- Equ. (4.4)}$$

Where, F is the F1-Score, Penalty to penalize the system. METEOR score is computed in several steps

5. Challenges and Future Directions

Despite significant advancements, TS faces several challenges. Current challenges in TS include ambiguity in generated summaries, computational inefficiencies, and the need for semantic understanding. Ambiguities arise from difficulties in modeling linguistic context and representing semantic meanings, leading to varied interpretations of summaries [47]. Additionally, the computational intensity of optimization algorithms used in Automatic Text Summarization (ATS) can hinder efficiency, particularly with complex techniques that require significant processing time [48].

- **Semantic Understanding**: Semantic understanding is crucial, as effective summarization requires capturing not only syntactic structures but also the underlying meaning and context of the text. Ensuring summaries accurately capture the meaning of the source text.
- **Coherence and Cohesion**: Coherence and cohesion in generated summaries are critical for readability and user comprehension; however, many existing systems still fall short in these areas, particularly in abstractive summarization. Maintaining logical flow and consistency in summaries.
- **Evaluation Limitations**: Evaluation of summarization quality presents its own set of challenges. Developing comprehensive and reliable evaluation metrics.

Future directions also include personalized and domain-specific summarization, as well as integrating external knowledge sources to enrich the summarization process [49]. By focusing on these innovative approaches, the field can overcome existing limitations and improve the effectiveness of text summarization techniques. Future research directions include:

- **Multimodal Summarization**: Current trends in multimodal summarization emphasize the integration of diverse data type's text, images, and audio to enhance the quality and relevance of summaries. Integrating text with other media forms (images, videos) for richer summaries.
- **Domain-Specific Summarization**: Domain-specific summarization techniques are increasingly tailored to meet the unique needs of fields such as medicine, law, and finance. In the medical domain, for instance, specialized models like MEDVOC optimize vocabulary for pre-trained language models, significantly enhancing the quality of medical text summaries by addressing out-of-vocabulary issues and improving fidelity in generated summaries.
- **Interactive Summarization**: Interactive summarization allows users to guide and customize the summarization process, enhancing the relevance and quality of the output. Allowing users to guide and customize the summarization process.

6. Conclusion

TS are a pivotal area in NLP, enabling efficient information processing and retrieval. While traditional methods laid the groundwork, neural network-based models have significantly advanced the field. The TS based study highlights the evolving landscape of summarization techniques, emphasizing the effectiveness of hybrid models that combine extractive and abstractive methods. The integration of BERT for extraction and GPT for abstraction has shown promising results in producing high-quality summaries, demonstrating the potential for improved information synthesis across various domains [17]. Earlier Studies indicates that Large Language Models (LLMs) can significantly enhance evaluation processes, aligning closely with human assessments, unlike traditional automatic metrics such as ROUGE and BERTScore, which often lack consistency and reliability [50]. Furthermore, a comprehensive analysis of abstractive techniques reveals that they yield clear, cohesive, and information-rich summaries, addressing the challenges of comprehending complex texts [7]. Overall, the findings suggest that a nuanced understanding of both extraction and abstraction methods, along with robust evaluation frameworks, is crucial for advancing text summarization in natural language processing. Continued research and innovation are essential to overcome existing challenges and harness the full potential of automated summarization.

References

1. Rosamma, K, S., Nagamma, Patil. "Pioneering Frontiers in Natural Language Processing: Exploring Promising Areas and Emerging Applications." (2023). doi: 10.1109/ocit59427.2023.10430927
2. Margot, Mieskes. "A Quantitative Study of Data in the NLP community." (2017). doi: 10.18653/V1/W17-1603
3. Liping, Zhao., Waad, Alhoshan., Alessio, Ferrari., Keletso, J., Letsholo., Muideen, A., Ajagbe., Erol-Valeriu, Chioasca., Riza, Theresa, Batista-Navarro. "Natural Language Processing (NLP) for Requirements Engineering: A Systematic Mapping Study Dataset." arXiv: Software Engineering, (2020). doi: 10.17632/ZS9VKK7SB7.1
4. Prof., Priyanka, Dhumal., Surajit, Sutar., Indraneel, Surve., Mirza, Munawwar., Vishal, Nanaware. "Text Summarization Using NLP." International Journal of Advanced Research in Science, Communication and Technology, (2024). doi: 10.48175/ijarsct-18650.
5. N.Srinivas, Rao. "Text Summarization Based on Semantic Similarity." Indian Scientific Journal Of Research In Engineering And Management,(2024). doi: 10.55041/ijsrem32218

6. Kishor, Bhimraoji, Sadafale., Sandeep, Thorat. "Review on Text Summarization using Clustering and Machine Learning-Deep Learning Models." (2024). doi: 10.1109/icc-robins60238.2024.10534014
7. Dr., Nagabhushan, S, V., Jeevitha, M., Manoj, Kumar., Archana, S., Likitha, Naveen. "Abstraction Based Text Summarization." International Journal of Advanced Research in Science, Communication and Technology, (2024). doi: 10.48175/ijarsct-18089
8. Jyoti, Srivastava., A., K., Srivastava., B., Muthu, Kumar., S.P., Anandaraj. "Text summarization using modified generative adversarial network1." (2024). doi: 10.3233/jifs-236813
9. Mr., Harshal, Bharati. "Text Summarization Using NLP." International Journal for Research in Applied Science and Engineering Technology, (2024). doi: 10.22214/ijraset.2024.56564
10. Mohamed, Elhadi. "Text Summarization: Taking Advantage of Structural Syntax, Term Expansion and Process Refinement." (2024). doi: 10.9734/bpi/ratmcs/v9/7558c
11. Arief, Agus, Sukmandhani., Yulyani, Arifin., Muhammad, Zarlis., Widodo, Budiharto. "Recent Trends for Text Summarization in Scientific Documents." (2023). doi: 10.1109/icced60214.2023.10425025
12. Y., S., Kiran., Kumar, Thejas. "Efficient Text Summarization using Natural Language Processing." (2023). doi: 10.1109/icdsaai59313.2023.10452601
13. Sakshi, Jawale., Pranit, Londhe., Prajwal, Kadam., Sushrut, Jadhav., Rushikesh, Kolekar. "Automatic Text Summarization." International Journal For Science Technology And Engineering, (2023). doi: 10.22214/ijraset.2023.51815
14. Dr., Vidyagouri, B, H., BibiSadiqa, M, D. "Text Summarization Using Machine Learning Algorithm." International journal of scientific research in computer science, engineering and information technology, (2022). doi: 10.32628/cseit228421
15. Abhishek, Bhola., Jyothirmayi, Mullapudi., Sirisha, Kollipara., Thanishka, Sanaka. "Text Summarization Based On Ranking Techniques." (2022). doi: 10.1109/IC3I56241.2022.10072962
16. Y., S., Kiran., Kumar, Thejas. "Efficient Text Summarization using Natural Language Processing." (2023). doi: 10.1109/icdsaai59313.2023.10452601
17. Prof., Priyanka, Dhumal., Surajit, Sutar., Indraneel, Surve., Mirza, Munawwar., Vishal, Nanaware. "Text Summarization Using NLP." International Journal of Advanced Research in Science, Communication and Technology, (2024). doi: 10.48175/ijarsct-18650
18. Xiao, Cheng., Weiming, Lu. "Shapley Value Guided Extractive Text Summarization." (2024). doi: 10.1109/icassp48485.2024.10447115
19. Siba, Prasad, Pati., Rasmita, Rautray. "Sentence Selection for Extractive Text Summarization using TOPSIS Approach." Procedia Computer Science, (2024). doi: 10.1016/j.procs.2024.04.144
20. Pooja, Gupta., Swati, Nigam., Rajiv, Singh. "Automatic Extractive Text Summarization using Multiple Linguistic Features." (2024). doi: 10.1145/3656471
21. Chandra, Shekar, G., Nithin, Datta, D., Geetha, Sri, Abhinay, P., Krishna, Siva, Prasad, M. "Extractive Text Summarization of Clinical Text Using Deep Learning Models." (2024). doi: 10.1109/ic-etite58242.2024.10493738
22. Utkarsh, Dixit., Sonam, Gupta., Arun, Kumar, Yadav., Divakar, Yadav. "Analyzing the Impact of Extractive Summarization Techniques on Legal Text." (2024). doi: 10.1007/978-981-99-6544-1_44
23. Utkarsh, Dixit., Sonam, Gupta., Arun, Kumar, Yadav., Divakar, Yadav. "Analyzing the Impact of Extractive Summarization Techniques on Legal Text." (2024). doi: 10.1007/978-981-99-6544-1_44
24. Muthu, Virumeshwaran., R, Thirumahal. "TF-IDF Vectorization and Clustering for Extractive Text Summarization." Journal of Information Technology and Digital World, (2024). doi: 10.36548/jitdw.2024.1.008

25. Ajinkya, Gothankar., Lavish, Gupta., Niharika, Bisht., Samiksha, Nehe., Prof., Monali, Bansode. "Extractive Text and Video Summarization using TF-IDF Algorithm." International Journal For Science Technology And Engineering, (2022). doi: 10.22214/ijraset.2022.40775
26. Shreyas, Ghorpade., Ayesha, Khan., Akhilanand, Chaurasia., V., Ramgopal, Rao., Aditi, Chhabria. "A Comparative Analysis of TextRank and LexRank Algorithms Using Text Summarization." Algorithms for intelligent systems, (2024). doi: 10.1007/978-981-97-0180-3_30
27. Mahesh, Bhutada., Shirish, Gund., Dipti, Pandit., Smita, Bhagwat., Raj, Dumbre., Pratham, Ankalwar. "Comparative Analysis of TextRank with GloVe and K-Means Clustering for Extractive Summarization." (2024). doi: 10.1109/i2ct61223.2024.10544082
28. Abhishek, Kumar. "Extractive Text Summarization." (2023). doi: 10.1109/icac3n60023.2023.10541763
29. Tuba, Gökhan., Malcolm, J, Price., Mark, Lee. "Graphs in clusters: a hybrid approach to unsupervised extractive long document summarization using language models." Artificial Intelligence Review, (2024). doi: 10.1007/s10462-024-10828-w
30. Ratish, Puduppully., Parag, Jain., Nancy, Chen., Mark, Steedman. "Multi-Document Summarization with Centroid-Based Pretraining." (2023). doi: 10.18653/v1/2023.acl-short.13
31. Raji, Ramachandran., Sanju, Jayachandran., Vaishnav, Das. "A Novel Method for Text Summarization and Clustering of Documents." (2022). doi: 10.1109/GCAT55367.2022.9972037
32. Yllias, Chali., Wenzhao, Zhu. "Abstractive Text Summarization Based on Neural Fusion." Proceedings of the ... International Florida Artificial Intelligence Research Society Conference, (2024). doi: 10.32473/flairs.37.1.135561
33. Y., S., Kiran., Kumar, Thejas. "Efficient Text Summarization using Natural Language Processing." (2023). doi: 10.1109/icdsaai59313.2023.10452601
34. Thomas, Scialom., Paul-Alexis, Dray., Sylvain, Lamprier., Benjamin, Piwowarski., Jacopo, Staiano. "Discriminative Adversarial Search for Abstractive Summarization." (2020).
35. Tatiana, Vodolazova., Elena, Lloret. "The Impact of Rule-Based Text Generation on the Quality of Abstractive Summaries." (2019). doi: 10.26615/978-954-452-056-4_146
36. Iman, Munire, Bilal., Bo, Wang., Adam, Tsakalidis., Dong-Nhat, Nguyen., Rob, Procter., Maria, Liakata. "Template-based Abstractive Microblog Opinion Summarization." Transactions of the Association for Computational Linguistics, (2022). doi: 10.1162/tacl_a_00516
37. Mehtap, Ülker., Ahmet, Bedri, Özer. "Abstractive Summarization Model for Summarizing Scientific Article." IEEE Access, (2024). doi: 10.1109/access.2024.3420163
38. Aditya, Saxena., Ashutosh, Ranjan. "Improving Sequence-to-Sequence Models for Abstractive Text Summarization Using Meta Heuristic Approaches." arXiv.org, (2024). doi: 10.48550/arxiv.2403.16247
39. Nikesh, Kumar., Suresh, Kumar. "Enhancing Abstractive Text Summarisation Using Seq2Seq Models: A Context-Aware Approach." (2024). doi: 10.1109/autocom60220.2024.10486110
40. Jyoti, Srivastava., A., K., Srivastava., B., Muthu, Kumar., S.P., Anandaraj. "Text summarization using modified generative adversarial network1." (2024). doi: 10.3233/jifs-236813
41. Jingui, Huang., Wenya, Wu., Jingyi, Li., Shengchun, Wang. "Text Summarization Method Based on Gated Attention Graph Neural Network." Sensors, undefined (2023). doi: 10.3390/s23031654
42. A., X., Hao., B., Y., Li., E., Lei, Li., C., Baofeng, Li., D., Y., Qin. "Spiking Neural Network Based Text Summarization System." (2023). doi: 10.1109/ic-nidc59918.2023.10390845
43. Dhruva, Khanwelkar., Harsh, More., NULL, AUTHOR_ID., NULL, AUTHOR_ID., Juhi, Rajani. "Transformer Based Text Summary Generation for Videos." (2024). doi: 10.1109/icctac61556.2024.10581200

44. Zaki, Muhammad, Zayyanu. "Revolutionising Translation Technology: A Comparative Study of Variant Transformer Models - BERT, GPT, and T5." Computer science and engineering : an international journal, (2024). doi: 10.5121/cseij.2024.14302
45. Christophe, Ropers., David, Dale., Prangthip, Hansanti., Gabriel, Mejia, Gonzalez., Ivan, Evtimov., Corinne, Wong., Christophe, Touret., Kristina, Pereyra., Seohyun, Sonia, Kim., Cristian, Canton-Ferrer., Pierre, Andrews., Marta, R., Costa-jussà. "Towards Red Teaming in Multimodal and Multilingual Translation." arXiv.org, (2024). doi: 10.48550/arxiv.2401.16247
46. Taisia, Glushkova., C, Zerva., André, F., T., Martins. "BLEU Meets COMET: Combining Lexical and Neural Metrics Towards Robust Machine Translation Evaluation." arXiv.org, (2023). doi: 10.48550/arXiv.2305.19144.
47. Sutriawan, Sutriawan., Supriadi, Rustad., Guruh, Fajar, Shidik., Pujiono, Pujiono., Muljono, Muljono. "Review of Ambiguity Problem in Text Summarization Using Hybrid ACA and SLR." Intelligent systems with applications, undefined (2024). doi: 10.1016/j.iswa.2024.200360
48. Muhammad, Hafizul, H., Wahab., Nor, Asilah, Wati, Abdul, Hamid., S., Subramaniam., Rohaya, Latip., Mohamed, Othman. "Decomposition–based Multi-Objective Differential Evolution for Extractive Multi-Document Automatic Text Summarization." Applied Soft Computing, (2023). doi: 10.1016/j.asoc.2023.110994
49. Abha, Kaushik., Shree, Harsh, Attri., Ravi, Shankar, Jha. "Exploring Text Summarization Techniques: A Review of Current Challenges and Future Directions." (2024). doi: 10.1109/icdt61202.2024.10489243
50. H.N., Nguyen., Haihua, Chen., Lavanya, Pobbathi., Junhua, Ding. "A Comparative Study of Quality Evaluation Methods for Text Summarization." (2024). doi: 10.48550/arxiv.2407.00747
51. Grusky, M. (2023) ‘Rogue scores’, Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) [Preprint]. doi:10.18653/v1/2023.acl-long.107.
52. Zhang, M. et al. (2024) ‘Rouge-SEM: Better evaluation of summarization using Rouge combined with semantics’, Expert Systems with Applications, 237, p. 121364. doi:10.1016/j.eswa.2023.121364.
53. Kumar, S. and Solanki, A. (2023) Rouge-SS: A New Rouge variant for evaluation of text summarization [Preprint]. doi:10.22541/au.168984209.92955863/v1.
54. Kathrin, Blagec., Georg, Dorffner., Milad, Moradi., Simon, Ott., Matthias, Samwald. (2022). 2. A global analysis of metrics used for measuring performance in natural language processing. doi: 10.48550/arXiv.2204.11574
55. Silva, A. et al. (2020) ‘Meteor: Learning memory and time efficient representations from multi-modal data streams’, Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 1375–1384. doi:10.1145/3340531.3411991.