

TEXT SUMMARIZATION USING NLP

G. Durgaprasanna¹

¹Department of Computer Science & Engineering GMRIT, Rajam, Andhra Pradesh, India

ABSTRACT

In the era of information overload, efficient data processing and understanding are essential. Text summarization, a crucial Natural Language Processing (NLP) technique, addresses this need by generating concise summaries of lengthy documents while retaining the core information. This paper examines two primary methods of text summarization: extractive and abstractive. Extractive summarization selects key sentences from the original content, ensuring accuracy and grammatical correctness. Abstractive summarization, on the other hand, generates new sentences to express the main ideas, resulting in more cohesive, human-like summaries. By leveraging NLP techniques such as tokenization, stop-word removal, and Term Frequency-Inverse Document Frequency (TF-IDF), this study enhances the effectiveness of extractive methods. Additionally, recent NLP advancements are improving the coherence and readability of abstractive summarization. This exploration contributes to the development of more advanced summarization tools, aiding users in managing and comprehending vast amounts of digital information.

Keywords: Text Summarization, Fluent summary generation, Natural Language Processing, Extractive summarization, Abstractive summarization, Information extraction, Content reduction, Machine learning, Data compression.

1. INTRODUCTION

In today's digital age, the amount of information available online is growing at an unprecedented rate, making it essential to find effective ways to simplify and make sense of this vast data. Text summarization, a key method in Natural Language Processing (NLP), helps tackle this issue by automatically shortening lengthy documents into brief summaries that still capture the main points. There are two main approaches to text summarization: extractive and abstractive. Extractive summarization works by selecting and piecing together important sentences from the original text, while abstractive summarization creates new sentences that convey the core ideas of the content. This paper compares these two methods, exploring their benefits and challenges. It uses some fundamental NLP techniques like breaking down the text into smaller parts (tokenization), removing common but unimportant words (stop-word removal), and using a method called TF-IDF (Term Frequency-Inverse Document Frequency) to highlight key content. Simple methods like scoring sentences and extracting important keywords are also used to improve the selection of key sentences in extractive summarization. The study finds that extractive methods are easier to implement and ensure the summary is grammatically correct since they rely on the original text. On the other hand, abstractive methods, although more complex, produce summaries that are less repetitive and flow more naturally, much like how a human would summarize the text. The paper also discusses recent advances in NLP, such as better text analysis and representation techniques, which have improved the quality and accuracy of both summarization approaches. These insights are valuable for future research and the development of more effective summarization tools.

1. Understanding Text summarization

Text summarization is a way to make long pieces of text shorter and more meaningful. It's a part of a field called Natural Language Processing (NLP), which helps computers to understand human language. Think of it as a tool that takes big texts and turns them into shorter, useful summaries. In a world where we have lots of information, this tool is like a superhero. There are two main ways to do text summarization: extraction and abstraction. Extraction is like picking out the most important parts of the text to make a summary. Abstraction is fancier – it means rewriting some parts of the text to create a summary that makes sense, almost like a human would do. This study looks at these methods and see where they're helpful. The techniques in NLP are like the magic that makes text summarization work. One of the foundational techniques is statistical analysis, which includes methods like TF-IDF (Term Frequency-Inverse Document Frequency) that assess the importance of words in a document. It is about looking at words and figuring out which ones are the most important in a text. Graph-based algorithms are another valuable tool, aiding in the identification of key sentences and relationships within the text. That means, it is like drawing a map of the text to see which sentences are connected and which ones are super important. In recent years, some techniques, such as Recurrent Neural Networks (RNNs) and Transformers, have emerged as powerful tools for generating abstractive summaries by understanding the context and semantics of the text. These NLP techniques are essential for several reasons in the context of text summarization. First, it helps in identifying the most relevant and important information within a text, which is crucial for crafting meaningful summaries. Second, it enable the system to understand the relationships between words and sentences, ensuring that the summary maintains clarity and readability. Third, it plays a pivotal role in enhancing the efficiency of text summarization, automating what would be a time-consuming and challenging task for humans.

In this work, it will take a closer look at these NLP techniques. In this study it shows how it make text summarization possible and why they're so important.

The contribution of the work is as follows:

Extractive Summarization:

Sentence Ranking: This technique involves ranking sentences within the source text based on their importance.

Keyword Extraction: Identifying important keywords or phrases within the text and selecting sentences that contain these keywords.

Abstractive Summarization:

Natural Language Generation (NLG): This often involves using deep learning models like Recurrent Neural Networks (RNNs) or Transformer-based models like BERT and GPT.

Sentence Compression: This technique involves compressing sentences and It's achieved by removing less important words or phrases.

Topic Modeling:

Latent Dirichlet Allocation (LDA): LDA is a statistical method used to selecting the most representative sentences or passages for each identified topic.

Word Embeddings: Word embedding is a method for representing words with large vectors in a continuous vector space that is used in text summarization and natural language processing (NLP). With the use of a given corpus, this approach seeks to capture the semantic links between words. Word embeddings are essential for improving comprehension of the meaning and relationships between words in a document when it comes to text summarization.

Word2Vec and GloVe: These word embedding techniques help capture the summarization to identify sentences with similar or related content.

Evaluation Metrics: It is used to assess the quality of generated summaries, including ROUGE (Recall-Oriented Understudy for Gisting Evaluation), BLEU (Bilingual Evaluation Understudy), and METEOR.

These metrics compare the machine-generated summary to human-written reference summaries to measure their similarity and quality. The next sections will examine more sophisticated approaches, such as neural network-driven abstractive summarization and TF-IDF analysis, as we look deeper into the specifics of NLP-based text summary. These methods seek to understand the text's subtle semantic differences in addition to condensing information. We attempt to uncover the real meaning of words and phrases via the lens of NLP, allowing for the construction of concise and educational summaries that accurately capture the core of the original content.

2. LITERATURE SURVEY TABLE

Table 1: Comparison table

Sl.no	Title	year	Description	Limitations	Advantages	Performance metrics	Gaps
1	Attention-based Transformer for Assamese Abstractive Text Summarization	2024	Improve Summarization Efficiency Evaluate Model Performance Contribute to NLP Research Develop an Abstractive Summarization Model	Dataset Availability Evaluation Metrics Morphological Complexity Low-Resource Language	Practical Applications High Accuracy Low Training Loss Improved Efficiency	Training loss: .0022, Accuracy: 47.15%.	Data scarcity, Evaluation metrics

2	A probabilistic approach for extractive summarization based on clustering cum graph ranking method	2024	Develop a New Summarization Method Improve Community Detection Enhance Sentence Selection Validate Performance	Overlapping Communities Nested Communities Link-Centric Detection Scalability Issues	Improves diversity, coverage, efficiency.	Precision: 80%, Recall: 88.89%, F1-Score: 84.21%	Dynamic community detection needed
3	Abstractive Summarization Model for Summarizing Scientific Article	2024	Novel Model Proposal Semantic Integrity Performance Comparison Focus on Current Topics	Struggles with math expressions	Versatility Semantic Integrity Improved Performance	ROUGE-L: 34.96	Complex terminologies handling
4	End to End Urdu Abstractive Text Summarization With Dataset and Improvement in Evaluation Metric	2024	Prepare dataset Implement Transformer model Enhance evaluation metric	Limited resources Complex morphology Lack of existing tools	Novel evaluation metric Contextual understanding Comprehensive dataset	ROUGE-1: 25.18 CA-RoBERTa Score: 20.61	Limited Urdu research Few existing datasets Lack of standardization
5	Mining Commonality and Specificity in Multi-Document Summarization	2024	Enhance summarization quality through commonality and specificity.	Dependency on clustering quality.	Improved coverage and diversity.	Average ROUGE-1: 1.54 ROUGE-L: 1.42.	Limited to extractive methods
6	A Survey of Text Summarization: Techniques, Evaluation and Challenges	2024	Analyze summarization methods	Challenges include semantic drift and ambiguity	Improves accessibility and understanding of information	Evaluated using ROUGE and BLEU metrics	Limited domain-specific knowledge and adaptability issues
7	Text Summarization	2024	Combine BERT and GPT for	Computationally intensive;	High-quality, concise, and	BLEU: .75, ROUGE: .85,	Need for real-time

	on Using NLP		summarizati on	requires large datasets	coherent summaries	F1: .80	processing solutions
8	Abstractive Text Summarizati on for the Urdu Language: Data and Methods	202 4	Corpus development , model evaluation, research promotion.	Low-resource language, computational cost, small prior studies.	Large corpus, diverse domains, public availability.	RROUGE-1: 46.7, ROUGE-2: 24.1, ROUGE-L: 48.7.	Low research exploration.
9	Abstractive Summarizati on Model for Summarizing Scientific Article	202 4	Develop a novel graph- based abstractive summarizati on model using SciBERT and GTN for scientific articles	Struggles with mathematical expressions, figures, and tables	Preserves document integrity, handles long documents, and generates informative summaries	ROUGE-L score: 34.96	Mathematica l expressions, figures, tables
10	Neural Abstractive Summarizati on for Long Text and Multiple Tables	202 4	Create concise summaries from long text and multiple tables, introduce FINDSum dataset, propose evaluation metrics	Dataset scarcity, identifying salient information, incorporating diverse content, processing efficiency.	Large-scale dataset, comprehensi ve summarizati on methods, improved performance over baselines.	ROUGE, Number Precision (NP), Number Coverage (NC), Number Selection (NS)	Dataset scarcity, integration complexity, efficiency.
11	Exploring the Landscape of Automatic Text Summarizati on: A Comprehensi ve Survey	202 3	Provide comprehensi ve ATS overview Examine challenges and applications Analyze methodologi es and datasets	Quality variability in summaries Dependence on training data Difficulty in evaluation standardizatio n	Saves time and effort Enhances information accessibility Supports decision- making processes.	ROUGE, BLEU, METEOR.	Need for standardized evaluation metrics Limited understandin g of context

12	Knowledge-Enhanced Graph Topic Transformer for Explainable Biomedical Text Summarization	2024	Improve explainability and accuracy of biomedical text summarization using domain-specific knowledge and graph neural topic models	Semantic gap, lack of global semantic information, black-box nature.	Enhanced explainability, better topic coherence, improved performance.	ROUGE-1, ROUGE-2, ROUGE-L (specific values not provided in the context).	Explainability, domain knowledge, coherence.
13	A Survey of Text Summarization Using NLP	2024	Develop robust cloud-based summarization.	Bias in training data.	Improved summarization efficiency.	ROUGE score, human evaluation.	Interpretability of models.
14	MFMMR-BertSum: Extractive Summarization Model	2023	Improve extractive summarization using BERT	May struggle with highly abstract content	Reduces redundancy; enhances summary coherence	ROUGE-1: 42.74, ROUGE-2: 19.85	Limited handling of context-rich summaries.
15	Knowledge-Enhanced Graph Topic Transformer for Explainable Biomedical Text Summarization	2024	Improve explainability and accuracy of biomedical text summarization using domain-specific knowledge and graph neural topic models	Semantic gap, lack of global semantic information, black-box nature.	Enhanced explainability, better topic coherence, improved performance.	ROUGE-1, ROUGE-2, ROUGE-L (specific values not provided in the context).	Explainability, domain knowledge, coherence.
16	Abstractive Text Summarization Techniques using NLP	2020	Categorize summarization techniques and analyze key algorithms.	High complexity, dependence on large datasets, and potential ambiguity in generated summaries.	Improved accuracy, better semantic understanding, and reduced redundancy in summaries.	ROUGE-1: 43.33 ROUGE-2: 20.21 ROUGE-L: 40.51	Lack of data diversity, slow processing speed, and challenges in model generalization across domains.

17	A Survey of the State-of-the-Art Models in Neural Abstractive Text Summarization	2021	Review recent models, frameworks, and challenges in neural abstractive text summarization.	Out-of-vocabulary words, factual inaccuracies, computational inefficiencies.	Efficient summarization, context understanding, diverse applications.	ROUGE scores: Up to 44.79, various datasets evaluated.	Lack of multilingual models, dataset limitations, evaluation method diversity
18	Graph-based Abstractive Biomedical Text Summarization	2022	Extract concepts, generate summaries	Domain specificity, resource dependency	Improved accuracy, coherent output	ROUGE-1: 59.60%, ROUGE-2: .7861	Data diversity, computational efficiency
19	Textual Entailment for Summarization Evaluation	2023	Evaluate abstractive summaries using entailment	Traditional metrics lack semantic assessment	Improved accuracy with entailment-based evaluation	BLEU: .754882 - ROUGE-1 F-score: .22535 - FEMS (BART): .768	Need for more sophisticated models.
20	A Survey of Automatic Text Summarization	2021	To review various methods in ATS.	Current methods often lack coherence and fluency.	Enhances efficiency in processing large datasets.	Utilizes ROUGE scores for evaluating summaries.	Insufficient integration of user feedback mechanisms.
21	Extractive Summarization of Call Transcripts	2022	Develop a method to summarize call transcripts using topic modeling, sentence selection, and punctuation restoration.	Not continuous texts Long with irrelevant sentences Ill-formed sentences	Improved readability, domain-specific summarization, effective punctuation restoration.	Rouge-1: 0.45, Rouge-2: 0.22, Rouge-L: 0.40, BLEU: 0.35, Accuracy: 78.85%, F1 Score: 92%	Domain adaptation, real-time processing, multi-lingual support.
22	Knowledge-Enhanced Graph Topic Transformer for Explainable Biomedical Text	2024	Improve explainability and accuracy of biomedical text summarization using domain-specific	Semantic gap, lack of global semantic information, black-box nature.	Enhanced explainability, better topic coherence, improved performance.	ROUGE-1, ROUGE-2, ROUGE-L (specific values not provided in the context).	Explainability, domain knowledge, coherence.

	Summarization		knowledge and graph neural topic models				
23	Advancements and Challenges in Machine Learning: A Comprehensive Review	2023	Understand theory, real-world applications, and challenges in ML.	Data privacy, computational cost, model interpretability.	Broad coverage, practical insights, comprehensive review.	Accuracy, Precision, Recall, F1 Score.	Data bias, scalability, robustness.
24	Summarization of Text and Image Captioning	2022	Enhance information retrieval and summarization efficiency	Complexity in labeled data collection process.	Superior performance in precision, recall, F-score.	Precision: .53, Recall: .73F1-score: .57.	Lack of hyperparameter tuning exploration mentioned.

3. METHODOLOGIES

A methodical technique to developing a Text summarization using NLP is described in this methodology. We will explore the process of gathering data, which includes building an extensive dataset with both actual and deepfake. We will examine the preparation stages and the methods by which the Text data was ready for model training. Next, we will examine the Natural Language processing model.

Conceptual Framework:

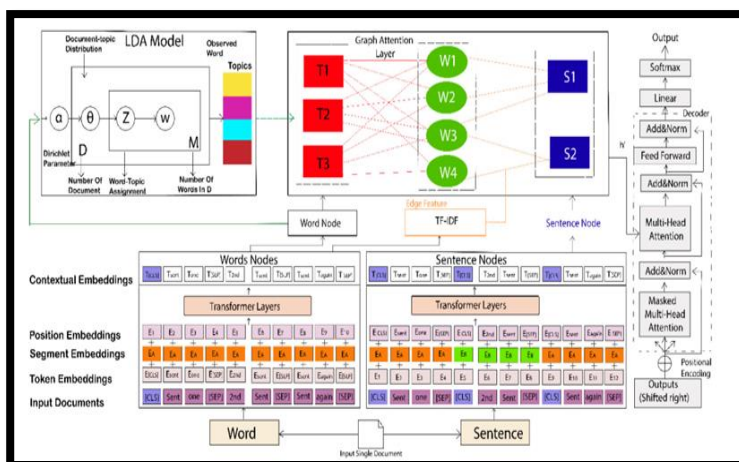


Fig 1: Architecture of proposed model

This architecture appears to illustrate a complex architecture for processing documents, combining multiple components such as Latent Dirichlet Allocation (LDA), Graph Attention Networks (GAT), and Transformers, to generate meaningful representations and output. Here's a step-by-step breakdown of each part of this architecture:

1. LDA Model (Latent Dirichlet Allocation)

- Purpose: The LDA model here is used for topic modeling, which means it's identifying themes or topics within the document.
- Components:
 - α (Dirichlet Parameter): Controls the document-topic distribution. It represents the prior distribution over topics for each document, influencing the spread of topics across documents.
 - θ (Document-Topic Distribution): For each document (D) , this distribution shows how likely each topic is within that document.
 - Z (Topic Assignment): This component indicates which topic is assigned to each word in a document.
 - W (Word-Topic Assignment): After identifying the topics, each word is assigned a topic based on the word-topic distribution.

- Topics: These are the final sets of themes derived from the document, represented by colors in the diagram.
- Output: The LDA model produces a topic distribution for each word in a document, helping identify which topics are more relevant.

2. Graph Attention Layer

- Purpose: It processes the document as a graph of nodes, with words and sentences as nodes, to capture relationships between them. The Graph Attention Network (GAT) enables focusing on important words and sentences based on their interconnections.
- Components:
 - Word Nodes (W1, W2, W3, W4): These nodes represent individual words within the document.
 - Topic Nodes (T1, T2, T3): These nodes represent the topics identified from the LDA model.
 - Sentence Nodes (S1, S2): These nodes represent sentences within the document.
- Edge Feature (TF-IDF): Term Frequency-Inverse Document Frequency (TF-IDF) is used as an edge feature between nodes, reflecting the importance of words relative to the document and across the corpus.
- Operation: The Graph Attention Layer takes the nodes and the edge features to create a graph where the model can attend to connections between words, topics, and sentences. It enables the model to dynamically weigh the connections, focusing on more informative relationships.

3. Embedding Layers

- Purpose: Embeddings provide a dense representation for each word and sentence, capturing contextual, positional, segment, and token information.
- Components:
 - Contextual Embeddings: Capture information about the context in which each word appears.
 - Position Embeddings: Encode the position of each word in the sentence, which helps the model understand the order of words.
 - Segment Embeddings: Distinguish between different parts (e.g., sentences) within the document.
 - Token Embeddings: Represent the actual word tokens in a vectorized form.
- Output: This layer produces embeddings that combine context, position, segment, and token information for each word and sentence.

4. Transformer Layers (for Words and Sentences)

- Purpose: The Transformer layers process the embeddings to produce high-level representations by applying self-attention mechanisms.
- Operation: For both word and sentence nodes, the Transformer uses self-attention to weigh the importance of each token relative to others. This allows it to create contextually enriched embeddings for both individual words and sentences.
- Output: The output from the Transformer layers are embeddings that capture the relationships and importance of words and sentences in the document.

5. Output Pathway (From Transformer Layers to Final Output)

- Masked Multi-Head Attention: Applies attention over the transformed embeddings, allowing the model to focus on various parts of the input simultaneously.
- Add & Norm: Normalizes the embeddings and adds residual connections, which help prevent information loss and maintain stability during training.
- Feed Forward Layers: Process the normalized embeddings further to add non-linearity and increase expressiveness.
- Decoder: This layer transforms the processed embeddings into the final representation suitable for the output layer.
- Linear Layer and Softmax: The final step in the output pathway, where the linear layer maps the embeddings to the desired output space, and Softmax is applied to get probability distributions for classification or generation tasks.

4. RESULTS AND DISCUSSIONS

Datasets: A wide range of datasets, each suitable to a particular topic and application case, are used to train text summarization algorithms. Datasets including news stories from a variety of sources covering a broad range of subjects are among the frequently used datasets. These datasets are crucial for training summarization models that help people remain up to date on current events by rapidly summarizing the most important aspects of news articles. Another important category consists of research articles and scientific publications from many areas. Summarization models are

trained on these kinds of datasets, enabling them to quickly and effectively extract significant information and knowledge from lengthy academic texts.

Metrics: The overlap between the n-grams (word sequences) in the generated summary and the reference (human-created) summary is measured using a set of metrics called ROUGE. By comparing machine-generated text with one or more reference texts, BLEU evaluates the text's quality. It compares the generated summary's n-gram precision to the reference summaries. When evaluating the quality of generated summaries, METEOR takes word order, precision, recall, synonymy, and splitting into consideration. This metric calculates the overlap between the generated summary and reference summary for n-grams, or word sequences of length 'n'. These standard metrics evaluate their harmonic mean (F1 score), recall (completeness of recovered information), and precision (accuracy of relevant information).

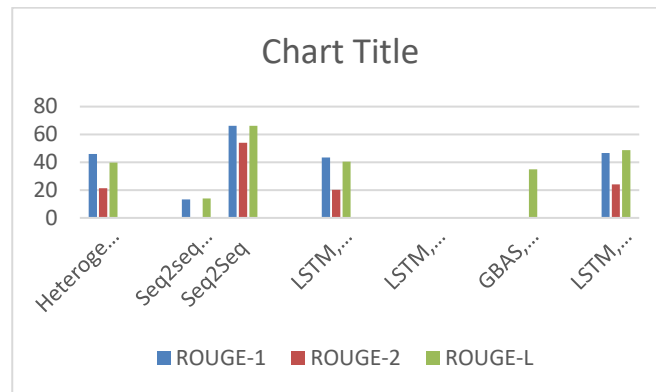


Fig 5: Result chart

5. CONCLUSION

NLP-powered text summarization represents a groundbreaking shift in how we manage and digest vast amounts of information by converting lengthy texts into coherent, concise summaries. This transformation relies primarily on two techniques: extractive and abstractive summarization, each with its own strengths and applications.

Extractive summarization works by identifying and selecting the most relevant sentences or phrases directly from the original text to form a summary. Techniques like TextRank utilize algorithms inspired by Google's PageRank, creating a network of sentences scored by relevance to ensure that key information is preserved. Frequency-based methods, such as term frequency-inverse document frequency (TF-IDF), further aid in selecting high-value content by highlighting frequently occurring terms, making extractive summarization an effective choice for technical or factual documents where preserving the original phrasing is essential. Although extractive summaries retain the style and tone of the source, they sometimes lack readability or flow, particularly when pulling sentences from different sections without rephrasing or restructuring. Abstractive summarization, on the other hand, generates new sentences that capture the main ideas of the text. This approach leverages advanced neural networks and transformer-based models such as BERT (Bidirectional Encoder Representations from Transformers) and T5 (Text-To-Text Transfer Transformer), which are designed to understand context and relationships between words through self-attention mechanisms. Abstractive methods allow these models to synthesize information into shorter, coherent phrases, creating more natural and readable summaries. This approach is particularly useful for summarizing narrative content or complex, lengthy documents where direct extraction might lead to redundancy or disjointed sentences. However, abstractive models require significant computational resources and can occasionally introduce inaccuracies or "hallucinations" by generating information that was not present in the source text.

Challenges and Future Directions: Both extractive and abstractive techniques present unique challenges. Ensuring accuracy and reliability is crucial, particularly for abstractive summarization, where misinterpretations can lead to summaries that diverge from the original meaning. Reducing bias and improving fairness in summarization models is essential, as these models learn from large datasets that may reflect real-world biases. Additionally, the demand for multilingual and multimodal summarization—capable of summarizing text in various languages or incorporating different media types like images or audio—continues to grow. Finally, research is focusing on creating user-controlled summarization, where users can customize summaries to emphasize certain aspects, providing more tailored information access. As NLP evolves, overcoming these challenges will broaden the applicability of summarization, making it more accessible, accurate, and adaptable across languages, industries, and contexts. With advancements in summarization techniques, NLP will continue transforming how individuals and organizations interact with information, making knowledge more efficient to access and understand in our information-rich world.

6. REFERENCES

- [1] Khaliq, A., Khan, A., Awan, S. A., Jan, S., Umair, M., & Zuhairi, M. F. A. (2024). Integrating Topic-Aware Heterogeneous Graph Neural Network with Transformer Model for Medical Scientific Document Abstractive Summarization. IEEE Access.
- [2] Goutom, P. J., Baruah, N., & Sonowal, P. (2024). Attention-based Transformer for Assamese Abstractive Text Summarization. Procedia Computer Science, 235, 1097-1104.
- [3] Ahmad, A., Ahmad, T., Masood, S., Siddiqui, M. K., Abd El-Rahiem, B., Plawiak, P., & Alblehai, F. (2024). A probabilistic approach for extractive summarization based on clustering cum graph ranking method. IEEE Access.
- [4] Ulker, M. and Ozer, A.B., 2024. Abstractive Summarization Model for Summarizing Scientific Article. IEEE Access.
- [5] Raza, H., & Shahzad, W. (2024). End to End Urdu Abstractive Text Summarization With Dataset and Improvement in Evaluation Metric. IEEE Access.
- [6] Ma, B. (2024). Mining both Commonality and Specificity from Multiple Documents for Multi-Document Summarization. IEEE Access.
- [7] Wibawa, A. P., & Kurniawan, F. (2024). A survey of text summarization: Techniques, evaluation and challenges. Natural Language Processing Journal, 7, 100070.
- [8] Sutar, S., Surve, I., Munawwar, M., Nanaware, V., & Dhumal, P. (2024). Text Summarization Using NLP.
- [9] Bhuyan, S. S., Mahanta, S. K., Pakray, P., & Favre, B. (2023). Textual entailment as an evaluation metric for abstractive text summarization. Natural Language Processing Journal, 4, 100028.
- [10] Ulker, M., & Ozer, A. B. (2024). Abstractive Summarization Model for Summarizing Scientific Article. IEEE Access.
- [11] Liu, S., Cao, J., Deng, Z., Zhao, W., Yang, R., Wen, Z., & Philip, S. Y. (2023). Neural abstractive summarization for long text and multiple tables. IEEE Transactions on Knowledge and Data Engineering.
- [12] Barros, T. S., Pires, C. E. S., & Nascimento, D. C. (2023). Leveraging BERT for extractive text summarization on federal police documents. Knowledge and Information Systems, 65(11), 4873-4903.
- [13] Biswas, P. K., & Jakubovich, A. (2022). Extractive summarization of call transcripts. IEEE Access, 10, 119826-119840.
- [14] Jang, H., & Kim, W. (2021). Reinforced abstractive text summarization with semantic added reward. IEEE Access, 9, 103804-103810.
- [15] Fan, J., Tian, X., Lv, C., Zhang, S., Wang, Y., & Zhang, J. (2023). Extractive social media text summarization based on MFMMR-BertSum. Array, 20, 100322.
- [16] Mahalakshmi, P., & Fatima, N. S. (2022). Summarization of text and image captioning in information retrieval using deep learning techniques. IEEE Access, 10, 18289-18297.
- [17] Batra, P., Chaudhary, S., Bhatt, K., Varshney, S., & Verma, S. (2020, August). A review: Abstractive text summarization techniques using NLP. In 2020 International Conference on Advances in Computing, Communication & Materials (ICACCM) (pp. 23-28). IEEE.
- [18] Syed, A. A., Gaol, F. L., & Matsuo, T. (2021). A survey of the state-of-the-art models in neural abstractive text summarization. IEEE Access, 9, 13248-13265.
- [19] Givchi, A., Ramezani, R., & Baraani-Dastjerdi, A. (2022). Graph-based abstractive biomedical text summarization. Journal of Biomedical Informatics, 132, 104099.
- [20] Kolambkar, V., Shingade, B., Matha, Y., Kasar, S., & Palve, P. (2024). A Survey of Text Summarization Using NLP. A Survey of Text Summarization Using NLP (March 24, 2024).
- [21] Mridha, M. F., Lima, A. A., Nur, K., Das, S. C., Hasan, M., & Kabir, M. M. (2021). A survey of automatic text summarization: Progress, process and challenges. IEEE Access, 9, 156043-156070.
- [22] Awais, M., & Nawab, R. M. A. (2024). Abstractive Text Summarization for the Urdu Language: Data and Methods. IEEE Access.
- [23] Xie, Q., Tiwari, P., & Ananiadou, S. (2023). Knowledge-enhanced graph topic transformer for explainable biomedical text summarization. IEEE journal of biomedical and health informatics.
- [24] Salam, M. A., Aldawsari, M., Gamal, M., Hamed, H. F., & Sweidan, S. (2024). MSG-ATS: Multi-level Semantic Graph for Arabic Text Summarization. IEEE Access.