Extractive Text Summarization for Malayalam News Articles

Devi B S, Dr. Rani Koshy

Department of Computer Science and Engineering, College of Engineering Trivandrum, Trivandrum, India

DOI: https://doi.org/10.51244/IJRSI.2026.1303000223

Received: 23 March 2026; Accepted: 28 March 2025; Published: 18 April 2026

ABSTRACT

The rapid increase in textual data across digital environments has made automatic text processing an essential component of Natural Language Processing (NLP). Extractive approaches involve evaluating, identifying, and selecting the most relevant sentences and are considered efficient, interpretable, and systematic alternatives to abstractive methods. Previous methods have struggled to capture meaningful semantic relationships and contextual relevance using statistical or rule-based techniques. To address these limitations, this study proposes a headline- guided extractive model that combines multilingual transformer embeddings with linguistic cues to improve relevance and information retention. The system selects sentences based on semantic similarity and syntactic importance, ensuring that the generated summaries are coherent and concise. Additionally, it reduces redundancy, thereby enhancing applicability in real-world tasks.

Index Terms—Text summarization, Malayalam news, headline- guided model, extractive summarization, transformer embed- dings.

INTRODUCTION

The extensive growth of digital information dissemination across the internet, social networks, and enterprise applications has resulted in a large volume of unstructured text that is difficult to interpret. The challenge of deriving meaningful information from large datasets has highlighted the importance of automatic text summarization within the field of Natural Language Processing (NLP) in the twenty-first century. The goal of automated text summarization is to generate a concise version of a document that preserves its essential content and semantic meaning while minimizing redundancy and irrelevant information. Text summarization techniques are broadly categorized into extractive and abstractive approaches. Extractive summarization is widely used in automated systems; however, Malayalam presents unique challenges due to its rich morphology, diverse writing styles, and limited linguistic resources. Previous work on Malayalam text summarization has largely relied on surface-level similarities. Methods such as graph-based sentence selection [1] have not effectively captured deeper contextual and semantic relationships. Although neural and transformer-based models offer improved contextual understanding and semantic representation, they typically require large datasets and substantial computational resources, which limits their applicability in low-resource settings [4]. Additionally,several approaches have been proposed, including statistical methods, frequency-based techniques, multiagent particle swarm optimization, and hierarchical self-attentive models. Despite their effectiveness, these methods rely heavily on word frequency and content-word distribution, limiting their ability to capture true semantic meaning [20]. Recent research emphasizes the importance of integrating contextual and linguistic information to enhance extractive summarization performance. Contextual modeling approaches [24] highlight that understanding inter-sentence relationships improves coherence and ensures that summaries accurately reflect the main themes of an article.

To address these challenges, this study proposes a headline-guided extractive summarization framework that utilizes transformer-based embeddings and linguistic signals to improve relevance and clarity. By leveraging the article's headline, the model selects sentences that align closely with the overall theme, ensuring concise and meaningful summaries. Additionally, heuristic-based selection mechanisms are em- ployed to reduce redundancy and enhance readability, enabling the system to handle real-world news processing tasks more effectively. The proposed framework also supports flexible sentence selection strategies,

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue III March 2026

allowing effective summarization in cases of limited or unevenly distributed information. It integrates contextual language modeling, headline-guided selection, and refinement processes to produce clear summaries of Malayalam news articles. This paper further examines the proposed framework by analyzing the impact of semantic modeling, linguistic features, and relevance mechanisms on summary quality and reliability. It also discusses key challenges, including data scarcity, mixed-language usage, and computational overhead. Finally, the study outlines future directions for developing scalable, interpretable, and intelligent summarization systems that promote broader language inclusion. This study presents a survey-driven analysis and a conceptual framework, with partial implementation to demonstrate feasibility.

Traditional Methods and Their Challenges

This section examines existing summarization systems, their methodologies, and the limitations they face in capturing semantic richness, contextual relevance, and efficiency across diverse document types.

A. Graph-Based Methods

Graph-based summarization models construct similarity graphs in which sentences are represented as nodes and edges are weighted based on lexical or semantic similarity. Early Malayalam systems employed cosine similarity and graph-theoretic centrality to identify informative sentences while maintaining coherence [1]. Subsequent studies [25] enhanced these approaches by incorporating linguistic features such as keyphrases, syntactic relations, and domain-specific ontologies, enabling the modeling of increasingly complex semantic relationships [8]. These improvements enhanced extraction quality, particularly in domains with high conceptual overlap, such as biomedical texts. However, graph-based methods still rely heavily on surface-level similarity, limiting their ability to capture deeper semantic connections. This often leads to inaccurate sentence ranking in cases involving paraphrasing, figurative language, or domain-specific abstractions. Performance may also degrade when similarity measures are noisy or sentence representations lack discriminative semantic features.

B. Statistical and Frequency-Based Methods

Statistical methods assess sentence importance using lexical frequency, sentence position, and simple scoring functions. Early Malayalam summarizers effectively utilized term-frequency signals, positional heuristics, and content word density for relevance modeling [2]. Later approaches incorporated redundancy filtering and frequency-based similarity to improve ranking stability in multi-document settings [4]. These low-cost methods are suitable for real-time and resource- constrained applications. However, they treat documents as a bag-of-words representation and fail to capture contextual meaning and complex semantic relationships. As a result, sentences containing frequent terms may be ranked highly despite lacking true importance. These approaches are therefore less effective for technical or semantically rich content.

C. TF-IDF and Keyphrase-Based Summarization Systems

TF–IDF enhances sentence ranking by emphasizing domain-relevant terms while reducing the weight of common tokens. Malayalam systems using TF–IDF have demonstrated improved performance over purely frequency-based methods [5], [16]. More advanced approaches integrate keyphrase extraction, title matching, synonym expansion, and lexical scoring to refine relevance estimation. Despite these improvements, such methods rely on surface lexical matching and are sensitive to vocabulary variation, particularly in morphologically rich or informal text. Their lack of deeper semantic understanding limits performance when key information is expressed through paraphrasing or domain-specific terminology.

D. Topic Modeling Approaches

Topic modeling approaches identify latent themes and select sentences that best represent them. LDA-based Malayalam summarizers have shown that ranking sentences based on topic probability improves coherence, especially in multitheme documents [9]. Later work combined topic distributions with clustering, similarity ranking, and redundancy reduction to enhance coverage and diversity [15]. However, topic modeling methods often suffer from instability, particularly with short or domain-specific texts. The inferred topics may not align well with sentence-level relevance, reducing summary precision. These methods also require large corpora, careful preprocessing, and parameter tuning for optimal performance.

E. Hybrid Approaches

Hybrid approaches combine statistical features, semantic similarity, graph-based ranking, and linguistic cues to improve extraction quality. Systems integrating TF–IDF, sentiment analysis, keyphrase extraction, and PageRank-based methods have shown improved accuracy and reduced noise [7]. More advanced frameworks incorporate clustering, supervised learning, and dynamic feature weighting to optimize relevance across document collections [12]. Despite their effectiveness, hybrid systems introduce significant complexity. Coordinating multiple components increases computational overhead and may propagate errors across system components. Parameter tuning is also challenging, limiting scalability and real-time applicability.

F. Evolutionary and Optimization-Based Summarization Systems

Evolutionary approaches treat summarization as a multi-objective optimization problem, balancing relevance, coverage, diversity, and redundancy. Genetic algorithms model sentence selection as chromosomes and optimize them using fitness functions based on TF–IDF, embeddings, or topic distributions [11], [22]. Techniques such as NSGA-II combined with clustering identify optimal summaries for multi-document scenarios [17]. While these methods can produce high-quality summaries, they are computationally expensive due to iterative optimization and repeated fitness evaluation. Their reliance on predefined similarity measures also limits robustness when vocabulary varies significantly across documents.

G. Neural Methods Without Transformers

Early neural summarization models introduced distributed representations capable of capturing deeper linguistic relationships than traditional methods. Malayalam systems utilized embeddings with architectures such as RBMs, enhanced with fuzzy logic and semantic similarity measures [3]. Other approaches employed Essence Vector representations to reduce noise and improve domain relevance [13]. Hierarchical RNNs and attention mechanisms further improved modeling of long-range dependencies and discourse structure [16]. However, these models require large training datasets and significant computational resources. Additionally, their lack of interpretability and limited generalization to diverse linguistic contexts remain key challenges.

H. Transformer and Foundation Models

Transformer-based models have significantly advanced summarization by enabling deep contextual understanding through self-attention mechanisms. Models such as BERT generate high-quality extractive summaries using bidirectional embeddings [10]. Foundation models, including GPT and LLaMA, extend these capabilities to abstractive summarization using large-scale pretraining and few-shot learning [18]. Despite their effectiveness, these models are computationally intensive and require substantial memory and processing power. Their outputs may also vary with input prompts and occasionally include hallucinated information. These limitations pose challenges for deployment in resource-constrained or high-stakes environments.

PROPOSED FRAMEWORK

This section presents the overall design of the proposed system and its main components. Fig. 1 illustrates how summaries are generated from Malayalam news articles using semantic similarity modeling, threshold-based extraction, and language model refinement. While certain components of the proposed framework have been partially implemented, the complete system remains under development.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue III March 2026

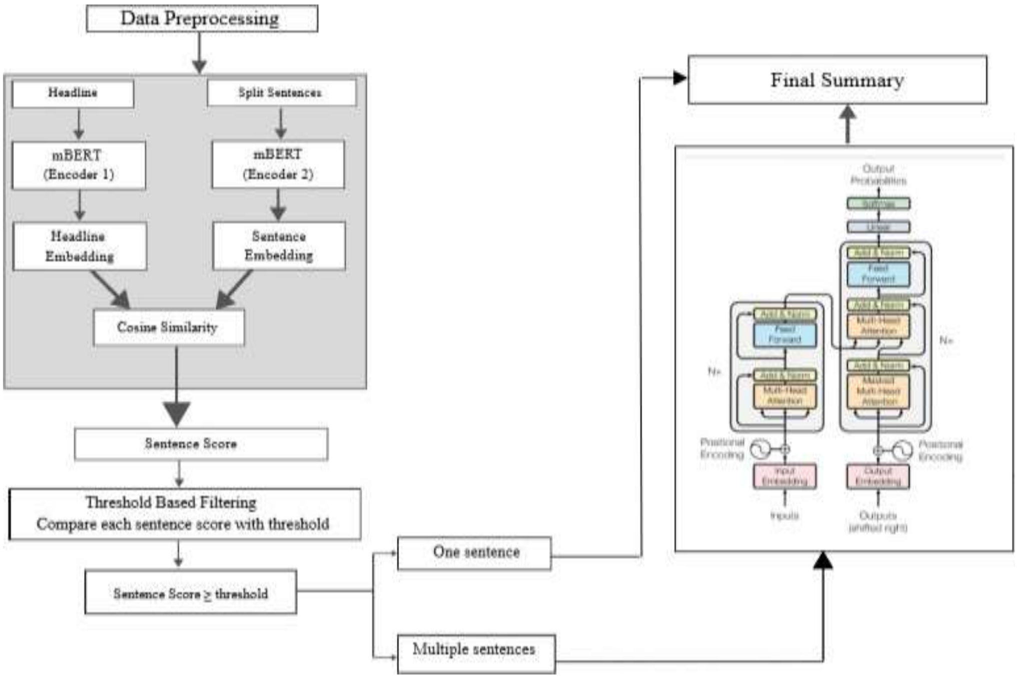

Fig. 1. Overall system design

A. Properties of Malayalam News Documents

Malayalam news articles exhibit several linguistic and structural characteristics that influence the overall process. These properties must be considered to ensure accurate information extraction.

- **Rich morphology:** Malayalam exhibits complex inflection and agglutination, resulting in multiple surface forms for a single lexical item.
- **Flexible sentence structure:** Sentence meaning is often determined by semantic roles rather than strict syntactic order.
- **Headline-body relationship:** Headlines strongly influence the central theme of news articles.
- **Contextual redundancy:** Articles frequently contain repeated or rephrased information, requiring effective handling of redundancy.
- **Domain variation:** Content spans multiple domains, such as politics, sports, and entertainment, necessitating adaptable models.

The proposed system (Fig. 1) addresses these challenges using sentence-level embeddings, similarity analysis, and language model refinement to preserve natural Malayalam phrasing.

B. Input Processing and Sentence Segmentation

The system processes Malayalam news articles containing both headlines and body text. A preprocessing pipeline normalizes the input by removing noise, including special symbols, inconsistent punctuation, and extraneous whitespace. The cleaned text is then segmented into sentences. Each sentence, along with the headline, is tokenized and converted into dense vector representations for semantic analysis. Additional pre-processing includes filtering non-Malayalam characters using Unicode-based regular expressions. Very short sentences are removed to reduce noise and improve extraction quality. This ensures that downstream components operate on consistently structured input data.

C. Headline-Guided Embedding Module

The embedding module employs the BERT-base multilingual cased model to generate contextual representations. Each sentence and headline is encoded into a 768-dimensional vector using mean pooling over the final hidden states. Semantic similarity between the headline embedding h and each sentence embedding s_i is computed using cosine similarity:

$$\frac{h \cdot s_i}{||h||.||s_i||} \quad (1)$$

This score serves as the primary criterion for ranking sentence relevance.

D. Threshold-Based Sentence Selection

Each sentence is assigned a final score based on its semantic similarity to the headline and its linguistic importance. Sentences are then selected using a predefined threshold. If no sentence exceeds the threshold, the highest scoring sentence is selected. If a single sentence exceeds the threshold, it is directly chosen. When multiple sentences meet the threshold, they are passed to a small language model for refinement. This strategy ensures that only the most relevant sentences are retained.

E. Summary Refinement Using a Small Language Model

The refinement stage employs IndicBART, a pretrained sequence-to-sequence transformer designed for Indian languages, which is fine-tuned using a learning rate of 3e-5, a batch size of 4, 10 epochs, and a maximum sequence length of 512 tokens. IndicBART processes the selected sentences to capture inter-sentence relationships and overall context. It integrates key information into a unified representation, reduces redundancy, and improves coherence by restructuring sentence flow. Instead of simply concatenating extracted sentences, the model generates a concise and fluent output aligned with the headline. Its encoder–decoder architecture effectively handles Malayalam's rich morphology and flexible syntax, resulting in improved readability and contextual accuracy.

F. User Interface and Deployment

The system includes a web-based interface that allows users to upload Malayalam news articles in plain text format. The backend processes the input, computes similarity scores, selects relevant sentences, and generates refined outputs for the user. The results are displayed through an accessible interface, enabling non-technical users to easily access these features for reading, research, and information retrieval.

COMPARATIVE ANALYSIS OF DIFFERENT METHODS

Table I compares existing extractive methods. Traditional statistical approaches are simple; however, they lack semantic depth. Machine learning, hybrid, and topic-modeling techniques improve contextual understanding and theme detection, but often compromise coherence or rely on handcrafted features. Advanced neural models, such as BERT-based CHIMA, generate high-quality summaries using contextual embeddings but are computationally expensive. These limitations highlight the need for approaches that effectively balance semantic accuracy, efficiency, and coherence, particularly for low-resource languages such as Malayalam.

TABLE I

COMPARISON OF DIFFERENT TEXT SUMMARIZATION METHODS

Paper	Year	Methodology	Result	Advantages	Disadvantages
An Extractive Malayalam Document Summarization Based on Graph Theoretic	2015	Graph-based model;cosine-similarity; sentence nodes; high-degree	Precision: 0.852; Recall: 0.891; F-measure: 0.8018	Improves accuracy; simple graph-based method; identifies key sentences	No semantic understanding; small dataset; lacks cohe- sion; relies on surface

Approach [1]		ranking		effectively	sim- ilarity
Extractive Multi-document Summarization System for Malayalam News [6]	2015	TF-IDF; WordNet semantic re-finement; cosine similarity; re-dundancy removal	F-measure: 0.33-0.75	Improved semantic relevance using WordNet; handles multi- document summarization	Surface-level features; sen-tences may sound discon- nected; simple cosine redun- dancy; manual weight choice
Malayalam Text Summarization: An Extractive Approach [2]	2016	Content-word frequency; sen-tence scoring; stopword re-moval	ROUGE-1: 0.53, ROUGE-2:0.57;	Easy implementation; works well for news; frequency-based ranking	Small dataset; no morphology; no semantic analysis; no coherence mechanism
Extractive Text Summarization Using Deep Learning [3]	2018	Restricted Boltzmann Machine (RBM); fuzzy logic; 9 hand- crafted features	Precision:0.88; Recall:0.80;F-measure: 0.84	Higher accuracy; hybrid model improves summary quality; captures more features	Manual feature engineering; limited semantics; computationally heavy; fuzzy rules fixed
Extractive Text Summarization Using Sentence Ranking [4]	2019	POS tagging; keyword frequency; sentence ranking; au- dio conversion	Generated summaries more similar to human summaries than MS Word outputs	Simple; uses POS + frequency; can generate audio summary	No semantic/context model- ing; no redundancy removal; small dataset; unclear weight formula
Efficient Text Summarization Using TF-IDF and Keyphrase Identification [5]	2021	TF-IDF; keyphrase list; title similarity; sentence position scoring	Accuracy: 95–96%; Precision:0.89–0.98; Recall: 0.79–0.89	High accuracy; effective across genres; uses multiple scoring factors	Small dataset; static word lists; purely frequency-based; manual threshold selection
A Hybrid Approach for Extractive Summarization of Medical Documents [7]	2021	Sentiment filtering (VADER); keyphrase extraction; POS + PageRank	ROUGE-1:0.82; ROUGE-2:0.74; ROUGE-L: 0.79	Neutral factual summaries; strong semantic scoring; high ROUGE	Rule-based sentiment; fixed phrase list; limited seman-tics; domain-specific
Framework for Multi- document Extractive Summary in Malayalam [8]	2021	TF-IDF + FastText + SIF em- beddings; noun-weighted Tex- tRank; MMR	ROUGE-1: 0.52, ROUGE-2:0.57;	Captures semantics; reduces redundancy; better than previ-ous Malayalam models	Small dataset; weak deep se-mantics; embedding quality sensitive; coherence issues

INTERNATIONAL JOURNAL OF RESEARCH AND SCIENTIFIC INNOVATION (IJRSI)
ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue III March 2026

Extractive Summarization Using LDA Topic Modelling [9]	2022	LDA-topic distribution;Malayalam stemmer; topic sentence scoring	ROUGE-1:0.64; ROUGE-2:0.52; ROUGE-L: 0.58	Good topic relevance; effective theme extraction	Ignores word order;shallow semantics; manually tuned parameters; some redundancy
Headline-Guided Extractive Summarization(CHIMA Model) [10]	2025	BERT-based-embeddingmodel; headline-body similarity;reranking layer	Improved recall by up to 130%; best BLEU, F1 among models tested	Strong contextual modeling; headline guidance improves sentence selection	Thai-only; surface similarity issues;logical flow sometimes inconsistent

CHALLENGES

Despite the strengths of the proposed framework, several challenges remain:

- **Language Complexity and Morphology:** Malayalam exhibits rich morphology, which can introduce variability in sentence structure and lead to errors in sentence embeddings. This complexity also poses challenges for part-of-speech (POS)-based scoring methods.
- **Model Interpretability:** Although the small language model generates fluent summaries, understanding how it selects and integrates information from multiple sentences remains challenging.
- **Redundancy and Over-Summarization:** Highly similar or redundant sentences may result in the omission of important information, leading to less informative summaries.
- **Domain-Specific Bias:** The framework is primarily designed for news articles and may not generalize effectively to other domains such as blogs or social media content.
- **Computational Resources:** Generating sentence embed- dings and running the small language model require significant computational resources, which may limit deployment on resource-constrained devices.

Fig. 2. Extractive Summarization

EXPERIMENTAL EVALUATION

A. Dataset Description

The dataset used in this work consists of approximately 89,000 Malayalam news articles, each containing a headline and body text. It spans multiple domains such as politics, sports, and entertainment, ensuring diversity in content. Each article is preprocessed and segmented into sentences for extractive summarization.

B. Current Implementation Results

Currently, only the extractive module has been implemented and evaluated using a similarity-based approach, while the remaining components are yet to be fully developed and integrated. A threshold of 0.84 was empirically determined and applied to select semantically relevant sentences. The remaining components of the proposed framework are currently under development and have not yet been evaluated experimentally.

C. Sample Output

As shown in Fig. 2, the extractive module identifies relevant sentences based on headline similarity. A comprehensive evaluation using standard metrics such as ROUGE, along with benchmarking against existing models, is part of ongoing work.

FUTURE WORK

The full implementation and integration of all modules described in the proposed framework remain incomplete and are designated as part of future work. Although research on extractive text summarization with language model refinement has advanced in the context of Malayalam news articles, several challenges and opportunities remain. Context-based summarization frameworks [24] emphasize the importance of modeling semantic relationships between sentences, demonstrating that capturing these connections can improve summary quality and coherence. Recent studies have further shown the effectiveness of enhancing traditional extractive methods with neural architectures [3], [13], resulting in more meaningful and concise summaries. Similarly, research on feature selection highlights the benefits of identifying discriminative linguistic attributes for improved sentence ranking [21]. Despite these advancements, existing methods continue to face persistent challenges related to interpretability, scalability, and factual consistency. Future research should focus on developing models that balance the interpretability of classical extractive approaches with the contextual capabilities of neural networks. Additionally, greater emphasis should be placed on model transparency, robustness, and multilingual adaptability. Further directions include optimizing models for real-time performance, enabling user-controlled customization of summary length and focus, and incorporating human-in-the-loop evaluation strategies. Moreover, adapting models for efficient deployment on edge devices and mobile platforms remains an important area for future work.

CONCLUSION

The study examined extractive, statistical, neural, and transformer-based approaches to automatic text processing for Malayalam news articles. Traditional frequency-based and graph-based methods are computationally efficient; however, they lack semantic depth and perform poorly on morphologically rich languages [1], [2]. Neural and topic-modeling approaches improve semantic representation but still face challenges related to redundancy, domain variability, and limited annotated data [3], [4].

To address these limitations, the proposed framework outlines a hybrid approach that combines similarity-driven extraction with language model refinement. This approach improves coherence, readability, and alignment with the article’s headline. Overall, the proposed framework presents a promising and adaptable direction for content condensation in low-resource languages, although full implementation and validation remain future work.

REFERENCES

1. E. B. Ajmal and R. P. Haroon, "An Extractive Malayalam Document Summarization Based on Graph Theoretic Approach," 2015 Fifth Inter- national Conference on e-Learning (econf), Manama, Bahrain, 2015, pp. 237-240, DOI: 10.1109/ECONF.2015.41
2. P. Krishnaprasad, A. Sooryanarayanan and A. Ramanujan, "Malayalam text summarization: An extractive approach," 2016 International Confer- ence on Next Generation Intelligent Systems (ICNGIS), Kottayam, India, 2016, pp. 1-4, DOI: 10.1109/ICNGIS.2016.7854008
3. N. S. Shirwandkar and S. Kulkarni, "Extractive Text Summarization Us- ing Deep Learning," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-5, DOI: 10.1109/ICCUBEA.2018.8697465
4. J. N. Madhuri and R. Ganesh Kumar, "Extractive Text Summarization Using Sentence Ranking," 2019 International Conference on Data Science and Communication (IconDSC), Bangalore, India, 2019, pp. 1-3, DOI: 10.1109/IconDSC.2019.8817040
5. R. P. Haroon, A. G. M, N. Ali and B. N. U, "An Efficient Text Summariza- tion Using Term and Inverse Frequency With Key Phrase Identification in Malayalam Language," 2021 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON- ECE), Dhaka, Bangladesh, 2021, pp. 145-148, DOI: 10.1109/WIECON-ECE54711.2021.9829671
6. Manju, K., David Peter S., and Sumam Mary Idicula, "An Extractive Multi-document Summarization System for Malayalam News Docu- ments," College of Engineering Cherthala & Cochin University of Science and Technology (CUSAT), 2015.
7. M. S. Ansary, "A Hybrid Approach for Extractive Summarization of Med- ical Documents," 2021 IEEE International Conference on Biomedical En- gineering, Computer and Information Technology for Health (BECITH- CON), Dhaka, Bangladesh, 2021, pp. 1-4, DOI: 10.1109/BECITH-CON54710.2021.9893674
8. K. Manju, D. P. S. Peter, and S. M. Idicula, "A Framework for Generating Extractive Summary from Multiple Malay- alam Documents," Inf., vol. 12, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:231733241
9. K. Manju, D. P. S. Peter, and S. M. Idicula, "Extractive summa- rization of Malayalam documents using latent Dirichlet allocation: An experience," Journal of Intelligent Systems,2022,Volume 31,url= https://api.semanticscholar.org/CorpusID:247765296
10. P. Kositcharoensuk, N. Sritrakool and P. N. Pratanwanich, "Headline- Guided Extractive Summarization for Thai News Articles," in IEEE Access, vol. 13, pp. 24368-24382, 2025, DOI: 10.1109/AC-CESS.2025.3538329
11. A´ . Herna´ndez-Castan˜eda, R. A. Garc´ıa-Herna´ndez, Y. Ledeneva and C. Milla´n-Herna´ndez, "Extractive Automatic Text Summarization Based on Lexical-Semantic Keywords," in IEEE Access, vol. 8, pp. 49896- 49907, 2020, DOI: 10.1109/ACCESS.2020.2980226
12. S. Ghodratnama, A. Beheshti, M. Zakershahrak and F. Sobhanmanesh, "Extractive Document Summarization Based on Dynamic Feature Space Mapping," in IEEE Access, vol. 8, pp. 139084-139095, 2020, DOI: 10.1109/ACCESS.2020.3012539.
13. K. -Y. Chen, S. -H. Liu, B. Chen and H. -M. Wang, "An Information Distillation Framework for Extractive Summarization," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 1, pp. 161-170, Jan. 2018, DOI: 10.1109/TASLP.2017.2764545
14. S. JUGRAN, A. KUMAR, B. S. TYAGI and V. ANAND, "Extractive Automatic Text Summarization using SpaCy in Python & NLP," 2021 International Conference on Advance Computing and Innovative Tech- nologies in Engineering (ICACITE), Greater Noida, India, 2021, pp. 582- 585, DOI: 10.1109/ICACITE51222.2021.9404712
15. A. R. Mishra, V. K. Panchal and P. Kumar, "Extractive Text Summariza- tion - An effective approach to extract information from Text," 2019 Inter- national Conference on contemporary Computing and Informatics (IC3I), Singapore, 2019, pp. 252-255, DOI: 10.1109/IC3I46837.2019.9055636
16. J. Chen and H. Zhuge, "Extractive Text-Image Summarization Using Multi-Modal RNN," 2018 14th International Conference on Semantics, Knowledge and Grids (SKG), Guangzhou, China, 2018, pp. 245-248, DOI: 10.1109/SKG.2018.00033
17. R. Alqaisi, W. Ghanem and A. Qaroush, "Extractive Multi-Document Arabic Text Summarization

Using Evolutionary Multi-Objective Opti- mization With K-Medoid Clustering," in IEEE Access, vol. 8, pp. 228206- 228224, 2020, DOI: 10.1109/ACCESS.2020.3046494

18. H. K. M, J. P and A. K. M, "Large Language Models for Indian Legal Text summarization," 2024 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2024, pp. 1-5, DOI: 10.1109/CONECCT62155.2024.10677069
19. C. Yongkiatpanich and D. Wichadakul, "Extractive Text Summarization Using Ontology and Graph-Based Method," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singa- pore, 2019, pp. 105-110, DOI: 10.1109/CCOMS.2019.8821755
20. K. Al-Sabahi, Z. Zuping and M. Nadher, "A Hierarchical Structured Self-Attentive Model for Extractive Document Summarization (HSSAS)," in IEEE Access, vol. 6, pp. 24205-24212, 2018, DOI: 10.1109/AC- CESS.2018.2829199.
21. Y. K. Meena, P. Deolia and D. Gopalani, " Optimal Features Set for Extractive Automatic Text Summarization," 2015 Fifth International Conference on Advanced Computing & Communication Technologies, Haryana, India, 2015, pp. 35-40, doi: 10.1109/ACCT.2015.123.
22. S. S. Benjumea and E. Leo´n, "Genetic Clustering Algorithm for Ex- tractive Text Summarization," 2015 IEEE Symposium Series on Compu- tational Intelligence, Cape Town, South Africa, 2015, pp. 949-956, doi: 10.1109/SSCI.2015.139.
23. Asgari, B. Masoumi and O. S. Sheijani, "Automatic text summa- rization based on multi-agent particle swarm optimization," 2014 Iranian Conference on Intelligent Systems (ICIS), Bam, Iran, 2014, pp. 1-5, doi: 10.1109/IranianCIS.2014.6802592.
24. R. Ferreira et al., "A Context Based Text Summarization System," 2014 11th IAPR International Workshop on Document Analysis Systems, Tours, France, 2014, pp. 66-70, doi: 10.1109/DAS.2014.19. keywords: Blogs;Context;Algorithm design and analysis;Abstracts;Gold;Standards;Educational institutions;Text Summarization;Text Summarization Evaluation;Document Engineering,
25. R. Ferreira et al., "A Four Dimension Graph Model for Automatic Text Summarization," 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Atlanta, GA, USA, 2013, pp. 389-396, doi: 10.1109/WI-IAT.2013.55.