IEESE International Journal of Science and Technology (IJSTE), Vol. 14 No. 3, September 2025, 14-19
ISSN : 2252-5297

Implementation of the TextRank Algorithm with WIDF for News Article Summarization

Shafira Halmahera

Department of Informatics Engineering, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang, **Indonesia**

Corresponding Author: Shafira Halmahera, e-mail: 210605110008@student.uinmalang.ac.id

Abstract. The rapid growth of information technology, especially the internet, has accelerated the spread of online news, making it increasingly difficult for readers to quickly capture the essence of articles. Automatic text summarization offers a solution by generating concise representations of lengthy documents. This study proposes an extractive summarization approach based on TextRank, enhanced with Weighted Inverse Document Frequency (WIDF) for more accurate term weighting and Latent Semantic Analysis (LSA) for dimensionality reduction. WIDF assigns weights by considering word frequency within documents and across a corpus, while LSA employs Singular Value Decomposition (SVD) to uncover deeper semantic relationships. Sentences are then ranked using TextRank and cosine similarity, with top-ranked sentences forming the summary. The method was evaluated on 3,000 news articles against human-written summaries. Results show that the best performance occurred at a 50% compression rate, achieving an average Recall of 0.6313, Precision of 0.3450, and F1-score of 0.4340, indicating the effectiveness of the proposed hybrid approach.

Key-Words: *Automatic Text Summarization; TextRank; WIDF; LSA; Extractive Summarization*

1. Introduction

In the Indonesian context, automatic summarization research is still relatively limited compared to English and other widely spoken languages. The IndoSum dataset provides a valuable benchmark for such research, enabling fair comparisons across methods [1]. This study focuses on evaluating how the integration of WIDF and LSA with TextRank performs in summarizing Indonesian news articles, highlighting trade-offs between recall, precision, and F1-score at different compression rates. The novelty of this study lies in its hybrid approach, tailored for Indonesian, where semantic richness and contextual nuances present unique challenges for summarization systems.

Recent studies have shown that integrating weighting techniques with TextRank improves performance. Weighted Inverse Document Frequency (WIDF) is a refinement of the classical IDF, assigning more discriminative power to words based on their occurrence across multiple documents [2]. Meanwhile, Latent Semantic Analysis (LSA) leverages Singular Value Decomposition (SVD) to uncover hidden semantic structures, allowing sentences to be represented in a reduced semantic space [3]. Combining these methods with TextRank is expected to enhance summarization outcomes, especially for morphologically rich languages such as Indonesian.

TextRank, introduced by Mihalcea and Tarau, has become one of the most widely used algorithms in extractive summarization due to its unsupervised and graph-based ranking nature [4]. It applies a similar principle to Google's PageRank algorithm, where sentences are treated as nodes in a graph and edges represent semantic similarity. However, its accuracy depends greatly on the representation of sentences, which in turn depends on how word importance and semantic structures are computed.

The exponential growth of online news has resulted in an overwhelming amount of information available to readers. This information overload often prevents readers from quickly identifying essential points, reducing efficiency in knowledge absorption. Automatic text summarization addresses this challenge by condensing large volumes of text into concise summaries while maintaining meaning integrity. Extractive summarization is more widely applied due to its lower complexity and higher interpretability. Various techniques for automatic summarization have been introduced, both in global studies and within Indonesian-language research [5][7].

2. Related Work

Automatic text summarization has been studied for decades, with extractive approaches often preferred due to their simplicity and efficiency. TextRank has been applied in multiple languages and domains and consistently produces concise summaries without requiring supervised learning. However, its performance strongly depends on the representation of sentence similarity [8].

Received: November 10, 2024. Revised: June 09, 2025.

Several studies attempt to enhance TextRank by incorporating weighting schemes such as TF-IDF to improve similarity matrices. Recent studies show that the choice of weighting method significantly affects sentence ranking performance, and WIDF has emerged as a promising approach because it gives more emphasis to salient terms across documents [2].

LSA is another technique used for enhancing extractive summarization. By applying SVD, LSA reduces dimensionality and uncovers latent semantic relationships among sentences, making sentence comparison more semantically meaningful [3]. Prior studies indicate that combining LSA with graph-based summarization can increase coverage of key information.

Research on Indonesian text summarization remains limited. IndoSum is currently one of the largest benchmarking datasets for summarization in Indonesian [1]. Comprehensive reviews mapping the evolution of summarization methods—from extractive to contemporary neural approaches—have been detailed in prior survey studies [9][10]. Recent advances incorporate neural and sequence-to-sequence models, including reinforcement learning and neural extractive methods [11][12]. Despite progress in neural summarization, hybrid extractive models remain relevant due to lower computational cost and interpretability, especially for low-resource languages such as Indonesian.

3. Materials and Methods

This research utilized the IndoSum dataset, a benchmark collection for Indonesian text summarization tasks, which consists of thousands of online news articles. From this dataset, 3000 articles were selected to ensure a manageable yet representative sample. Each article was paired with a human-written summary to provide ground truth references for evaluation. The diversity of topics and article lengths within IndoSum makes it a suitable dataset for testing summarization algorithms.

The overall workflow of the proposed summarization model is illustrated in Figure 1. The system is designed to process news articles through several stages, starting from preprocessing and term weighting to semantic analysis, sentence ranking, and summary generation. Each component plays a crucial role in ensuring that the final summary captures both the statistical significance and semantic coherence of the original text.

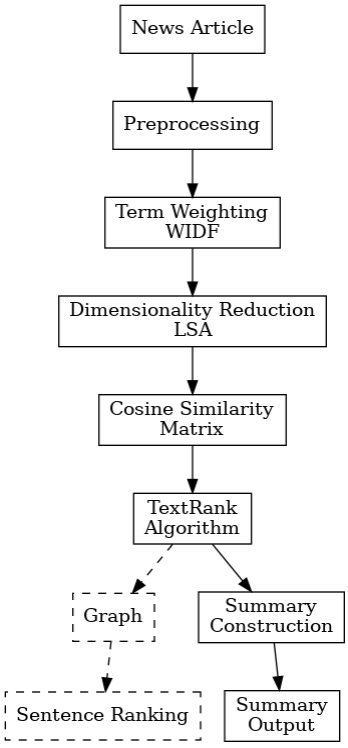

Figure 1. System Design of the Research Framework

Preprocessing

The preprocessing phase ensured that raw news articles were standardized for analysis. This included sentence tokenization, text normalization, case folding, stopword removal, and word tokenization. These steps minimized noise and established consistency in word and sentence representation, especially important for morphologically rich languages like Indonesian.

Weighted Inverse Document Frequency (WIDF)

Weighted Inverse Document Frequency assigns higher weights to words that appear frequently in a single document but less often across the corpus. The WIDF value for a word t in document d is calculated as:

$$WIDF(t,d) = tf(t,d) \times \log\left(\frac{N}{df(t)}\right) \quad (1)$$

where $tf(t,d)$is the term frequency of term $t$in document d, $df(t)$is the number of documents containing term t, and $N$is the total number of documents.

Latent Semantic Analysis (LSA)

Latent Semantic Analysis reduces the dimensionality of the WIDF matrix Ausing Singular Value Decomposition (SVD):

$$A \approx U_k \Sigma_k V_k^T \quad (2)$$

where U_k, Σ_k, and $V_k$are truncated matrices preserving the most significant latent semantic features.

Cosine Similarity

Sentence similarity is measured using cosine similarity:

$$\text{CosSim}(S_i, S_j) = \frac{\mathbf{S_i} \cdot \mathbf{S_j}}{\|\mathbf{S_i}\| \, \|\mathbf{S_j}\|} \quad (3)$$

where $\mathbf{S_i}$and $\mathbf{S_j}$are vector representations of sentences $S_i$and S_j.

TextRank Algorithm

Sentences are modeled as nodes in a graph with cosine similarity as edge weights. The TextRank score of sentence $S_i$is calculated iteratively:

$$TR(S_i) = (1-d) + d \sum_{S_j \in In(S_i)} \frac{w_{ji}}{\sum_{S_k \in Out(S_j)} w_{jk}} TR(S_j) \quad (4)$$

where $d$is the damping factor (commonly 0.85), $In(S_i)$is the set of sentences linking to S_i, and $w_{ji}$is the similarity weight.

Evaluation (ROUGE-1)

Summaries were evaluated using ROUGE-1, measuring unigram overlap between generated and reference summaries:

$$Recall = \frac{|G \cap R|}{|R|}, Precision = \frac{|G \cap R|}{|G|}, F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

where $G$is the set of unigrams in the generated summary and $R$is the set of unigrams in the reference summary.

4. Results

The evaluation of the automatic summarization system was conducted at two compression rates, namely 30% and 50%. The purpose of this experiment was to examine how different compression levels influence the balance between Recall, Precision, and F1-score.

At the lower compression rate (30%), the system produced shorter summaries that were closer in length to human-written references, reducing redundancy and maintaining compactness. In contrast, at the higher compression rate (50%), the summaries captured a larger portion of the original information but included more irrelevant sentences, leading to a decrease in Precision.

To provide a clearer illustration of these quantitative results, Figure 2 presents a comparison of the key evaluation metrics (Recall, Precision, and F1-score) at both compression levels. This visualization highlights the trade-off between information coverage (Recall) and content accuracy (Precision), with the F1-score serving as an indicator of balance between the two.

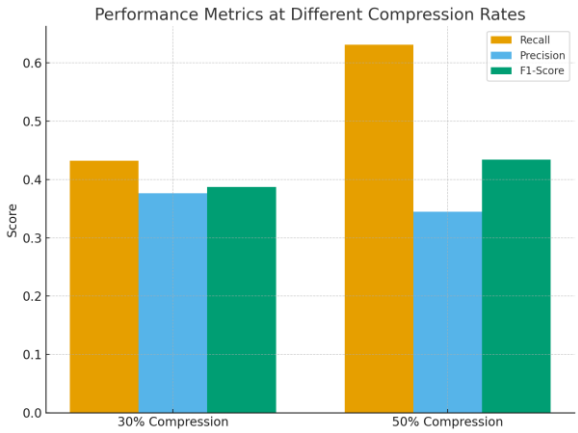

Figure 2. Recall, Precision, and F1-score at 30% and 50% compression rates.

The results of the experiments demonstrated distinct differences between the two compression rates tested. At a 30% compression rate, the generated summaries were more concise and closer in length to human-written references. This produced a balanced trade-off between Recall and Precision, with values of 0.4320 and 0.3765, respectively, and an F1-score of 0.3871. Although Recall was lower, the summaries tended to contain fewer redundant sentences and offered a level of compactness that is often desirable in practical applications.

At the 50% compression rate, system-generated summaries achieved higher Recall, reaching 0.6313, indicating that the summaries were able to capture a larger portion of the information contained in the reference summaries. However, this improvement came at the cost of Precision, which dropped to 0.3450, reflecting the inclusion of more irrelevant or less significant content. Nevertheless, the F1-score of 0.4340 was higher than that of the 30% compression rate, suggesting that the longer summaries offered a more comprehensive representation of the original text.

A comparison with baseline TextRank without WIDF and LSA revealed that the hybrid method consistently outperformed the baseline in terms of ROUGE scores. This confirms the contribution of WIDF in better weighting of important words and the role of LSA in capturing latent semantic structures. The integration of these techniques enhanced the semantic quality of the summaries, enabling the system to rank sentences more effectively based on content relevance.

In addition to quantitative evaluation, qualitative inspection of selected summaries indicated that the hybrid approach produced summaries that were more coherent and contextually meaningful. While the baseline TextRank often selected sentences that were loosely connected, the hybrid method produced summaries that preserved logical flow and conveyed the main points more effectively. This suggests that the proposed method not only improves evaluation metrics but also enhances readability and interpretability for end users.

5. Discussion

This section provides a detailed analysis of the experimental results obtained from the proposed hybrid text summarization approach. The evaluation focuses on three key metrics—Recall, Precision, and F1-

score—which serve as standard indicators of system performance in extractive summarization tasks. Table 1 presents the results at two different compression rates, offering insights into how varying the length of the generated summaries influences the balance between coverage and accuracy. By examining these results, it becomes possible to identify the inherent trade-offs between capturing more information from the source text and maintaining the relevance of the extracted sentences.

Furthermore, this discussion aims to contextualize the findings within both theoretical and practical perspectives. On the theoretical side, the results provide evidence of how feature engineering through Weighted Inverse Document Frequency (WIDF) and semantic analysis via Latent Semantic Analysis (LSA) enhance the capabilities of the baseline TextRank algorithm. On the practical side, the outcomes highlight the potential applications of the system in domains such as media and education, where concise and coherent summaries are increasingly valuable. The following subsections elaborate on these points in detail, starting with an interpretation of the results presented in Table 1.

Table 1. Summarization performance at different compression rates.

Compression Rate	Recall	Precision	F1-score
30%	0.4320	0.3765	0.3871
50%	0.6313	0.3450	0.4340

The results indicate a clear trade-off between Recall and Precision across different compression rates. The 50% compression rate produced longer summaries, allowing the system to capture a greater portion of the source content, resulting in a higher Recall score. However, this also increased the likelihood of including less relevant sentences, leading to a decrease in Precision. Conversely, the 30% compression rate generated shorter summaries with tighter sentence selection and a more balanced performance between Recall and Precision, although this configuration yielded a slightly lower F1-score. These findings suggest that determining the most suitable compression level should be aligned with the intended use case—for example, applications prioritizing completeness may prefer higher Recall, while applications prioritizing conciseness may benefit from higher Precision.

The improvement achieved by the proposed hybrid method demonstrates the effectiveness of combining WIDF and LSA in enhancing the sentence ranking process. WIDF ensures that highly informative terms receive greater weight, while LSA contributes to understanding the deeper semantic relationships among sentences. This mechanism allows the system to produce summaries that are not only statistically significant but also semantically coherent. In addition, previous findings indicate that summarization systems that incorporate semantic representation—rather than relying solely on surface-level frequency features—tend to generate summaries with stronger coherence and contextual relevance, as semantic relationships among sentences are explicitly modeled [6]. This behavior is also reflected in the proposed model, where the integration of semantic analysis enables the selection of sentences not only based on their importance but also on their contribution to the overall meaning and flow of the summary.

In addition to its academic contribution, the proposed summarization system offers practical value. In media contexts, it can assist in generating concise news digests automatically, accelerating information delivery to end users. In education and research environments, it supports document review processes by reducing long passages into essential key points. Future development may include integrating contextual embeddings from transformer-based models and exploring abstractive summarization to produce even more natural summaries that resemble human writing.

6. Conclusion

This study set out to evaluate the effectiveness of a hybrid extractive summarization approach combining TextRank with Weighted Inverse Document Frequency (WIDF) and Latent Semantic Analysis (LSA). Through experiments conducted on 3000 Indonesian news articles from the IndoSum dataset, the findings demonstrated that this hybrid method improves the quality of summaries compared to baseline TextRank. The integration of WIDF and LSA enhanced both the statistical weighting of terms and the semantic representation of sentences, yielding more informative summaries.

The results revealed that compression rates strongly influence summary quality. At a 50% compression rate, the system achieved higher Recall values, ensuring broader coverage of important content, although Precision decreased due to the inclusion of less relevant sentences. At a 30% compression rate, summaries were shorter and more selective, producing a better balance between Recall and Precision but a

slightly lower F1-score. These insights underscore the need to align summarization settings with application requirements, whether prioritizing comprehensiveness or conciseness.

Beyond quantitative performance, qualitative analysis indicated that the proposed hybrid approach produced summaries that were more coherent and contextually relevant. This finding highlights the potential of combining frequency-based weighting and semantic analysis to address challenges in summarizing Indonesian texts, where linguistic complexity poses unique obstacles. The contributions of this study lie not only in improving summarization metrics but also in offering practical insights into designing summarization systems for morphologically rich languages.

Future research can build on this work by exploring abstractive summarization approaches and integrating deep learning-based embeddings such as BERT, RoBERTa, or other transformer models. Such methods may further improve the ability of systems to capture meaning beyond surface-level representations. In addition, domain-specific datasets and cross-lingual studies could expand the applicability of summarization systems. Ultimately, this research provides a foundation for advancing automatic summarization in Indonesian, with potential benefits for media, education, and information management sectors.

References

[1] K. Kurniawan and S. Louvan, "IndoSum: A New Benchmark Dataset for Indonesian Text Summarization," arXiv preprint arXiv:1810.05334, 2018. Available: https://arxiv.org/abs/1810.05334

[2] Pratama, Septya Egho, et al. "Weighted inverse document frequency and vector space model for hadith search engine." *Indonesian Journal of Electrical Engineering and Computer Science* 18.2 (2020): 1004-1014. Available: http://doi.org/10.11591/ijeecs.v18.i2.pp1004-1014

[3] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to Latent Semantic Analysis," *Discourse Processes*, vol. 25, no. 2–3, pp. 259–284, 1998. Available: https://doi.org/10.1080/01638539809545028

[4] R. Mihalcea and P. Tarau, "TextRank: Bringing order into text," in *Proc. EMNLP*, 2004.

[5] D. Gunawan and F. Amalia, "Review of the Recent Research on Automatic Text Summarization in Bahasa Indonesia," *IOP Conf. Series: Materials Science and Engineering*, vol. 407, 2018. DOI: 10.1109/IAC.2018.8780512

[6] H. Widyassari et al., "Automatic text summarization using deep learning and semantic graph," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, 2022. Available: https://doi.org/10.1016/j.jksuci.2020.05.006

[7] K. E. Dewi and N. I. Widiastuti, "The Design of Automatic Summarization of Indonesian Texts Using a Hybrid Approach", *J. teknol. inf. pendidik.*, vol. 15, no. 1, pp. 37–43, May 2022. Available: https://doi.org/10.24036/jtip.v15i1.451

[8] J. J. Sihombing, D. Siahaan, and A. Napitupulu, "Implementation of TextRank Algorithm with TF-IDF Features for Automatic Summarization," *J. Sci. Appl. Informatics*, vol. 2, no. 2, 2024.

[9] A. Nenkova and K. McKeown, "A survey of text summarization techniques," in *Mining Text Data*, Springer, 2012.

[10] V. Gupta and G. Lehal, "A survey of text summarization extractive techniques," *J. Emerg. Technol. Web Intell.*, vol. 2, no. 3, pp. 258–268, 2010.

[11] R. Paulus, C. Xiong, and R. Socher, "A Deep Reinforced Model for Abstractive Summarization," arXiv preprint arXiv:1705.04304, 2018.

[12] Xingxing Zhang, Mirella Lapata, Furu Wei, and Ming Zhou. 2018. Neural Latent Extractive Document Summarization. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 779–784, Brussels, Belgium. Association for Computational Linguistics.