

Comparative Analysis of TextRank and LexRank for Indonesian Multi-Document News Summarization Using a Graph-Based Approach

Cut Mulya Tarisa^{1*}, Rizal Tjut Adek², Kurniawati³

^{1, 2, 3}Department of Informatics Engineering, Universitas Malikussaleh
Aceh, Indonesia.

*correspondence E-mail : cut.220170049@mhs.unimal.ac.id

ARTICLE INFO

Article History:

Submitted 07-05-2026
Received 08-06-2026
Published 09-06-2026

Keywords:

Multi-Document
Summarization;
TextRank;
LexRank;
Graph-Based Summarization;
ROUGE Evaluation;

ABSTRACT

The rapid growth of digital news content has triggered the phenomenon of information overload, making it difficult for readers to extract essential information from multiple sources. Although graph-based algorithms have been widely implemented, in-depth comparative studies analyzing the effectiveness of TextRank and LexRank on Indonesian multi-document news corpora remain relatively limited. This study aims to design and evaluate a multi-document news summarization system using a graph-based ranking approach. The methodology involves preprocessing, Term Frequency-Inverse Document Frequency (TF-IDF) weighting, cosine similarity calculation, and sentence ranking using the TextRank and LexRank algorithms. Experimental results indicate that the TextRank algorithm achieved superior performance, obtaining a ROUGE-1 score of 0.35 and a ROUGE-L score of 0.28, compared to LexRank, which yielded scores of 0.25 and 0.18, respectively. The findings also demonstrate that a damping factor of 0.85 represents the most stable parameter for balancing convergence stability and the exploration of inter-sentence relationships within the graph.

INTRODUCTION

The digital shift has catalyzed an exponential expansion of online news platforms, facilitating a continuous cycle of news production and distribution across diverse media channels [1]. This surge has introduced a substantial hurdle known as information overload, where audiences frequently struggle to distill core insights from a vast array of information sources [2]. Consequently, the development of Automated Text Summarization (ATS) has become a pivotal focus within Natural Language Processing (NLP), aimed at condensing extensive texts into concise versions while preserving their fundamental informational essence.

In the news domain, the fact that a single event is often reported by multiple sources with varying perspectives makes multi-document summarization highly relevant, yet significantly more complex due to the challenges of information redundancy [3]. Prior research has investigated the efficacy of graph-based algorithms, such as TextRank by Mihalcea and Tarau (2004) and LexRank by Erkan and Radev (2004), as effective extractive methods [4]. Nevertheless, the performance of these algorithms on Indonesian-language news corpora—particularly in multi-document scenarios involving an analysis of the damping factor

parameter—requires more rigorous investigation compared to studies focused solely on single-document summarization [5].

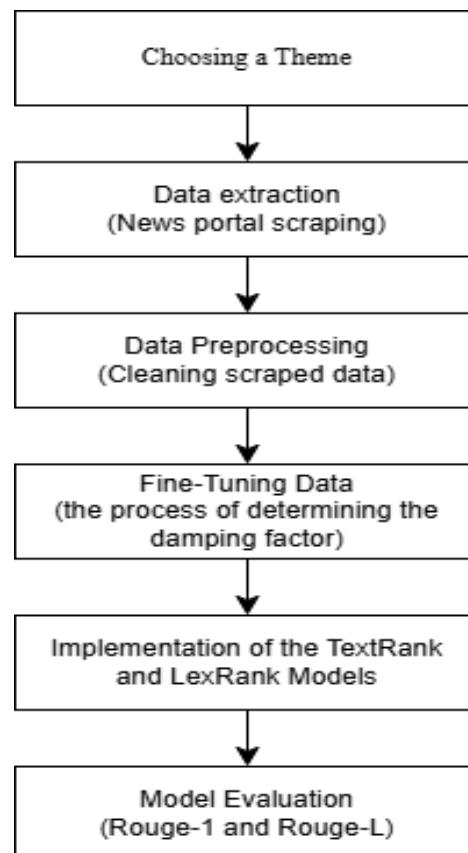
This study contributes to the field by evaluating the comparative performance of TextRank and LexRank in mitigating information redundancy within Indonesian news datasets. The focus lies in providing a granular analysis of how graph connectivity and cosine similarity weighting influence the identification of the most representative sentences across multiple source documents [6].

METHOD

This study employs a quantitative experimental methodology to evaluate the performance of graph-based summarization models. The system workflow is systematically designed to ensure data integrity from various online news portals, including Detik.com, Kompas.com, and Serambi News.

2.1. System Architecture and Preprocessing

The research process commences with data collection through web scraping, utilizing the Requests and BeautifulSoup libraries. Subsequently, a preprocessing phase is conducted to improve the quality of text representation. This stage encompasses data cleaning, lowercasing, and stopword removal to mitigate noise before the graph construction phase. Following preprocessing, each sentence is represented using the TF-IDF weighting scheme [4]. The system architecture is visualized in the following flowchart:



(Figure 1: System Architecture Diagram)

2.2. Graph Modeling and Evaluation

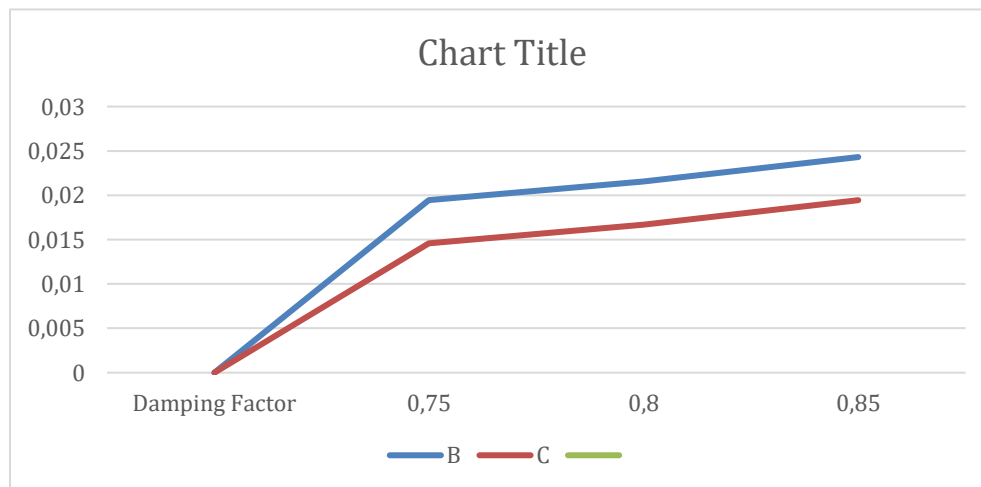
After the text has been processed, each sentence is represented as a node, and the similarity levels between sentences are mapped as edges. These edges are calculated using cosine similarity within a TF-

IDF vector space [7]. The TextRank algorithm is applied to rank sentences based on the resulting similarity graph [8], while the LexRank algorithm is utilized to determine the most central sentences within the network [9]. Ultimately, the quality of the generated summaries is assessed using ROUGE-1 and ROUGE-L metrics by comparing the system's output against human-generated summaries [3].

RESULT AND DISCUSSION

3.1. Analysis of the Damping Factor Parameter

The damping factor (D) plays a pivotal role in regulating the stability of weight calculations within the constructed graph. This parameter effectively controls the probability of transitions between nodes, thereby influencing the convergence of the ranking process. In this study, experimental testing was conducted across a range of values from 0.75 to 0.95 to identify the most effective configuration. The results indicate that a damping factor of 0.85 tends to yield the most stable and optimal evaluation scores. This specific value provides an ideal balance between the exploration of inter-sentence relationships and the convergence stability of the algorithm, leading to its selection for the multi-document news summarization system.



(Figure 2. Comparative analysis of ROUGE scores across various damping factor values, illustrating peak performance at 0.85.)

Table 1: Determination of D Value

Algorithm	Range L	Range 1	Value D	Keyword
Textrank	13,36	10,92	0,95	Local Elections
Lexrank	6,27	6,72		
Textrank	12,4	10,72	0,9	
Lexrank	8,17	6,72		
Textrank	12,4	10,72	0,85	
Lexrank	8,17	6,72		
Textrank	12,4	10,72	0,8	
Lexrank	9,93	8,23		
Textrank	12,4	10,72	0,75	
Lexrank	9,93	8,23		
Textrank	17,49	12,1	0,95	

Lexrank	16,04	11,4		
Texrank	17,49	12,1	0,9	
Lexrank	16,04	11,4		
Texrank	17,49	12,1	0,85	
Lexrank	16,52	11,5		
Texrank	17,49	12,1	0,8	
Lexrank	16,52	11,5		
Texrank	17,49	12,1	0,75	
Lexrank	16,52	11,5		
Texrank	14,96	11,44	0,95	dpr
Lexrank	14,96	11,44		
Texrank	14,96	11,44	0,9	
Lexrank	14,96	11,44		
Texrank	14,96	11,44	0,85	
Lexrank	14,96	11,44		
Texrank	14,96	11,44	0,8	
Lexrank	14,96	11,44		
Texrank	14,96	11,44	0,75	
Lexrank	14,96	11,44		
Texrank	23,77	19,22	0,95	president
Lexrank	16,84	13,15		
Texrank	23,77	19,22	0,9	
Lexrank	16,84	13,15		
Texrank	23,77	19,22	0,85	
Lexrank	16,84	13,15		
Texrank	24,13	18,74	0,8	
Texrank	13,36	10,92	0,95	Local Elections
Lexrank	6,27	6,72		
Texrank	12,4	10,72	0,9	
Lexrank	8,17	6,72		
Texrank	12,4	10,72	0,85	
Lexrank	8,17	6,72		
Texrank	12,4	10,72	0,8	
Lexrank	9,93	8,23		
Texrank	12,4	10,72	0,75	
Lexrank	9,93	8,23		
Texrank	17,49	12,1	0,95	aceh
Lexrank	16,04	11,4		
Texrank	17,49	12,1	0,9	
Lexrank	16,04	11,4		
Texrank	17,49	12,1	0,85	
Lexrank	16,52	11,5		
Texrank	17,49	12,1	0,8	
Lexrank	16,52	11,5		
Texrank	17,49	12,1	0,75	
Lexrank	16,52	11,5		
Texrank	14,96	11,44	0,95	dpr
Lexrank	14,96	11,44		

Texrank	14,96	11,44	0,9	
Lexrank	14,96	11,44		
Texrank	14,96	11,44	0,85	
Lexrank	14,96	11,44		
Texrank	14,96	11,44	0,8	
Lexrank	14,96	11,44		
Texrank	14,96	11,44	0,75	
Lexrank	14,96	11,44		
Textrank	23,77	19,22	0,95	president
Lexrank	16,84	13,15		
Texrank	23,77	19,22	0,9	
Lexrank	16,84	13,15		
Texrank	23,77	19,22	0,85	
Lexrank	16,84	13,15		
Texrank	24,13	18,74	0,8	

The selection of $\$D = 0.85\$$ indicates an ideal balance between the exploration of inter-sentence relationships within the graph and the convergence stability of the algorithm.

3.2. Performance Comparison and the Influence of Document Volume

Performance evaluation reveals significant discrepancies between the results of the two algorithms:

Table 2. Performance evaluation reveals significant discrepancies

METODE	ROUGE-1	ROUGE-L
TextRank	0,35	0,28
LexRank	0,25	0,18

Experimental results indicate that TextRank tends to be more effective than LexRank in multi-document summarization scenarios. This superiority is attributed to TextRank's weighted graph mechanism, which fully accounts for similarity values during each iteration. As the number of source documents increases, graph complexity rises due to the high levels of information redundancy across various news sources. TextRank is better equipped to reduce this redundancy by assigning lower weights to sentences with repetitive connectivity, ensuring that only the most representative information units are selected. In contrast, LexRank which is based on an unweighted graph tends to equalize relationships between sentences as long as they exceed a specific threshold, which may lead to the selection of redundant sentences in more extensive datasets.

3.3 ROUGE Score Justification and Limitations

A ROUGE-1 score of 0.35 is categorized as a competitive result for extractive multi-document summarization. This value demonstrates that the system successfully captures essential information substances that align with human standards, despite lexical differences arising from the inherently abstractive nature of human-generated summaries. The limitations of this study include the reliance on lexical metrics, which cannot yet capture deep semantic meaning, and a dataset that remains restricted to the political news domain.

3.4. Implementasi Interface Sistem

The system interface is designed to be simple, responsive, and user-friendly to support efficient interaction. It consists of key components covering the search process and result evaluation, providing an integrated experience. The following are the key components of the system:

1) Keyword Search Form

At the top of the homepage, there is a search bar for entering keywords. The search process uses an exact-match method on news headlines to avoid errors caused by partial word similarities, ensuring that a search for “Iran” does not return results like “Sindiran” and maintaining search accuracy.



Figure 3. News Search Form

2) List of Search Results

After the keyword is entered, the system displays up to five relevant news articles from 2025: all of them if there are fewer than five, the five most recent ones if there are more, and a notification if none are available. Each news article is displayed as a card containing the title, date, source, excerpt, and a link to the original article.



Figure 4. List of News Search Results

3) Summary of Results

After the news list is displayed, the user can start the summarization process by clicking the provided button. The system then generates two summary versions using TextRank and LexRank methods. The summaries are presented as coherent paragraphs, where each sentence is labeled with source identifiers such as (b1), (b2), etc., indicating the origin of each sentence from the corresponding news documents in the dataset.

Hasil Ringkasan
TextRank Partai Gema Bangsa □ yang dideklarasikan sebagai Partai Gerakan Mandiri Bangsa □ juga menyatakan arah perjuangan mereka yang bertumpu pada tiga visi besar, yakni Indonesia Mandiri, Desentralisasi Politik, dan Indonesia Reform. (b1) Partai Gema Bangsa sendiri memandang desentralisasi politik sebagai kebutuhan yang tidak bisa dihindari untuk mewujudkan tatanan demokrasi yang egaliter, terbuka, dan demokratis. (b1) Sejalan dengan itu, Rofiq dalam pidatonya juga menegaskan bahwa kemerdekaan tidak berarti menutup diri dari dunia, tetapi juga berdiri tegak dan menentukan arah bangsa dengan kekuatan sendiri. (b1) Ketua Umum Partai Gema Bangsa Ahmad Rofiq pada acara deklarasi, menekankan bahwa partai tersebut tidak dibangun untuk menjadi milik segelintir orang. (b1) Politik di Indonesia ini pengorbanan. (b3) Waktu proses: 0.269 detik
LexRank Partai Gema Bangsa □ yang dideklarasikan sebagai Partai Gerakan Mandiri Bangsa □ juga menyatakan arah perjuangan mereka yang bertumpu pada tiga visi besar, yakni Indonesia Mandiri, Desentralisasi Politik, dan Indonesia Reform. (b1) Partai Gema Bangsa sendiri memandang desentralisasi politik sebagai kebutuhan yang tidak bisa dihindari untuk mewujudkan tatanan demokrasi yang egaliter, terbuka, dan demokratis. (b1) Sejalan dengan itu, Rofiq dalam pidatonya juga menegaskan bahwa kemerdekaan tidak berarti menutup diri dari dunia, tetapi juga berdiri tegak dan menentukan arah bangsa dengan kekuatan sendiri. (b1) Ketua Umum Partai Gema Bangsa Ahmad Rofiq pada acara deklarasi, menekankan bahwa partai tersebut tidak dibangun untuk menjadi milik segelintir orang. (b1) Politik di Indonesia ini pengorbanan. (b3) Waktu proses: 0.036 detik

Figure 5. Summary Results

4) Evaluation and Recommendations for the Method

At the bottom of the summary results, the system displays a quality evaluation using the ROUGE-1 and ROUGE-L metrics in the form of a comparison table between the TextRank and LexRank methods, allowing users to analyze the differences and providing a recommendation for the best method based on the highest ROUGE-1 score to help users understand the effectiveness of each approach.



Figure 6. Evaluation and Recommendations for the Method

CONCLUSION

This study concludes that a graph-based architecture utilizing the TextRank and LexRank algorithms is effective in producing informative multi-document Indonesian news summaries. Experimental results indicate that TextRank tends to offer higher reliability, particularly when supported by an optimal damping factor of 0.85.

The academic contribution of this research lies in the establishment of optimized graph parameterization for local news corpora, which offers practical benefits by enabling audiences to process complex information flows more efficiently. Furthermore, the scientific implications of this work reaffirm that graph-based approaches remain a robust and computationally cost-effective solution for automated text summarization. Future research is encouraged to explore hybrid methodologies that integrate graph structures with pre-trained language models, such as BERT or Transformer, alongside the integration of semantic embeddings to enhance summarization quality through a deeper understanding of textual meaning.

REFERENCES

- [1] L. Chen, "Digital Transformation and Innovation of the News Media," *International Journal of Education and Humanities*, vol. 17, no. 1, pp. 154-158, 2024.
- [2] P. C. Fernández, "Effects of information overload on news consumer behavior," *International Visual Culture Review*, vol. 14, no. 1, pp. 1-11, 2023.
- [3] m. nasari and a. s. girsang, "Automated multi-document summarization using extractive-abstractive approaches," *International Journal of Informatics and Communication Technology (IJ-ICT)*, vol. 13, no. 3, pp. 400-409, 2024.
- [4] j. j. sihombing, arnita and s. iskandar, "Implementation of Text Summarization on Indonesian Scientific Articles Using TextRank Algorithm with TF-IDF Web-Based," *Journal of Soft Computing Exploration*, vol. 5, no. 4, pp. 310-319, 2024.

- [5] T. B. Kusumaa, I. M. W. and I. P. G. H. Suputra, "Implementasi LexRank dan BERT2GPT dalam Auto Summarization Teks Bahasa Indonesia," *Jurnal Nasional Teknologi Informasi dan Aplikasinya (JNATIA)*, vol. 4, no. 1, pp. 19-25, 2025.
- [6] M. Adha and m. Nasucha, "Peringkasan Teks Putusan Pengadilan Berbahasa Indonesia Menggunakan Algoritma TextRank," *Science and information technology SINTECH JOURNAL*, vol. 8, no. 1, pp. 12-27, 2025.
- [7] M. Habibi and Sumarsono, "Implementation of Cosine Similarity in an Automatic Classifier for Comments," *JISKA*, vol. 3, no. 2, pp. 110-118, 2018.
- [8] Yugandhar, R. Tezashri and Kusum, "TEXT SUMMARIZATION USING TEXTRANK ALGORITHM," *International Journal for Research Trends and Innovation (IJRTI)*, vol. 8, no. 7, pp. 323-334, 2023.
- [9] G. Erkan and D. R. Radev, "LexRank: Graph-based Lexical Centrality as Salience in Text Summarization," *Journal of Artificial Intelligence Research (JAIR)*, vol. 22, no. -, pp. 47-479, 2004.

WEBSITE / DATA SOURCES (WEB SCRAPING)

- Detik.com. (2026). *Online news portal*. <https://www.detik.com>
- Kompas.com. (2026). *Online news portal*. <https://www.kompas.com>
- Serambi News. (2026). *Online news portal*. <https://www.serambinews.com>