Enhancing multi-document summarization through topic–pattern-based sentence selection

Shaufiah [a,b,1,*], Yuefeng Li [a,2], Richi Nayak [a,3], Yutong Wu [c,4]

[a] Queensland University of Technology, 2 George Street, Brisbane QLD 4000, Australia
[b] Telkom University, Bandung, Indonesia
[c] Australian e-Health Research Center, CSIRO, Brisbane, Australia
[1] shaufiah@telkomuniversity.ac.id; [2] y2li@qut.edu.au; [3] rnayak@qut.edu.au; [4] yutong.wu@csiro.au
[*] corresponding author

ARTICLE INFO

Article history
Received December 26, 2025
Revised February 1, 2026
Accepted March 8, 2026
Available online May 31, 2026

Keywords
Multi-Document Summarization
Extractive Summarization
Topic Modeling
Pattern Mining

ABSTRACT

The growing volume of digital text content requires automated summarization systems as fundamental tools that enable users to access information at high speed. Automatic Multi-Document Summarization (MDS) requires systems to produce a summary that combines essential information from multiple documents. The extractive methods, which rely on lexical signals and sentence-based rules, yield only repetitive results because they cannot capture complex thematic relationships. This study developed an improved extractive MDS model that combines topic modeling with pattern-based semantic indicators and a method to choose diverse sentences. The model employs LDA to identify concealed thematic structures, retrieves typical word patterns to improve topic models, and selects topics via a greedy algorithm that reduces redundancy to achieve appropriate salience and coverage. The proposed system achieves better results than classical baselines on the DUC 2006 and DUC 2007 datasets, outperforming Lead, CLASSY04, KL-SUM, LexRank, TextRank, and PETMSUM. The system demonstrates superior performance over all baseline methods, achieving better results on the ROUGE-1, ROUGE-2, and ROUGE-SU4 evaluation metrics. The results show that extractive summarization tasks achieve their best performance when topic–pattern representations are combined with diversity-aware scoring methods.

© 2026 The Author(s).
This is an open access article under the CC–BY-SA license.

1. Introduction

The rapid expansion of digital text in every digital sector has created an urgent need for automated methods that are capable of organizing, filtering, and condensing information efficiently [1], [2]. Automatic Document Summarization (ADS) is the process of extracting essential information from a single or multiple documents to produce a short, concise, informative, and comprehensive summary. In recent years, ADS has been developed to address information overload. Based on the number of documents processed, ADS can be classified into two broad categories: single-document summarization (SDS) and multi-document summarization (MDS) [3]. According to its approach, ADS can also be classified as extractive and abstractive summarization. Extractive methods produce a summary by selecting the existing most important lexical and syntactic units from text, such as words, phrases, and sentences. Abstractive methods construct a summary by generating new sentences that are similar to human- written text using a complex natural language processing approach [4], [5]. The majority of automated summarization systems still use the extractive method because it is simple, widely applicable, and requires less effort for deep understanding of natural language processing.

Early studies in automatic summarization date back to Luhn's study in 1958 [6] followed by Edmunson's improvements in 1969 through the introduction of features, such as sentence position and cue phrases [7], [8]. Since then, extractive summarization has remained used due to its interpretability and computational efficiency. However, the method has also been criticized for relying heavily on shallow lexical cues and simple heuristics, which often fail to capture deeper thematic structures and introduce redundancy [3].

A later survey by [9], [10], [11] highlights two difficulties in MDS. The first requirement needs methods that eliminate duplicate information, and the second requirement needs to monitor how the main subject transforms between different document collections. The development of advanced topic–pattern-based models became necessary because current models fail to produce accurate results at their current level of performance. To address these problems, several researchers have explored modeling techniques that capture semantic signals rather than only surface-level word statistics. One example of topic modeling methods is Latent Dirichlet Allocation (LDA) in [12], [13] that uncovers latent thematic structure.

Pattern mining [14], [15] serves as a method that identifies recurring lexical patterns that point to important subjects. The present topic-driven summarization approaches lack a unified framework that manages both topic value assessment and word pattern identification and minimizes duplication [3], [9], [15], [16] The research introduces an improved extractive MDS system that draws on the PETMSUM framework. The model consists of three parts that implement topic significance, pattern-based cues, and diversity-aware sentence selection to create more detailed summaries while avoiding duplicate information [17]. The model's behavior under various sentence budget conditions becomes observable through multiple summary-length experiments.

In short, the contributions of this paper are as follows:

- The new sentence evaluation system evaluates core information by performing a thorough analysis of particular language components.
- The system selects sentences through a method that chooses information that adds novelty to the content while keeping the discussion focused on the main topic. Thus, the final summary is balanced.
- The evaluation process tested various summary length options to identify which sentence organization method generated the most effective summary results.

This paper is structured as four sections. We briefly review related work in Section 2. In section three, we describe the method in detail. The results and discussion section appears in Section 4, followed by Section 5, which contains the conclusion together with limitations and future research directions.

2. Related Work

Automatic text summarization has evolved through numerous methodologies. The earlier work began with statistical techniques, and then expanded into topic-modeling-based, graph-based, and neural summarization, both abstractive and extractive. This section discusses previous studies that provide the basis for this research and mainly focus on extractive multi-document summarization (MDS) and its challenges.

2.1. Classical and Statistical Extractive Summarization

Extractive summarization approaches extract the most essential sentences from the document(s) and provide them in a short-form summary. The importance of the content is based on several decision criteria, such as statistical measures, word frequency, title, cue words, sentence location, and sentence length. Luhn's pioneering model identified important words based on frequency statistics [6], and Edmundson built on this by aligning cue words, positional indicators, and title-based features [7]. These methods established the foundation for subsequent extractive techniques that used external heuristics

to determine sentence importance. These methods are computationally efficient, but still struggle to capture semantic relationships in large text collections and generate redundant summaries[2], [9]. These challenges have spurred the development of more advanced models capable of capturing thematic coherence across multiple documents.

2.2. Topic-Modelling-Based Summarization

Topic representation approaches create a representation of a document or set of documents and then identify important sentences based on that representation. Numerous topic modeling techniques are commonly applied to topic representation in extractive summarization. The example shows how word frequency-based term weighting produces TF*IDF (Term Frequency * Inverse Document Frequency) [18], Latent Semantic Analysis approaches [19] that apply the Singular Value Decomposition (SVD), and Bayesian models that represent the content of the document using probabilistic topic models [16], [20], [21].

The Latent Dirichlet Allocation (LDA) model allows researchers to identify concealed thematic patterns that documents share between their clusters. The main subjects in documents enable summarization systems to identify essential sentences that contain the fundamental information. [14], [22] Summarization systems frequently use these topic distributions to identify sentences most representative of major themes. Wang et al. [16] demonstrated that MDS content selection achieves better results through researcher-based sentence-level topic identification. The system achieves better results through its advanced capabilities, yet it maintains multiple essential restrictions which multiple studies [3].

Pattern mining provides an alternative approach for detecting common lexical patterns that co-occur with thematic coherence. PETMSUM [3], [17] PETMSUM [17] incorporates such patterns into topic scoring, improving the precision of sentence selection, and further [23] extended this idea by applying closed patterns to strengthen topical relevance. Despite their benefits, topic-driven models still lack solutions to data redundancy and to weak connections between pattern signals and topic importance or sentence diversity.

2.3. Graph-Based Summarization

Graph-based approaches play an important role in MDS because they explicitly model relationships among sentences, allowing them to detect redundant information and maintain content diversity [19], [24] Their popularity is largely due to strong performance in settings where documents contain a high degree of overlap [8], [25], [26], [27]. The traditional graph models depend primarily on lexical similarity to analyze text, but this approach fails to detect complex semantic and conceptual connections between words [3]. The current research focuses on two new approaches that combine graph neural networks (GNNs) with hybrid graph–transformer models to solve this problem. The models in [24], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], and GSum [41] improve discourse-level modeling through their combination of graph structures with contextual embeddings.

2.4. Neural Extractive and Abstractive Summarization

The introduction of BERT [42], and other pretrained transformer models established a new standard for document representation through their transformer architecture. The development led to BERTSUM [43], which applied these contextual embeddings to extractive summarization, achieving excellent performance on single-document summarization tasks. However, when applied to multi-document datasets, neural extractive systems do not always generalize well unless they are carefully fine-tuned. The writing styles in DUC collections become most apparent because these collections contain documents that authors wrote in different ways and documents about different subjects.

The development of abstractive models has advanced through three main architectural stages, beginning with pointer-generator networks [44] and sequence-to-sequence RNNs [45], and culminating in the introduction of PEGASUS [46]. These models generate summaries with a natural flow that appear as if written by people. The models demonstrate strong performance but need large amounts of training information to function and their decision-making processes remain unexplained. The current

limitations of extractive methods make them useful for specific applications that require stable operations, transparent systems, and flexible domain management.

2.5. Challenges in Multi-Document Summarization

The process of Multi-document summarization (MDS) presents unique difficulties which distinguish it from single-document summarization tasks. Multiple studies, including surveys [3], [9], [27] and DUC evaluation [47], [48], [49] demonstrate that systems need to address three main problems: handling duplicate information across documents, tracking changes in document topics, and handling different evaluation metrics for selection tasks. Most existing approaches tackle only parts of this problem. Graph-based models succeed in detecting duplicate content, but they do not create successful links between sentences and essential thematic elements.

The methods that focus on specific topics identify general concepts, but they fail to detect specific word choices. Improved semantic understanding in neural models comes with limitations, as they require large amounts of supervised training data and strong computing power to operate across diverse real-world environments.

2.6. Summary and Research Gap

The present summarization approaches operate independently of one another because they fail to create a unified framework that integrates topic importance, discriminative lexical pattern modeling, and explicit redundancy management into a single evaluation system. The research by PETMSUM [17] linked topic models to pattern-enhanced representations through a fundamental connection, but its redundancy management system operated with limited capabilities and failed to achieve complete integration within a unified optimization framework. The research develops an integrated framework for extractive summarization that combines three fundamental elements into one decision system. The proposed model improves content diversity by linking thematic content to penalty values that detect redundant information at each sentence-selection stage. The new system connects its elements through a core methodological advancement that exceeds the capabilities of previous topic–pattern-based methods.

3. Method

To strengthen the semantic representation used in extractive multi-document summarization, this study adopts a topic–pattern modeling approach that integrates two complementary signals: the latent themes captured by LDA and the lexical patterns that frequently appear across documents. The framework employs LDA to produce topic distributions that help identify recurring lexical patterns, enabling the detection of essential relationships among text segments. The system discovers semantic relationships between words that standard topic probability analysis fails to detect through these patterns. The system uses these patterns to assess sentence meaning with enhanced precision and understanding of the context. The model produces an expanded set of topic–pattern features through its integration of thematic cues with pattern information, which directs both candidate sentence scoring and diversity-aware selection.

3.1. Definition

Given a document set $\boldsymbol{D} = \{\boldsymbol{d_1}, \boldsymbol{d_2}, \ldots, \boldsymbol{d_n}\}$ that discusses a common theme, the aim of extractive MDS is to create a concise summary using a selected subset of sentences from $\boldsymbol{D}$. Let $\boldsymbol{S}$ be the full pool of candidate sentences, and let $\boldsymbol{S'} \subseteq \boldsymbol{S}$ denote the sentences chosen for the final summary. The objective is to balance three factors: capturing the main ideas, covering the core topics, and keeping redundancy to a minimum [6], [7], [8], [49].

3.2. Overview of the Proposed System

The proposed system builds on ideas from both traditional summarization techniques and more recent topic-oriented models. The proposed system leverages the strengths and seeks to avoid the weaknesses of MDS outlined in Section 2. An overview of the system design is shown in Fig. 1, which

can be understood through the following steps. First, preprocess and import documents into MALLET, then perform topic modeling using Latent Dirichlet Allocation (LDA) [50] to extract thematic information from the document set. Next step, do pattern mining with the FP-growth algorithm to generate topic–pattern representations. Next, score each sentence based on how well it aligns with the combined topic and pattern features. Next, redundancy reduction through similarity-based penalties applied during scoring, followed by sentence selection using a greedy strategy that balances relevance and diversity. Last, summary generation, in which the system selects a fixed number of sentences, α, to form the final output.

3.3. Topic Extraction Using LDA

LDA is applied to the document set D to estimate latent topics represented by topic–word distributions ϕ_k and document–topic proportions $\theta_{(d)}$ [22]. Each candidate sentence is mapped to a sentence–topic relevance vector by aggregating the topic probabilities of its constituent words. This results in a coarse-grained semantic representation that reflects the sentence's thematic alignment with discovered topics.

3.4. Pattern Mining and Topic-Pattern Representation

While LDA identifies broad thematic tendencies, it often fails to capture specific lexical cues that provide additional semantic detail. To address this, the FP-growth algorithm [51], is used to extract frequent lexical patterns from the document set. FP-growth is widely used in text mining due to its efficiency in identifying recurrent term combinations [52], [53]. These lexical patterns are then aligned with LDA topic–word distributions to construct topic–pattern vectors that highlight topically coherent and semantically informative associations, following the principles of pattern-enhanced topic modeling [17], [54].

The resulting sentence representation integrates both topic relevance and pattern activation, enabling more discriminative sentence scoring.

3.5. Unified Sentence Scoring Framework

The proposed model uses three elements to create informative summaries that prevent duplicate information through its single scoring system. The system evaluates sentences by combining all three signals to determine both topic salience and diversity. For each candidate sentence s, the final score is defined as:

$$Score(s) = \alpha \cdot TS(s) + \beta \cdot PR(s) - \delta \cdot Red(s, S) \quad (1)$$

where $TS(s)$measures the strength of the sentence's alignment with dominant LDA topics in the document cluster; $PR(s)$ quantifies the activation of frequent topic-coherent lexical patterns which the FP-growth mining technique identifies; and $Red(s, S)$represents the maximum cosine similarity between a sentence s and the already selected summary set S. The system distributes its attention between different data aspects through the weighting parameters α, β, δ.

This formulation ensures that sentences are not selected solely because they are central (as in graph-based models) or solely because they match dominant topics (as in topic-only models). Instead, a sentence is preferred only if it simultaneously: aligns strongly with salient topics, activates discriminative lexical patterns, and the selected content receives new information which differs from what has been presented before.

The system implements the redundancy penalty as a dynamic process which runs during the greedy selection phase. After each sentence is selected, similarity scores are recomputed for remaining candidates, and their final scores are updated using the diversity control parameter δ. The unnormalized saliency scale that TS(s)and PR computation require needs a δ value of 5 to maintain equal importance between relevance and diversity terms. The system blocks users from choosing additional options that produce similar sentences that contain strong thematic connections.

The model allows users to manage information diversity through a single parametric equation that achieves maximum coverage with minimum redundancy. The proposed framework uses a single evaluation approach that differs from conventional centrality-based systems yet provides an exact mathematical model for describing pattern-topic connections in PETMSUM-inspired models. The proposed system is illustrated in Fig. 1, and the detailed algorithm is shown in Fig. 2.

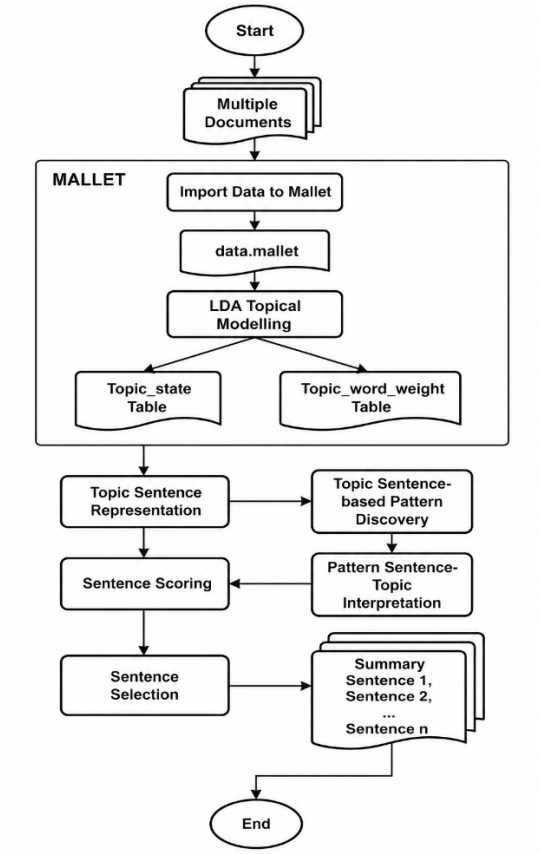

Fig. 1. Complete pipeline from preprocessing to selection of the proposed system.

Algorithm 1. PETMSUM-Unified Topic Sentence Selection

Input:
- $\{SR_v\}_{v=1}^{V}$: top-n ranked sentence sets per topic
- L : required number of sentences
- α, β, δ : weighting parameters

Output: Summary S of size L

$S \leftarrow \emptyset$
$PL \leftarrow \emptyset$
{— Candidate Pool Construction —}
for $v = 1$ to V **do**
  for each sentence s in SR_v **do**
    if $s \notin PL$ **then**
      Insert s into PL
    end if
  end for
end for
{— Initial Unified Score Computation —}
for each $s \in PL$ **do**
  Compute topic significance $TS(s)$
  Compute pattern activation $PR(s)$
  $Score(s) \leftarrow \alpha \cdot TS(s) + \beta \cdot PR(s)$
end for
$PL' \leftarrow PL$
{— Greedy Selection with Diversity Penalty —}
while $|S| < L$ **do**
  $s^* \leftarrow \arg\max_{s \in PL'} Score(s)$
  $S \leftarrow S \cup \{s^*\}$
  for each $x \in PL', x \neq s^*$ **do**
    $Score(x) \leftarrow Score(x) - \delta \cdot sim(x, s^*)$
  end for
  Remove s^* from PL'
end while
return S

Fig. 2. PETMSUM- Unified Topic Sentence Selection Algorithm.

3.6. Redundancy-Aware Sentence Selection

The selection process works step by step, beginning with the system picking the highest-scoring sentence from the candidate pool. Second, redundancy penalties are applied to the remaining sentences based on their semantic similarity. Third, the scores are updated, and the process repeats.

This continues until the summary reaches the set sentence limit. The approach ensures that the selected sentences are both topically important and sufficiently distinct from one another, following the principles of the PETMSUM framework.

3.7. Summary Length Criterion

The proposed system generates summaries with a specific number of sentences, whereas PETMSUM produces summaries of approximately 250 words. The proposed system generates summaries with a fixed number of sentences, enabling better comparison between baseline models.

3.8. Experimental Setup

3.8.1. Dataset

The proposed model was evaluated using two benchmark datasets from the Document Understanding Conference (DUC): DUC 2006 and DUC 2007 [47], [48]. These datasets are standard for multi-document and query-focused summarization research.

The DUC 2006 collection contains 50 document clusters, each comprising 25 news articles. DUC 2007 includes 45 clusters, each with 25 articles. The number of sentences per cluster ranges from 5 to 79 in DUC 2006 and from 9 to 125 in DUC 2007. Each cluster is provided with multiple human-generated reference summaries: 10 in DUC 2006 and 4 in DUC 2007. Both datasets specify a target summary length of approximately 250 words.

A brief overview of dataset characteristics is provided in Table 1.

Table 1. DUC Dataset Description.

Parameter	Dataset	
	DUC 2006	DUC 2007
Number of Documents Sets	50	45
Number of articles in each set	25	25
Min number of sentences per document set	5	9
Max number of sentences per document set	79	125
Length of Summary (in words)	250	250

3.8.2. Baseline

To provide a comprehensive comparison, the proposed model was evaluated against several established extractive summarization baselines, as shown in Table 2. These baselines represent the evolution from early rule-based and statistical approaches to modern neural methods. The proposed model is compared against all baselines using identical preprocessing and evaluation conditions.

Table 2. Comparison of Baseline Model and Proposed System.

Component / Aspect	Lead [52]	CLASSY 04 [53]	KL-SUM [20]	LexRank [24]	TextRank [25]	PETMSUM [17]	BERTSUM [44]	Proposed Model
Descriptions	Selects the first few sentences of each document. Effective for structured news but less robust for heterogeneous text.	Employs clustering and statistical-linguistic techniques to select representative sentences	Minimizes KL-divergence between the source documents and summary, preserving word-distribution similarity	Graph-based centrality methods that score sentences based on similarity graphs.	Graph-based centrality methods that score sentences based on similarity graphs.	Utilizes pattern-enhanced topic modeling for query-focused MDS, integrating topic salience, pattern activation, and redundancy reduction.	A neural extractive model using pretrained BERT encoders, often achieving strong performance for single-document summarization.	The proposed algorithm applies unified topic–pattern scoring with a greedy diversity penalty to select salient, non-redundant sentences under a predefined sentence budget in MDS.
Model Paradigm	Heuristic	Clustering + statistical	Probabilistic divergence minimization	Graph-based	Graph-based	Topic–pattern-based	Neural (Transformer-based)	Topic–pattern-based (enhanced)
Designed for Multi-Document	No	Yes	Yes	Yes	Yes	Yes	Primarily single-document	Yes

Component / Aspect	Lead [52]	CLASSY 04 [53]	KL-SUM [20]	LexRank [24]	TextRank [25]	PETMSUM [17]	BERTSUM [44]	Proposed Model
Topic Modeling Integration	No	No	No	No	No	Yes (LDA-based)	No (implicit via embeddings)	Yes (LDA with refined weighting)
Pattern-Based Lexical Modeling	No	No	No	No	No	Yes	No	Yes (integrated into scoring)
Sentence Centrality Mechanism	Positional bias	Cluster-based selection	Distribution alignment	Eigenvector centrality	PageRank-based ranking	No	Classification-based scoring	No (topic–pattern scoring)
Redundancy Handling Strategy	None	Limited (cluster filtering)	Implicit via distribution match	Similarity threshold	Similarity threshold	Greedy diversity penalty	Learned implicitly via training	Refined greedy diversity-aware penalty
Supervised Training Required	No	No	No	No	No	No	Yes	No
Computational Requirement	Low	Low–Moderate	Moderate	Moderate	Moderate	Moderate	High (GPU recommended)	Moderate
Interpretability	High	High	High	High	High	High	Limited (black-box encoder)	High
Explicit Sentence Budget Control	Fixed positional	Heuristic	Word-distribution guided	Length heuristic	Length heuristic	Word-count (-250 words)	Length fixed during training	Explicit sentence-number control (3–25)

3.8.3. Experimental Setup

To ensure consistency and equal experimental conditions, all experiments were conducted using default parameters. No fine-tuning was applied to BERTSUM. The proposed system experimental setup was configured as shown in Table 3.

Table 3. Experimental Setup Parameters.

Parameter	Search Space	
Number of topics	DUC 2006 : 20 topics	DUC 2007 : 25 topics
Candidate pool (highest-ranked sentences per topic)	Top 27 – 50	
Pattern mining	FP-growth	
Summary length (Fixed sentence budgets)	3, 5, 10, 15, 20, and 25	
Redundancy control	Cosine similarity	
Penalty weight (δ)	5	
Evaluation	ROUGE-1, ROUGE-2, ROUGE-SU4	

The diversity penalty weight δ was set to 5 after preliminary tuning to balance coverage and redundancy suppression. Cosine similarity in [0,1] was used for redundancy computation, so deduction per selection step I was bounded by δ. Since the saliency scores are computed on an unnormalized scale, this value preserves the proportional influence of the relevance and diversity terms.

3.8.4. Evaluation Metric

Lin (2004) developed ROUGE, which includes evaluation metrics and software tools for automatic summary assessment through Recall-Oriented Understudy of Gisting Evaluation. ROUGE relies on the number of overlapping contiguous words, N-grams, and word pairs between the system-generated summary to be evaluated and the ideal summaries. To assess the quality of the generated summaries, this study employs the three ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metrics, which are extensively used in extractive multi-document summarization research [55]:

ROUGE-1: The system-generated summary and reference summary are evaluated through unigram (word-level) overlap assessment. The metric measures how well the model identifies essential content words in individual texts.

ROUGE-2: The system evaluates its ability to preserve reference bigrams through its bigram overlap calculation, which assesses its maintenance of reference word sequences and their associated semantic relationships.

ROUGE-SU4: A skip-bigram measure allowing skips of up to four words in addition to unigrams. The ROUGE-SU4 metric holds special value for multi-document Summarization because it handles the different word patterns that appear between documents. The SU4 system enables researchers to conduct flexible semantic similarity assessments, as implemented in multiple DUC/TAC evaluation studies.

All scores are reported in recall form, which is standard in MDS evaluation, because the goal is to measure how much essential information from the reference summary the system successfully captures. ROUGE metrics automatically assess the quality of a summary by comparing it to a human-created gold standard. We use the ROUGE 2.0 toolkit [56] to evaluate the performance of our summarization system. The ROUGE-2.0 toolkit was used with default settings, including Porter stemming and case normalization. ROUGE-1, ROUGE-2, and ROUGE-SU4 recall scores were reported following standard DUC evaluation practice. For ROUGE-SU4, the maximum skip distance was set to 4 as defined in the original configuration. ROUGE scores were computed against all available reference summaries and averaged across document topics. Unlike some recent summarization studies that report F1 variants of ROUGE, this work strictly follows the recall-based DUC evaluation setting to ensure direct comparability with prior MDS benchmarks.

4. Results and Discussion

The results of the ROUGE evaluation for DUC 2006 and DUC 2007 are presented in Table 4. The proposed model demonstrates performance. The performance evaluation reveals several consistent patterns regarding how the proposed summarization model behaves under different sentence-length configurations. In general, integrating topic-enhanced scoring with redundancy-controlled sentence selection improves ROUGE recall scores.

4.1. Trend Across ROUGE Metrics

Fig. 3 reports the ROUGE-1, ROUGE-2, and ROUGE-SU4 recall curves of the proposed framework evaluated across varying sentence budgets on DUC 2006 and DUC 2007. It demonstrates that ROUGE-1, ROUGE-2, and ROUGE-SU4 scores steadily increase as the number of selected sentences grows from 3 to approximately 10–15. The model achieves improved topic coverage by processing additional document clusters that contain their full sentence content. The essential information becomes visible when all curves achieve their peak at the 15-sentence point. The scoring system applies redundancy penalties to eliminate statements that add minimal value. As expected, ROUGE-2 and ROUGE-SU4 have lower absolute values than ROUGE-1 because of the stricter nature of bigram and skip-bigram matching. The two metrics show the same monotonic pattern, which proves that the model maintains both phrase-level coherence and lexical relevance.

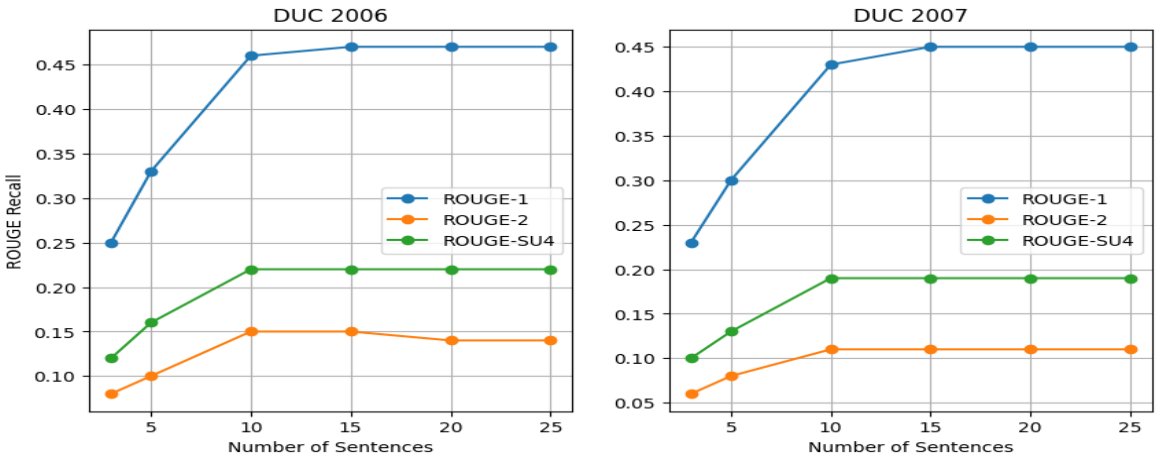

Fig. 3. Length-sensitivity analysis showing ROUGE-1, ROUGE-2, and ROUGE-SU4 recall trends across sentence budgets (3–25 sentences) for DUC 2006 and DUC 2007 datasets.

4.2. Consistency Across Datasets

Across both DUC 2006 and DUC 2007, three consistent behaviors emerge. First, the system produces superior results as text length increases from 3 to 10 sentences because it allows the inclusion of additional information about multiple subjects. Second, the system reaches its highest coverage between 10 and 15 sentences, representing the saturation point. Third, the ROUGE metrics show an identical pattern, indicating that the proposed system operates with stability while effectively handling redundant information. The proposed system demonstrates stability by preserving its patterns throughout the process, demonstrating its ability to process datasets with different sentence structures and writing styles.

4.3. Comparison with Baseline Systems

The results in Table 4 demonstrate that the proposed system outperforms LexRank, TextRank, KL-SUM, and the DUC baseline system on the ROUGE-1 and ROUGE-SU4 metrics.

Table 4. Rouge Evaluation Result on DUC 2006 and DUC 2007 Data.

System	DUC 2006			DUC 2007		
	Recall					
	Rouge-1	*Rouge-2*	*Rouge-SU4*	*Rouge-1*	*Rouge-2*	*Rouge-SU4*
Lead (1998)	NA	NA	NA	0.31	0.06	0.11
CLASSY04	NA	NA	NA	0.41	0.10	0.15
DUC Baseline	0.30	0.05	0.10	0.40	0.10	0.15
PETMSUM	0.39	0.08	0.14	0.43	0.12	0.17
KL-SUM	0.36	0.03	0.12	0.36	0.03	0.11
LexRank	0.37	0.04	0.12	0.36	0.03	0.11
TextRank	0.37	0.04	0.12	0.36	0.03	0.11
3 Sentences	0.25	0.08	0.12	0.23	0.06	0.11
5 Sentences	0.33	0.10	0.16	0.30	0.08	0.13
10 Sentences	0.46	0.15	0.22	0.43	0.11	0.19
15 Sentences	0.47	0.15	0.22	0.45	0.11	0.19
20 Sentences	0.47	0.14	0.22	0.45	0.11	0.19
25 Sentences	0.47	0.14	0.22	0.45	0.11	0.19

The model produces superior results because it employs topic–pattern representations, which enable it to capture specific lexical relationships that standard methods based on sentence centrality and surface similarity cannot. The model also surpasses PETMSUM in most ROUGE scores. For instance, PETMSUM records a ROUGE-2 score of 0.08 on DUC 2006, whereas Systems 3 and 4 reach 0.15. The

ROUGE-SU4 metric increases from 0.14 in PETMSUM to 0.22 across Systems 3 through 6. The research findings demonstrate that the enhanced scoring system, together with better diversity management in the proposed framework, produces positive results.

Most of the proposed system's ROUGE scores also surpass PETMSUM. For instance, PETMSUM records a ROUGE-2 score of 0.08 on DUC 2006, whereas the 10 Sentences System and 15 Sentences System reach 0.15. Similarly, ROUGE-SU4 increases from 0.14 (PETMSUM) to 0.22 in the 10-, 15-, 20-, and 25-Sentence Systems. These results highlight the benefit of the refined scoring strategy and improved diversity management incorporated in the proposed framework.

4.4. Length Constraint Justification

The research assesses system performance using predetermined sentence budgets, which replace the standard 250-word DUC evaluation protocol, to investigate how summary length influences the results. The analysis in Fig. 4 shows the average number of words for each sentence arrangement for fair evaluation purposes. The 15-sentence configuration enables researchers to create summaries that meet the DUC 250-word limit for standard baseline evaluations with equal assessment.

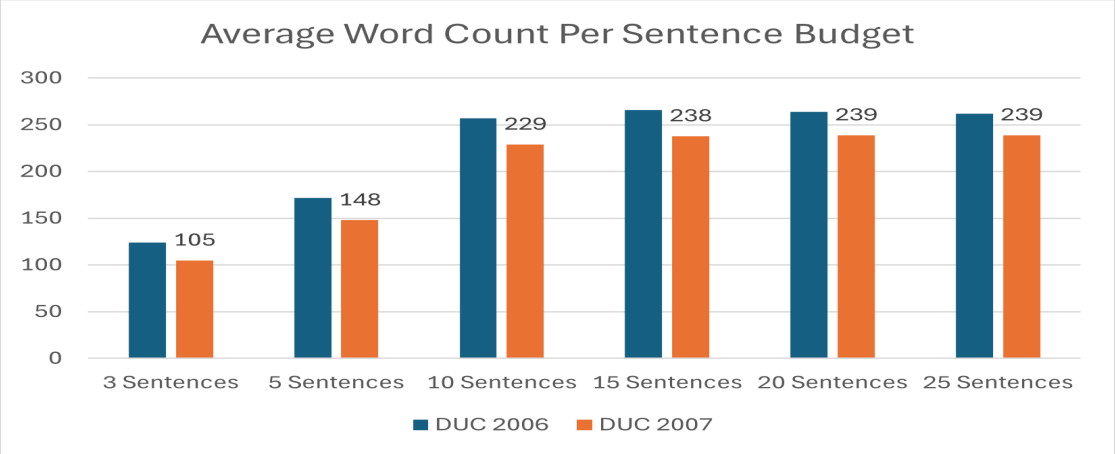

Fig. 4. Average Word Count Per Sentence Budget.

4.5. Contribution Analysis of System Components

The proposed framework comprises various components that can be studied through performance trend analysis across different baselines and sentence-length settings. The analysis reveals more about the model behavior through its results without requiring additional experimental work. The topic modeling component serves as the primary mechanism for selecting sentences based on their global importance for identifying major themes. The topic-driven mechanism enables better thematic exploration across document clusters because it operates differently from graph-based methods, such as LexRank and TextRank, which rely on sentence importance and word similarity. The system achieves higher ROUGE-1 recall, indicating it performs better at identifying essential content words.

The pattern-based lexical representation system enables improved sentence discrimination by conducting extensive analysis. The analysis of topic distributions reveals how different themes connect to each other, but it also reveals word patterns that standard topic models fail to detect. The evaluation results demonstrated that ROUGE-2 and ROUGE-SU4 achieved superior performance to PETMSUM and KL-SUM because the model enhanced its ability to activate patterns, which resulted in better phrase coherence and skip-bigram coverage.

The system needs a sentence-selection method that identifies unnecessary information to maintain summary diversity as the number of available sentences increases. The ROUGE scores reach their highest values at sentence lengths of 10–15, indicating that the diversity penalty successfully eliminates repetitive information. The proposed framework operates at a distinct operational level from systems that generate unregulated, repetitive patterns in their expanded summaries, because it maintains reliable operation by managing both the number of topics and the amount of duplicated information. The effectiveness of

redundancy control is influenced by the penalty parameter δ. The empirical tuning process showed that δ values that fell between moderate and extreme levels generated the most effective results for achieving complete coverage and producing various outcomes.

The system components operate as a single unit to deliver complete coverage through topic modeling, with pattern mining enhancing discrimination and redundancy-aware selection producing diverse results. The system performs better than classical extractive methods because their interaction yields superior results that surpass PETMSUM's performance without using neural networks or supervised learning.

4.6. Qualitative Comparison of Extracted Summaries

The quantitative assessment was supported by a qualitative assessment that investigated user behavior when choosing sentences using various summarization methods. An example document cluster from the DUC dataset was analyzed by comparing summaries generated by PETMSUM and the proposed system. The PETMSUM system generates summaries that contain relevant information but sometimes repeat themselves because it selects sentences based on topic relevance rather than uniqueness.

The system selects sentences that contain multiple subtopics and avoids repeating information that already exists. The behavior results from the system, which combines pattern-based lexical cues with its redundancy-aware selection mechanism. The qualitative observation aligns with the quantitative results reported in Table 4, which show that the proposed system achieves higher ROUGE-SU4 scores. The system produces better results, showing improved preservation of original phrases and reduced repetition, thus demonstrating that the method yields better numerical results and more diverse, easier-to-understand summaries.

4.7. Complexity Analysis

The computational complexity of the proposed model consists of LDA inference O(NK), FP-growth pattern mining that depends on the support threshold, and Greedy sentence selection $O(|S|^2)$ in the worst case due to pairwise similarity. Compared to transformer-based neural models that require large-scale pretraining and GPU acceleration, the proposed approach operates efficiently on CPU-based environments, making it suitable for constrained deployment settings.

4.8. Non-Comparable Neural Baseline

BERTSUM exhibits inconsistent performance across the datasets; on DUC 2006, ROUGE-1 reaches 0.29, while on DUC 2007, it reaches 0.15. BERT achieves relatively high ROUGE-2 and ROUGE-SU4 scores, which are 0.23 on DUC 2006, but performs poorly on DUC 2007 across all metrics (ROUGE-2 = 0.07 and ROUGE-SU4 = 0.08). This behavior is expected because:

- BERT was applied without multi-document adaptation and without dataset-specific fine-tuning. All documents for each topic were concatenated prior to encoding, which may result in truncation due to model input limits.
- The DUC creates difficulties for neural extractive models because these systems lack the ability to connect information between different documents during their training process.

The results show that these neural methods perform well, yet the research confirms that extractive methods using topic patterns remain essential for creating summaries from multiple documents.

4.9. Overall Interpretation

The results confirm several important findings: A summary requires 10 to 15 sentences that present all required information without repeating any details. The proposed method outperforms PETMSUM by maintaining stronger lexical associations, resulting in improved ROUGE-2 and ROUGE-SU4 scores. The proposed method outperforms graph-based baselines, including LexRank and TextRank, by identifying crucial topical connections that standard similarity assessment methods cannot detect. The system demonstrates excellent performance in identifying cluster content through its combination of pattern representation with diversity-aware selection methods. The research results show that the developed method outperforms the current reference techniques. These findings suggest that sentence-

budget–aware extractive summarization remains a competitive alternative when interpretability and domain independence are required. The following section discusses limitations and future work.

Fig. 3 reports the ROUGE-1, ROUGE-2, and ROUGE-SU4 recall curves of the proposed framework evaluated across varying sentence budgets on DUC 2006 and DUC 2007. It demonstrates that ROUGE-1, ROUGE-2, and ROUGE-SU4 scores steadily increase as the number of selected sentences grows from 3 to approximately 10–15. The model improves topic coverage by processing additional document clusters that contain their full-sentence content. The essential information becomes visible when all curves achieve their peak at the 15-sentence point. The scoring system applies redundancy penalties to eliminate statements that add minimal value. As expected, ROUGE-2 and ROUGE-SU4 have lower absolute values than ROUGE-1 because of the stricter nature of bigram and skip-bigram matching. The two metrics show the same monotonic pattern, which proves that the model maintains both phrase-level coherence and lexical relevance.

4.10. Limitations

The proposed model outperforms strong baselines; however, some limitations remain. The model depends on two main components: topic modeling and frequent pattern extraction, as these methods require statistical properties of the input corpus. The process of extracting topics and patterns becomes less accurate when analyzing domains that employ intricate vocabulary, impose word-choice restrictions, and exhibit irregular language patterns. This limitation contributes to reduced bigram coherence, as observed in DUC 2007. Second, the pattern interpretation component uses heuristic weighting to evaluate three elements: frequency, topic specificity, and distributional spread. The system achieves its purpose, but it fails to detect complete semantic connections that advanced embedding-based models could identify in complex contexts. The redundancy penalty system increases textual diversity but fails to account for how different sentences relate in a discourse, resulting in the removal of essential information.

5. Conclusion

The research develops an improved system for multi-document extractive summarization that combines three key elements to retrieve essential information from documents. The model generates better sentence representations by combining LDA-derived topic knowledge with standard lexical patterns, thereby improving topic identification and content selection. The system selects sentences by scoring them, combining the scores with a diversity penalty to find essential information while minimizing duplicate content. The proposed method achieves superior results compared to several traditional summarization methods in experiments conducted on the DUC 2006 and DUC 2007 datasets. The proposed method outperforms LexRank, TextRank, KL-SUM, and the standard DUC baseline in summarization tasks. It also gives clear improvements over PETMSUM, especially on ROUGE-2 and ROUGE-SU4. The model produced dependable results when processing summaries of varying lengths, demonstrating its ability to operate flexibly and with stability. The system enhances its extractive multi-document summarization capabilities by combining topic–pattern representations with diversity-aware selection methods, but it continues to struggle with precise topic identification and pattern usage, as well as with maintaining discourse coherence. The research offers multiple avenues for scientists to advance their findings. The system would achieve better semantic pattern detection by implementing neural topic representations, an advanced topic modeling approach. A hybrid summarization framework that combines extractive and abstractive methods would generate summaries that preserve the document's organization and maintain a readable format. The model needs additional evaluation with datasets from different languages and specific domains, including legal, clinical, and medical text, to achieve better generalization and stability. The ROUGE-based evaluation system provides standardized comparison methods, but researchers should consider using human assessment metrics, such as readability and coherence, in their future studies.

Acknowledgment

The authors would like to thank the Indonesia Endowment Fund for Education (LPDP) and Telkom University for their support in their study.

Declarations

Author contribution. All authors contributed equally to this paper, except for the development of the proposed model; the main author, who also serves as the corresponding author, played a full role in this task, and all authors read and approved the final paper.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

References

[1] M. Allahyari *et al.*, "Text Summarization Techniques: A Brief Survey," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 10, pp. 397–405, Oct. 2017, doi: 10.14569/IJACSA.2017.081052.

[2] Z. Huang, X. Chen, Y. Wang, J. Huang, and X. Zhao, "A survey on biomedical automatic text summarization with large language models," *Inf. Process. Manag.*, vol. 62, no. 5, p. 104216, Sep. 2025, doi: 10.1016/j.ipm.2025.104216.

[3] A. Nenkova and K. McKeown, "A Survey of Text Summarization Techniques," in *Mining Text Data*, vol. 9781461432, Boston, MA: Springer US, 2012, pp. 43–76. doi: 10.1007/978-1-4614-3223-4_3.

[4] R. Barzilay and K. R. McKeown, "Sentence Fusion for Multidocument News Summarization," *Comput. Linguist.*, vol. 31, no. 3, pp. 297–328, Sep. 2005, doi: 10.1162/089120105774321091.

[5] X. Zhao, T. Wu, X. Zheng, and R. Li, "Discussions on observer design of nonlinear positive systems via T–S fuzzy modeling," *Neurocomputing*, vol. 157, no. June, pp. 70–75, Jun. 2015, doi: 10.1016/j.neucom.2015.01.034.

[6] H. P. Luhn, "The Automatic Creation of Literature Abstracts," *IBM J. Res. Dev.*, vol. 2, no. 2, pp. 159–165, Apr. 1958, doi: 10.1147/rd.22.0159.

[7] H. P. Edmundson, "New Methods in Automatic Extracting," *J. ACM*, vol. 16, no. 2, pp. 264–285, Apr. 1969, doi: 10.1145/321510.321519.

[8] D. R. Radev, E. Hovy, and K. McKeown, "Introduction to the Special Issue on Summarization," *Comput. Linguist.*, vol. 28, no. 4, pp. 399–408, Dec. 2002, doi: 10.1162/089120102762671927.

[9] M. Gambhir and V. Gupta, "Recent automatic text summarization techniques: a survey," *Artif. Intell. Rev.*, vol. 47, no. 1, pp. 1–66, Jan. 2017, doi: 10.1007/s10462-016-9475-9.

[10] U. Ihsan, H. Ashraf, and N. Jhanjhi, "Survey on Multi-Document Summarization: Systematic Literature Review." [Online]. Available: https://arxiv.org/abs/2312.12915.

[11] E. Alanzi and S. Alballaa, "Query-Focused Multi-document Summarization Survey," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 6, pp. 822–833, Jun. 2023, doi: 10.14569/IJACSA.2023.0140688.

[12] A. S. Karnyoto, M. M. Henry, and B. Pardamean, "LDA Topic Modeling for Bioinformatics Terms in arXiv Documents," *Procedia Comput. Sci.*, vol. 245, pp. 229–238, Jan. 2024, doi: 10.1016/j.procs.2024.10.247.

[13] S. S. Mim, D. Logofatu, G. Guerrero-Contreras, and I. Medina-Bulo, "Leveraging Topic Modeling and Extractive Summarization for Unlocking Insights from NeurIPS Papers," in *2024 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, IEEE, Sep. 2024, pp. 1–6. doi: 10.1109/INISTA62901.2024.10683818.

[14] N. M. AbdelAziz, A. A. Ali, S. M. Naguib, and L. S. Fayed, "Clustering-based topic modeling for biomedical documents extractive text summarization," *J. Supercomput.*, vol. 81, no. 1, p. 171, Jan. 2025, doi: 10.1007/s11227-024-06640-6.

[15] R. K. Roul, Navpreet, and S. Nalband, "Unified Scoring and Topic Modeling: A Combined Approach for Superior Multi-document Summarization," in *Lecture Notes in Networks and Systems*, vol. 1344 LNNS,

Springer Science and Business Media Deutschland GmbH, 2025, pp. 481–493. doi: 10.1007/978-981-96-5958-6_42.

[16] D. Wang, S. Zhu, T. Li, and Y. Gong, "Multi-document summarization using sentence-based topic models," in *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers on - ACL-IJCNLP '09*, Morristown, NJ, USA: Association for Computational Linguistics, 2009, p. 297. doi: 10.3115/1667583.1667675.

[17] Y. Wu, Y. Gao, Y. Li, Y. Xu, and M. Chen, "Mining Topical Relevant Patterns for Multi-document Summarization," in *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, IEEE, Dec. 2015, pp. 114–117. doi: 10.1109/WI-IAT.2015.136.

[18] D. R. Radev, H. Jing, M. Styś, and D. Tam, "Centroid-based summarization of multiple documents," *Inf. Process. Manag.*, vol. 40, no. 6, pp. 919–938, Nov. 2004, doi: 10.1016/j.ipm.2003.10.006.

[19] Y. Gong and X. Liu, "Generic text summarization using relevance measure and latent semantic analysis," in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA: ACM, Sep. 2001, pp. 19–25. doi: 10.1145/383952.383955.

[20] A. Haghighi and L. Vanderwende, "Exploring content models for multi-document summarization," in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics on - NAACL '09*, Morristown, NJ, USA: Association for Computational Linguistics, 2009, p. 362. doi: 10.3115/1620754.1620807.

[21] D. H.-T. Asli Celikyilmaz, "A hybrid hierarchical model for multi-document summarization | Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics," 2010, pp. 815–824. [Online]. Available: https://aclanthology.org/P10-1084/.

[22] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003, [Online]. Available: https://dl.acm.org/doi/10.5555/944919.944937.

[23] J.-P. Qiang, P. Chen, W. Ding, F. Xie, and X. Wu, "Multi-document summarization using closed patterns," *Knowledge-Based Syst.*, vol. 99, no. March, pp. 28–38, May 2016, doi: 10.1016/j.knosys.2016.01.030.

[24] R. Mihalcea and D. Radev, *Graph-based Natural Language Processing and Information Retrieval*. Cambridge University Press, 2011. doi: 10.1017/CBO9780511976247.

[25] G. Erkan and D. R. Radev, "LexRank: Graph-based Lexical Centrality as Salience in Text Summarization," *J. Artif. Intell. Res.*, vol. 22, pp. 457–479, Dec. 2004, doi: 10.1613/jair.1523.

[26] R. Mihalcea, "Unsupervised large-vocabulary word sense disambiguation with graph-based algorithms for sequence data labeling," in *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT '05*, Morristown, NJ, USA: Association for Computational Linguistics, 2005, pp. 411–418. doi: 10.3115/1220575.1220627.

[27] A. P. Widyassari *et al.*, "Review of automatic text summarization techniques & methods," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 4, pp. 1029–1046, Apr. 2022, doi: 10.1016/j.jksuci.2020.05.006.

[28] S. Mulla and N. F. Shaikh, "Over comparative study of text summarization techniques based on graph neural networks," *Web Intell.*, vol. 22, no. 2, pp. 231–248, Apr. 2024, doi: 10.3233/WEB-230014.

[29] Z. Qilin, W. Yu, X. Jian, Z. Qilin, W. Yu, and X. Jian, "Extractive Document Summarization Model Based on Heterogeneous Graph and Keywords," *J. Univ. Electron. Sci. Technol. China, 2024, Vol. 53, Issue 2, Pages 259-270*, vol. 53, no. 2, pp. 259–270, Mar. 2024, [Online]. Available: https://www.juestc.uestc.edu.cn/article/doi/10.12178/1001-0548.2023019.

[30] Z. Jalil, M. Nasir, M. Alazab, J. Nasir, T. Amjad, and A. Alqammaz, "Grapharizer: A Graph-Based Technique for Extractive Multi-Document Summarization," *Electronics*, vol. 12, no. 8, p. 1895, Apr. 2023, doi: 10.3390/electronics12081895.

[31] C. Zhou *et al.*, "Exploring Contextual Word-level Style Relevance for Unsupervised Style Transfer," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 7135–7144. doi: 10.18653/v1/2020.acl-main.639.

[32] M. Yasunaga, R. Zhang, K. Meelu, A. Pareek, K. Srinivasan, and D. Radev, "Graph-based Neural Multi-Document Summarization," in *Proceedings of the 21st Conference on Computational Natural Language*

Learning (CoNLL 2017), Stroudsburg, PA, USA: Association for Computational Linguistics, 2017, pp. 452–462. doi: 10.18653/v1/K17-1045.

[33] E. Davoodijam, N. Ghadiri, M. Lotfi Shahreza, and F. Rinaldi, "MultiGBS: A multi-layer graph approach to biomedical summarization," *J. Biomed. Inform.*, vol. 116, p. 103706, Apr. 2021, doi: 10.1016/J.JBI.2021.103706.

[34] V. Giri, M. M. Math, S. Kusal, and S. Patil, "Extractive Text Summarization Employing a Graph-Based Ranking System with Context Transfer," in *Communications in Computer and Information Science*, vol. 2434 CCIS, Springer Science and Business Media Deutschland GmbH, 2025, pp. 3–19. doi: 10.1007/978-3-031-84602-1_1.

[35] N. Shabani *et al.*, "A Comprehensive Survey on Graph Summarization With Graph Neural Networks," *IEEE Trans. Artif. Intell.*, vol. 5, no. 8, pp. 3780–3800, Aug. 2024, doi: 10.1109/TAI.2024.3350545.

[36] D. Wang, P. Liu, Y. Zheng, X. Qiu, and X. Huang, "Heterogeneous Graph Neural Networks for Extractive Document Summarization," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 6209–6219. doi: 10.18653/v1/2020.acl-main.553.

[37] Y. Liu and Z. Gong, "Cycling topic graph learning for neural topic modeling," *Knowledge-Based Syst.*, vol. 310, no. February, p. 112905, Feb. 2025, doi: 10.1016/j.knosys.2024.112905.

[38] Q. Ye, B. Y. Lin, and X. Ren, "CrossFit: A Few-shot Learning Challenge for Cross-task Generalization in NLP," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2021, pp. 7163–7189. doi: 10.18653/v1/2021.emnlp-main.572.

[39] H. ; Zhao *et al.*, "A Multi-Granularity Heterogeneous Graph for Extractive Text Summarization," *Electron. 2023, Vol. 12, Page 2184*, vol. 12, no. 10, p. 2184, May 2023, doi: 10.3390/ELECTRONICS12102184.

[40] T. UÇKAN, "A hybrid model for extractive summarization: Leveraging graph entropy to improve large language model performance," *Ain Shams Eng. J.*, vol. 16, no. 5, p. 103348, Apr. 2025, doi: 10.1016/j.asej.2025.103348.

[41] S. R. Bowman and G. Dahl, "What Will it Take to Fix Benchmarking in Natural Language Understanding?," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2021, pp. 4843–4855. doi: 10.18653/v1/2021.naacl-main.385.

[42] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[43] Y. Liu and M. Lapata, "Text Summarization with Pretrained Encoders," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, pp. 3728–3738. doi: 10.18653/v1/D19-1387.

[44] A. See, P. J. Liu, and C. D. Manning, "Get To The Point: Summarization with Pointer-Generator Networks," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2017, pp. 1073–1083. doi: 10.18653/v1/P17-1099.

[45] R. Nallapati, B. Zhou, C. dos Santos, C. Gulcehre, and B. Xiang, "Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond," in *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2016, pp. 280–290. doi: 10.18653/v1/K16-1028.

[46] J. Zhang, Y. Zhao, M. Saleh, and P. J. Liu, "PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization," PMLR, Nov. 2020, pp. 11328–11339. [Online]. Available: https://dl.acm.org/doi/abs/10.5555/3524938.3525989.

[47] H. T. Dang, "Overview of DUC 2006," in *Proceedings of the Document Understanding Conference (DUC 2006)*, National Institute of Standards and Technology (NIST), 2006, pp. 1–10. [Online]. Available: https://duc.nist.gov/pubs/2006papers/duc2006.

[48] H. T. Dang, "Overview of the DUC 2007 Summarization Task," in *Proceedings of the Document Understanding Conference (DUC 2007)*, National Institute of Standards and Technology (NIST), 2007, pp. 1–38. [Online]. Available: https://duc.nist.gov/duc2007/proceedings/DUC2007_Summarization_Task.

[49] P. Over, H. Dang, and D. Harman, "DUC in context," *Inf. Process. Manag.*, vol. 43, no. 6, pp. 1506–1520, Nov. 2007, doi: 10.1016/j.ipm.2007.01.019.

[50] A. McCallum, "MALLET: A Machine Learning for Language Toolkit." [Online]. Available: http://mallet.cs.umass.edu/.

[51] H. Sun, B. Li, and B. Han, "A novel keyphrase extraction method by combining FP-growth and LDA," in *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, IEEE, Jul. 2017, pp. 1764–1768. doi: 10.1109/FSKD.2017.8393033.

[52] X. Yu, S. Zhou, and A. Liu, "Smart Tourist Attraction Data Mining Based on FP-Growth Algorithm," in *2024 International Conference on Internet of Things, Robotics and Distributed Computing (ICIRDC)*, IEEE, Dec. 2024, pp. 127–132. doi: 10.1109/ICIRDC65564.2024.00029.

[53] N. Ramakrishnan, M. Nair J., D. Jayaprakash, H. Ananthakrishnan, and S. Rani S., "Hypergraph based clustering for document similarity using FP growth algorithm," in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, IEEE, May 2019, pp. 332–336. doi: 10.1109/ICCS45141.2019.9065630.

[54] Y. Wu, Y. Li, Y. Xu, and W. Huang, "Mining Topically Coherent Patterns for Unsupervised Extractive Multi-document Summarization," in *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, IEEE, Oct. 2016, pp. 129–136. doi: 10.1109/WI.2016.0028.

[55] C.-Y. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries," 2004, pp. 74–81. [Online]. Available: https://aclanthology.org/W04-1013/.

[56] K. Ganesan, "ROUGE 2.0: Updated and Improved Measures for Evaluation of Summarization Tasks," vol. 1, pp. 1–8, Mar. 2018, [Online]. Available: https://github.com/kavgan/ROUGE-2.0.