Journal of Universal Computer Science, vol. 32, no. 11 (2026), 1590-1611
submitted: 14/8/2025, accepted: 11/2/2026, appeared: 28/10/2026 CC BY 4.0

SciRelate: A Framework for Related Work Generation with TopicFusion-MP for Subsection Heading Extraction

Army Justitia
(National Cheng Kung University, Tainan, Taiwan
Universitas Airlangga, Surabaya, Indonesia
https://orcid.org/0000-0002-4306-1634, army-j@fst.unair.ac.id)

Hei-Chia Wang
(National Cheng Kung University, Tainan, Taiwan
https://orcid.org/0000-0002-5790-7506, hcwang@mail.ncku.edu.tw)

Abstract: Writing the Related Work section is a challenging task for inexperienced researchers, requiring significant time, effort, and skill. This study aims to ease this burden by introducing an ICT-based framework that automates the generation of cohesive and well-structured Related Work sections. We propose SciRelate, a novel framework that combines abstractive multidocument summarization and automated keyword-based subsection heading extraction. SciRelate comprises three main modules: (1) document clustering using Latent Dirichlet Allocation (LDA), (2) related Work generation via a fine-tuned Bidirectional and Autoregressive Transformer (BART) model, and (3) Topic keyword extraction using TopicFusion-MP, a hybrid method that integrates LDA, TextRank, and MPNET embeddings. Experiments were conducted using the NudtRwG dataset, and BART was evaluated using 10-fold cross-validation. LDA clustering effectively identified subsection headings aligned with ground truth structures. The BART model achieved high ROUGE scores: 81.53 (ROUGE-1), 76.19 (ROUGE-2), and 79.03 (ROUGE-L)**.** TopicFusion-MP outperformed six baseline keyword extraction methods in four scenarios, reaching a top similarity score of 0.6398**.** This study presents a comprehensive, automated pipeline for generating Related Work sections, integrating clustering, summarization, and keyword extraction in a unified framework. SciRelate highlights the potential of ICT tools to support scientific writing, particularly for new researchers.

Keywords: Abstractive multidocument summarization, BART summarizer, Information and communication technology, Related work generation, Subsection heading extraction, TopicFusion-MP
Categories: I.7, I.2.7, I.5.3
DOI: 10.3897/jucs.168819

1 Introduction

The *Related Work* section of a scientific article synthesizes prior studies to provide essential background and a thematic overview of the field [Chen and Zhuge, 2016]. It highlights research gaps or advantages of the current study and supports methodological choices with citations [Chen et al., 2022]. However, new researchers often struggle to organize diverse sources into a coherent narrative. Automatic summarization systems offer a potential solution by assisting in drafting Related Work sections from multiple articles.

Abstractive multidocument summarization is vital in academic research for producing coherent and original summaries that overcome the inherent limitations of extractive methods. Since this task entails multiple lengthy source documents, several transformer-based architectures have been developed for processing them. PEGASUS accommodates inputs of up to 1,024 tokens [Zhang et al., 2020], and PRIMERA supports up to 4,096 tokens [Xiao et al., 2022], whereas versions such as BigBird [Zaheer et al., 2020], the Longformer encoder–decoder (LED) [Beltagy et al., 2020], and LongT5 [Guo et al., 2021b] provide increased input capacities of up to 16,384 tokens. Empirical evidence indicates that the bidirectional autoregressive transformer (BART) [Lewis et al., 2020] outperforms these models in summarizing related work [Justitia and Wang, 2025]. Therefore, we adopt the pretrained BART model for generating *Related Work* sections.

Structurally, the *Related Work* section may be presented as a single cohesive unit or divided into subsections based on topic complexity and discussion. Grouping by approach, methodology, or thematic [Aliyu et al., 2020], it helps readers better comprehend the structure and content of the text [Monreal et al., 2004]. In cognitive psychology, subsection headings have been shown to activate readers' prior knowledge and support memory processes related to text organization [Lemarié et al., 2012]. In this study, we extract keywords from a collection of text and utilize them as headings for each subsection.

Keyword extraction techniques fall into three primary approaches: statistical methods, topic modeling, and word graph-based models. Term frequency–inverse document frequency (TF–IDF) is a popular statistical method due to its simplicity but is less effective in technical domains. Nonetheless, TF–IDF depends largely on word frequency, rendering it less precise in specialized or technical domain texts. Researchers have implemented several improvements by modifying word weights [Wang, 2024] and positional weights [Li, 2021] or by integrating a weight-balancing method [Zhuohao et al., 2021] and contextual word embedding [Khan et al., 2022]. Latent Dirichlet Allocation (LDA) is a prevalent technique utilized in topic modeling [Song et al., 2009]. Nonetheless, LDA has deficiencies in terms of low topic interpretability [Harandizadeh et al., 2022], susceptibility to document length [Guo et al., 2021a], and sparsity in short documents [Rashid et al., 2019], necessitating modifications using alternative methodologies for improved outcomes. TextRank, a modification of PageRank, is a graph-based unsupervised model utilized for ranking terms. TextRank was introduced to assess the significance of words by analyzing their interconnections within a graph rather than relying solely on frequency, hence providing equitable consideration of essential although infrequently occurring terms [Gulati et al., 2023]. The integration of TextRank with alternative methodologies yields increased efficacy in keyword extraction [Zhang et al., 2024] and text summarization [Gulati et al., 2023] tasks. In our study, we amalgamate the benefits of LDA and TextRank for keyword extraction; furthermore, we incorporate MPNET contextualized sentence embedding to augment the semantic representation of words and sentences [Song et al., 2020] and identify the most representative keywords on the basis of cosine similarity without requiring labeling [Reimers and Gurevych, 2019].

This study aims to introduce an abstractive multidocument summarization system for the *Related Work* section of scientific articles using a pretrained BART model. Furthermore, we identify subsection headings from the generated text by employing a

method that integrates LDA, TextRank, and MPNET sentence embeddings. Prior to this, we grouped referenced papers with similar topics using an unsupervised clustering algorithm. Specifically, we investigate the performance of LDA and agglomerative hierarchical clustering (AHC) in grouping documents according to the number of topics present in the dataset. The key contributions of this study are as follows:

1. We cluster referenced papers using LDA and agglomerative hierarchical clustering (AHC) and evaluate their performance. The performance is assessed by comparing the number of subsection headings in the dataset with the number of clusters produced by LDA and AHC;
2. We fine-tune a pretrained BART model to perform abstractive multidocument summarization for automatically related work generation. The hyperparameters are optimized to maximize the ROUGE scores for the generated text;
3. We identify subsection headings within the generated Related Work text using a topic fusion keyword extraction method that integrates LDA, TextRank, and MPNET embeddings;
4. We compare our topic fusion keyword extraction approach with a prior keyword extraction method by calculating the cosine similarity between the extracted subsection headings and the generated Related Work text.

2 Literature Review

Writing the Related Work section involves multidocument summarization to distill essential and relevant information related to a specific research topic or problem. Although extractive summarization approaches have been widely proposed [Chen and Zhuge, 2019, Deng et al., 2021, Hoang and Kan, 2010, Hu and Wan, 2014, Wang et al., 2020b, 2020c], abstractive summarization is considered a more natural solution for this task because of its ability to paraphrase while preserving the original meaning and context [Chen et al., 2022].

Abstractive summarization has emerged from advancements in neural network technologies, particularly sequence-to-sequence architectures and transformer-based models, for text generation. To date, the output of related work summarization has generally focused on two distinct forms: citation sentence generation and related work paragraphs. Studies such as [AbuRa'ed et al., 2020, Ge et al., 2021, Gu and Hahnloser, 2022, Luu et al., 2021, Xing et al., 2020] have introduced approaches that allow only a single citation sentence to be generated. In contrast, [Jung et al., 2022] explored a method for generating multiple citation sentences and identifying the intent category of each citation. However, all of these works produce citations referring to a single source paper. Research by [Li et al., 2022, 2024, Mandal et al., 2024] has advanced this field through the generation of multiple cited texts spanning multiple cited papers. Nevertheless, these approaches still yield isolated citation sentences.

In subsequent work, methods for generating coherent paragraphs composed of several interrelated citation sentences were explored. For example, in [Chen et al., 2021, 2022] the target text was defined as a related work paragraph containing at least two citation sentences. [Chen et al., 2021] introduced the relation-aware related work generator (RRG), which comprises a hierarchical encoder, relationship modelling, and

a related work generator. They introduced graph-based relationship modelling to capture interpaper relations. This architecture was later improved by replacing the hierarchical encoder with a graph encoder and a target-aware graph encoder (TEG) [Chen et al., 2022]. During the decoding stage, they incorporated target-level attention, reference-level attention, and keyphrase-level attention. Their experiments demonstrated that the TEG outperformed the original RRG architecture. Moreover, [Liu et al., 2023] focused on modelling the relationships among citation sentences and referenced papers for sentence ordering and inserted transitional content to connect them. [Li and Ouyang, 2025] explored a large language model (LLM) prompting approach to perform a task similar to that proposed by [Liu et al., 2023]. [Martin-boyle et al., 2009] introduced ScholaCite, a GPT-4-based system that clusters citations by topic, identifies the rationale behind each group in supporting the current related work task, and generates related work text using the citations, topic clusters, and corresponding rationales. Research in related work summarization continues to focus on exploring various approaches to produce more accurate and relevant summaries. Within this context, the present study aims to expand the exploration space by identifying subsection headings from the generated paragraphs. This approach is expected to support a more systematic structure in the Related Work section.

3 Proposed Framework

Our proposed framework, SciRelate, enables users to provide a customized collection of abstracts from referenced papers retrieved through their searches. SciRelate is built with three main modules to generate the related work text: modules that cluster documents, generate text, and extract topics for subsection headings. Figure 1 presents an overview of the SciRelate framework and its modular architecture. The subsequent explanations detail the functionality of each module.

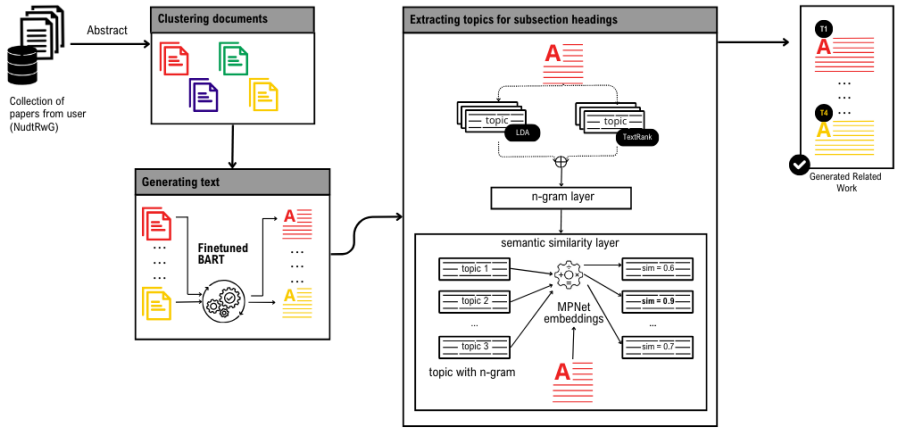


Figure 1: SciRelate framework

3.1 Clustering documents

SciRelate allows users to upload abstracts from referenced papers in Microsoft Excel (.xlsx) or comma-separated values (.csv). The input consists of mandatory attributes (i.e., authors' names, article titles, source titles, abstracts, and publication years) and optional attributes (i.e., volume, issue, start page, end page, DOI, and DOI link). The abstract serves as the primary input for generating the Related Work section. The first module clusters documents using an unsupervised method, where the number of clusters defines the candidate subsection headings. Two methods, LDA and AHC, are tested without requiring a preset cluster number, and the one yielding the best performance is used in the system.

3.1.1 LDA cluster

LDA is a topic-modelling technique used to uncover latent topics within a text corpus. LDA employs a Bayesian methodology that characterizes each text as a probabilistic distribution across many subjects, with each topic represented as a probabilistic distribution over the words in the corpus [Blei et al., 2003].

A collection of abstracts is transformed into a numerical representation as a bag-of-words matrix. We quantify the frequency of each word in the collection and produce a document-term matrix. We control word frequency by filtering out overly common and rare words. Common words, such as "the", "is", and "it", are often stop words, whereas rare words provide less value for topic modelling. To manage this filtering, we set df_{max} and df_{min} thresholds. df_{max} controls the maximum threshold for the occurrence of words in documents, whereas df_{min} controls the minimum threshold for the occurrence of words in documents. For df_{min}, we consider the number of reference documents provided by the user. If the user inputs too few reference documents, the minimum word occurrence threshold must also be adjusted. We choose two different values for df_{max} and df_{min}, as shown in Eqs. 1 to 4:

Values for df_{max}

$$df_{max} = 0.8 \tag{1}$$

$$df_{max} = 0.9 \tag{2}$$

Values for df_{min}

$$df_{min} = \left\lceil \frac{|documents|}{2} \right\rceil \tag{3}$$

$$df_{min} = \begin{cases} 2, & |documents| > 2 \\ \left\lceil \frac{|documents|}{2} \right\rceil, & otherwise \end{cases} \tag{4}$$

$df_{max} = 0.9$ indicates that any word present in more than 90% of the texts is eliminated. For the first value, $df_{min} = \left\lceil \frac{|documents|}{2} \right\rceil$ eliminates terms that occur in fewer than half of the total texts. For the second value, if the number of documents exceeds two, terms that occur in two documents are eliminated. Otherwise, the formula for the first value is applied.

One of the main challenges in using LDA is determining the optimal number of topics. This study aims to minimize user intervention by eliminating the need for manual configuration of the number of clusters. To achieve this goal, a coherence score optimization mechanism is employed to determine the most suitable number of subheading clusters. The coherence score measures the semantic consistency of a topic by evaluating the similarity between words within it [Rishu and Kukreja, 2024]. The Gensim library is used to compute the coherence score using the 'c_v' parameter. The 'c_v' parameter, which integrates the indirect cosine measure with NPMI and a Boolean sliding window of size 110 [He et al., 2016], has been shown to outperform other coherence measures [Röder et al., 2015]. The optimal number of clusters is selected when the coherence score reaches its maximum value for *k* values ranging from 1 to 5.

3.1.2 AHC

AHC is a bottom-up hierarchical clustering method that begins by treating each data point as its own cluster and iteratively merging the most similar pairs until a larger structure emerges. Unlike divisive (top-down) clustering, AHC requires no predefined number of clusters [Vijaya et al., 2019] and demonstrates better computational efficiency [Chehreghani, 2021]. Abstracts are numerically represented using an LDA-like framework, including word frequency control. Ward's linkage is employed for merging, minimizing intracluster variance to create compact, balanced clusters [Ridwan and Retnawati, 2021]. Compared to single or complete linkage**,** Ward's method provides more stable results, reduces the chain effect, and improves clustering performance [Ridwan and Retnawati, 2021].

3.2 Generating the Related Work section

We fine-tune a pretrained BART-large-CNN model to summarize each document cluster and draft the Related Work section (Figure 2). To keep the summary on topic, the target paper's abstract is prepended to the abstracts of its references, with <REV_BREAK> as a separator. The concatenated text is converted to token embeddings E_t; position embedding E_p are then computed via the standard sine-cosine formulas (Eqs. 5-6).

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right) \tag{5}$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right) \tag{6}$$

where pos is the position, i is the dimension, and d_{model} is the embedding size. For a sequence of length n, the final embedding E_i at position i combines token embedding E_t and position embedding E_p, as shown in Eq. 7. E_i; then, it fed into BART's attention layer.

$$E_i = E_t + E_p \tag{7}$$

Hyperparameter	BART
Number of epochs	10
Random seed	42
Train batch size	4
Optimizer	AdamW
Learning rate	5e-5
Weight decay	0.01
Gradient accumulation steps	1
Max input length	512
Max output length	256

Table 1: Hyperparameter settings for the BART model

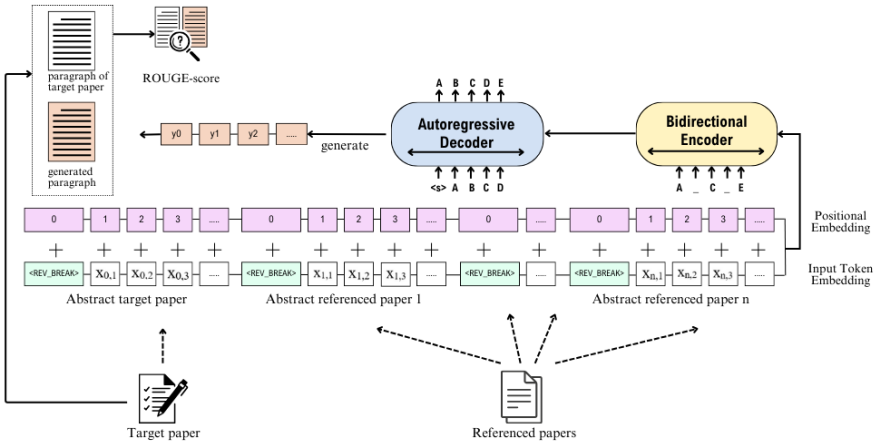


Figure 2: Architecture of the pretrained BART model used in this study

BART enhances text generation through a denoising mechanism that introduces noise—such as rotation, permutation, infilling, masking, and deletion—into the input [La Quatra and Cagliero, 2023]. Its bidirectional encoder processes the corrupted text to create rich representations, while the decoder autoregressively reconstructs the original text using past outputs [Lewis et al., 2020]. BART's architecture includes 12 encoder and decoder layers, 16 attention heads, a 1024 hidden size, a 4096 feedforward size, and 406 million parameters. Table 1 presents its hyperparameter settings.

3.3 Extracting keywords for subsection headings

The third module extracts keywords from the generated text to serve as subsection headings. Our method, TopicFusion-MP, combines LDA, TextRank, and MPNET embeddings to identify high-quality keywords from a given text. In addition to flexible topic modelling, LDA uncovers latent topics [Rishu and Kukreja, 2024]. Moreover,

TextRank ranks terms based on co-occurrence or syntax [Mihalcea and Tarau, 2004] and MPNET embedding captures word dependencies and addresses positional inconsistencies [Maslim et al., 2024] by integrating the *masked language model* (MLM) from BERT and *permutation language modelling* (PLM) from XLNet. In keyword extraction tasks, MPNet outperforms other embeddings [Ashqar and Mutlu, 2023].

In this module, we apply stopword removal, tokenization, and lemmatization. After lemmatization, we use the bag-of-words (BoW) model to represent the word frequency within the document. Once the document–word matrix is constructed, the words are distributed into several topics. LDA models the probability distribution of a word w in a document d through Eq. 8 as follows:

Steps	**Result**
1. Topics by LDA	["language", "neural", "processing", "networks", "models"]
2. Topics by TextRank	["neural networks", "Machine learning models", "natural language processing"]
3. Combined topics	["natural language processing", "Machine learning models", "language", "neural", "processing", "neural networks", "networks", "models"]
4. N-gram candidates	a. Tokenization and lemmatization: ["machine", "learn", "model", "neural", "network", "wide", "use", "natural", "language", "process"] b. N-gram tokens generation: - Unigram: Machine, learn, model, neural, network, wide, use, natural, language, process - Bigram: Machine learn, learn model, model neural, neural network, network wide, … - Trigram: Machine learn model, learn model neural, model neural network, neural network wide, … c. N-gram selection with the combined topics: - machine learn (passed → contains *"machine"*) - model neural (passed → contains *"neural"*) - wide use (rejected → does not contain any keyword) - natural language (passed → contains *"natural"* and *"language"*) d. Combine and remove duplicates ["machine", "learn", "neural", "network", "machine learn", "model neural", "natural language", "language process", "machine learn model", "learn model neural", "natural language process"]
5. Semantic similarity using MPNet embeddings	Select a top-1 topic by calculating cosine semantic similarity between the given text and the n-gram candidate topics

Table 2: Steps and results of TopicFusion-MP

$$P(w|d) = \sum_{z} P(w|z)\, P(z|d) \tag{8}$$

where $P(w|z)$ represents the probability of word w appearing in topic z and $P(z|d)$ denotes the probability of topic z appearing in document d. LDA subsequently yields a list that represents the key terms extracted from the text.

In TextRank, words are converted into nodes and linked based on proximity or syntax. Each word starts with an initial weight $\omega_{sen} = 1$, which is updated using a PageRank-based formula [Wang et al., 2020a], as shown in Eq. 9.

$$\omega_{sen}(Sen_i) = \frac{1-\alpha}{d} + \alpha * \sum_{Sen_j \in In(Sen_i)} \frac{Sim_{ji}}{\sum_{Sen_k \epsilon Out(Sen_i)} Sim_{jk}} \omega_{sen}(Sen_j) \tag{9}$$

where $\omega_{sen}(Sen_i)$ represents the weight of sentence Sen_i, α is a damping coefficient (usually 0.85), and d is the total number of sentences. Sim_{ji} represents the semantic similarity between sentence Sen_j and Sen_i. $In(Sen_i)$ indicates a set of sentences that point to Sen_i; otherwise, $Out(Sen_i)$ indicates a set of sentences to which Sen_i points. This weight adjustment is performed iteratively to update the word weights until convergence. These weights indicate the importance level of each word or phrase. Like LDA, TextRank produces a set of words or phrases that represent the key terms in the text.

Next, we concatenate salient terms from LDA and TextRank into a unified key term list. In a separate process, we generate n-gram topic candidates (up to trigrams) and retain only those containing words from this list. To filter semantically, we compute cosine similarity between each candidate and the generated text using MPNet embeddings, applying a threshold of 0.4. Table 2 presents the results for the given text: *"Machine learning models such as neural networks are widely used in natural language processing."*

4 Experiments and analysis

Our empirical experiments are presented in this section. First, the dataset is introduced, followed by three different experiments. The results of the experiments are presented along with an analysis of the parameters.

4.1 Dataset

We conducted our experiments using the NudtRwG dataset [Wang et al., 2019], which is a set of published papers in computational linguistics and natural language processing conferences. NudtRwG is composed of 2,084 target papers and over 52,000 referenced papers. Due to its comprehensiveness in providing target papers, ground truth, and corresponding reference papers, this dataset offers a large amount of high-quality data for this task compared with other datasets. Given a target paper D, D^{rw} is the related work text of D, and D^a represents the abstract of D. D has a reference paper set $R^a = \{R_1^a \oplus R_2^a \oplus ... \oplus R_n^a\}$, where n is the number of abstracts a from the reference paper and where $\oplus$ indicates concatenation. Since the text in D^{rw} of the NudtRwG dataset is still original, we separate each paragraph to produce D_i^{rw}. D_i^{rw} is the i^{th} paragraph in

the Related Work section of D. If D has i paragraphs, then D inserts i rows into the data array in the dataset. Thus, the number of rows in the dataset represents the number of paragraphs in the entire dataset. Since our model predicts text in one paragraph, n in R_n^a represents the number of referenced papers cited in the corresponding paragraph. Therefore, the final seq2seq input follows this notation $< D^a \oplus R^a, D_i^{rw} >$. Table 3 shows the descriptive statistics of the NudtRwG dataset.

ΣParagraphs of RW	Σ_{RPs} in RW			Σ_{words} of source papers			Σ_{words} of target papers			Σ_{RPs} in each paragraph of RW			Σ_{SHs} in RW		
	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg
7,889	5	96	25	56	4,847	625	7	875	95	1	34	5	1	8	1.35

*Note: RW denotes related work, RPs denotes referenced papers, and SHs denotes subsection headings

Table 3: Descriptive statistics of the NudtRwG dataset

Method	Scenario	Parameter setup
LDA	LDA_1	max_df set to Eq. 1, min_df set to Eq. 3
	LDA_2	max_df set to Eq. 1, min_df set to Eq. 4
	LDA_3	max_df set to Eq. 2, min_df set to Eq. 3
	LDA_4	max_df set to Eq. 2, min_df set to Eq. 4
AHC	AHC_1	max_df set to Eq. 1, min_df set to Eq. 3, distance_threshold = 1.5
	AHC_2	max_df set to Eq. 1, min_df set to Eq. 3, distance_threshold = 1.6
	AHC_3	max_df set to Eq. 1, min_df set to Eq. 3, distance_threshold = 1.7
	AHC_4	max_df set to Eq. 1, min_df set to Eq. 4, distance_threshold = 1.5
	AHC_5	max_df set to Eq. 1, min_df set to Eq. 4, distance_threshold = 1.6
	AHC_6	max_df set to Eq. 1, min_df set to Eq. 4, distance_threshold = 1.7
	AHC_7	max_df set to Eq. 2, min_df set to Eq. 3, distance_threshold = 1.5
	AHC_8	max_df set to Eq. 2, min_df set to Eq. 3, distance_threshold = 1.6
	AHC_9	max_df set to Eq. 2, min_df set to Eq. 3, distance_threshold = 1.7
	AHC_{10}	max_df set to Eq. 2, min_df set to Eq. 4, distance_threshold = 1.5
	AHC_{11}	max_df set to Eq. 2, min_df set to Eq. 4, distance_threshold = 1.6
	AHC_{12}	max_df set to Eq. 2, min_df set to Eq. 4, distance_threshold = 1.7

Table 4: Parameter settings for the LDA and AHC methods

4.2 Experiments

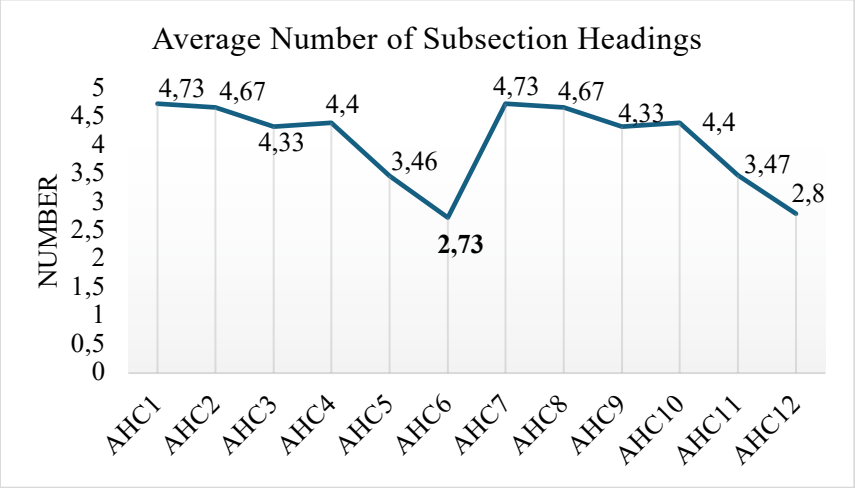
4.2.1 Experiment 1: Different clustering methods

This experiment evaluates clustering methods to find optimal topic-based clusters for forming subsections. The ideal clusters group similar documents while balancing the number of subsections with the number of cited papers. We adjust key parameters, i.e., df_{max}, df_{min}, and distance_threshold. The distance_threshold controls the cluster merging. Table 4 summarizes the parameter settings for each scenario.

	LDA1	**LDA2**	**LDA3**	**LDA4**
$\bar{n}_{sh}$	**1.8**	3.13	1.93	3

Table 5: LDA results

On the basis of Table 5, among the LDA scenarios, LDA1 has the lowest average number of subsection headings, with $\bar{n}_{sh} = 1.8$. Among the LDA methods, LDA1 is the most suitable and appropriate for this study, especially for the NudtRwG dataset, as the number of SHs generated by LDA1 closely approximates the average SH value of 1.35, as shown in Table 3. Using the same criteria, AHC6, which yields $\bar{n}_{ss} = 2.7$, performs the best in determining the number of clusters, as shown in Figure 3. When comparing both methods, LDA1 outperforms AHC6 by producing a more compact cluster structure. Therefore, we employ LDA as the clustering method in the document clustering module.



*Note: *sh* denotes the subsection heading

Figure 3: AHC results

Coherence score was employed as the standard validation metric for the clustering method, as described in the preceding section. We selected two articles to illustrate the coherence score results obtained from the clustering step using the best selected method, LDA1. As shown in Table 6, LDA grouped the referenced papers into four experimental clusters. For Article ID 102, the highest coherence score, i.e., 0.3030, was

Article ID 102, number of referenced papers: 20 articles

$\bar{n}_{cluster}$	**Coherence score**
1	0.2079
2	0.2823
3	0.2282
4	**0.3030**

Optimal number of clusters: 4

Article ID 121, number of referenced papers: 13 articles

$\bar{n}_{cluster}$	**Coherence score**
1	0.2339
2	**0.2593**
3	0.2547
4	0.2444

Optimal number of clusters: 2

Table 6: Coherence scores examples for article 102 and article 121

achieved when the LDA1 produced four clusters, whereas for Article ID 121, the optimal score was obtained with two clusters. These cluster counts subsequently informed the number of subheadings defined in our framework.

4.2.2 Experiment 2: Generating the Related Work section

Using the hyperparameter configuration specified in Table 1, a pretrained BART model was trained, and10-fold cross-validation was employed. The model's performance was assessed on the basis of validation loss as well as the ROUGE evaluation metrics, specifically ROUGE-1, ROUGE-2, and ROUGE-L. As indicated in Table 7, the optimal performance was recorded on the eighth fold, yielding ROUGE-1, ROUGE-2, and ROUGE-L scores of 81.530, 76.194, and 79.026, respectively, and a validation loss of 0.039. The model performs well in generating related work tasks, achieving average scores of 0.316, 68.130, 56.493, and 61.655 for the validation loss, ROUGE-1, ROUGE-2, and ROUGE-L, respectively.

Fold	**Validation Loss**	**ROUGE-1**	**ROUGE-2**	**ROUGE-L**
#1	2.123	33.696	6.753	17.957
#2	0.622	37.298	9.599	20.527
#3	0.110	53.342	33.959	41.948
#4	0.059	73.278	64.646	68.731
#5	0.047	79.860	73.858	76.958
#6	0.044	79.862	74.174	77.118
#7	0.039	80.965	75.443	78.263
#8	**0.039**	**81.530**	**76.194**	**79.026**
#9	0.037	81.404	75.891	78.714
#10	0.037	80.065	74.408	77.308
Avg	0.316	68.130	56.493	61.655

Table 7: Comparative results of each fold when ROUGE-1, ROUGE-2, and ROUGE-L are used

4.2.3 Experiment 3: Subsection heading extraction

This experiment aims to compare the performance of TopicFusion-MP with that of alternative approaches. We explore various combinations of the title and/or abstract of the source papers. Accordingly, the input text for this experiment consists of the concatenated content from the selected combination of title and/or abstract along with the generated related work text. Cosine similarity is employed to measure the semantic similarity between the identified subsection headings and the text generated by the BART algorithm in the second experiment. This evaluation approach is adopted because the documents are first clustered to determine the topic of each subsection, necessitating an alternative method to assess performance. We compare our proposed method with the following approaches:

1. RAKE [Rose et al., 2010] identifies the candidate keywords by splitting text into specific words or phrases and then ranks them on the basis of word frequency and co-occurrence metrics.
2. YAKE [Campos et al., 2020] selects and ranks candidate keywords on the basis of statistical text features, including capital terms, term position, term frequency, relatedness to the context, and different sentences;
3. TopicRank [Bougouin et al., 2013] clusters the candidate keywords into coherent topics using hierarchical clustering and ranks them on the basis of their importance in a word co-occurrence graph.
4. SingleRank [Wan and Xiao, 2008] is an unsupervised keyword extraction method in which a word graph is constructed on the basis of word co-occurrences and the PageRank algorithm.
5. LDA [Blei et al., 2003] computes the word distribution for each topic and extracts the words with the highest probability weights.
6. TextRank [Mihalcea and Tarau, 2004] is an unsupervised method used for keyword extraction using a graph-based ranking model.

Table 8 presents the results of using various topic extraction methods to generate subsection headings. Statistical methods such as RAKE, YAKE, and LDA exhibit limitations in identifying topics within the context of complex scientific documents, yielding similarity scores ranging from 0.20 to 0.28. Graph-based methods, including TopicRank, SingleRank, and TextRank, perform moderately better, with average similarity scores ranging from 0.30 to 0.50, thus slightly outperforming statistical approaches. Contrary to these findings, studies such as [Campos et al., 2020] and [Shapira et al., 2021] reported that YAKE outperformed TopicRank, SingleRank, and TextRank when evaluated against ground truth topics in terms of the precision, recall, and F1-measure. However, our experiments demonstrate that, compared with that of statistical methods, graph-based methods yield greater similarity with the text context. Both LDA and TextRank are less suitable as standalone approaches for extracting subsection headings. TopicFusion-MP consistently outperforms all the other methods across all the scenarios, with significant improvements in similarity scores:

Method	Scenario-1 Source paper: none	Scenario-2 Source paper: Title	Scenario-3 Source paper: Abstract	Scenario-4 Source paper: Title, Abstract
RAKE [Rose et al., 2010]	0.2256	0.2387	0.2083	0.2031
YAKE [Campos et al., 2020]	0.3496	0.4662	0.3369	0.4591
TopicRank [Bougouin et al., 2013]	0.3431	0.3619	0.2803	0.2938
SingleRank [Wan and Xiao, 2008]	0.4903	0.4893	0.5015	0.4992
LDA [Blei et al., 2003]	0.2789	0.2551	0.2153	0.2468
TextRank [Mihalcea and Tarau, 2004]	0.4414	0.4263	0.4148	0.4093
TopicFusion-MP	0.6398	0.6277	0.5817	0.6003
Average	0.3955	**0.4093**	0.3626	0.3873

Table 8: Comparative results of TopicFusion-MP and prior methods

$\text{scenario-1} = 0.6398, \text{scenario-2} = 0.6277, \text{scenario-3} = 0.5817, \text{and scenario-4} = 0.6003.$

Among the four tested scenarios, $scenario - 2$ yielded the highest average similarity score, with a value of 0.4093. The title of the source paper provides concise yet informative content, enabling topic extraction methods to effectively identify relevant topics without introducing excessive noise. In contrast, abstracts offer a more detailed and comprehensive summary of the scientific article, but the increased lexical variety carries the risk of introducing additional linguistic noise [Tullu, 2019], which may reduce topic identification accuracy by including less relevant terms. Although the TopicFusion-MP score slightly decreased in $scenario - 2$ compared with that in $scenario - 1$, it still achieved a substantially higher similarity score (0.6277) than all the other methods did in the same scenario. This suggests that incorporating the title offers a net positive effect, contributing to more accurate topic identification than does relying solely on abstracts.

5 Discussion

An article from the NudtRwG dataset titled "*Word sense induction using lexical chain-based hypergraph model*" was considered [Qian et al., 2014]. Figure 4 presents a comparison between the original subsection headings and related work texts (Figure 4a) and the subsection headings and related work texts generated by our system (Figure 4b). Our LDA clustering method formed a single cluster on the basis of the abstracts of the referenced papers, which consequently led to the generation of the subsection

heading "*word sense induction*". The original subsection headings were "*word sense induction*" and "*lexical chain extraction*". This finding highlights a limitation of the LDA method in identifying fine-grained topics, particularly when the topics are semantically close and interrelated. In such cases, the resulting topic cluster may be too broad, causing important subtopics (in this case, *lexical chain extraction*) to be subsumed under a more general topic, namely, *word sense induction*.

2.1 **Word sense induction**

A number of diverse approaches to WSI have been proposed so far. Context features are often represented in a variety of forms such as co-occurrence of words within phrases (Pantel and Lin, 2002; Dorow and Widdows, 2003), parts of speeches (Purandare and Pedersen, 2004), and grammatical relations ....

2.2 **Lexical chain extraction**

The lexical chain method is an important technique in natural language processing. A lexical chain is a sequence of semantic related words in text and provides important clues about the text structure and topic. It has formed ....

There are mainly two approaches for lexical chain extraction. One focuses on the use of knowledge resources like WordNet (Hirst and St-Onge, 1998) or thesauri (Morris and Hirst, 1991) as background information..

(a)

Word sense induction

At the word sense induction task, a number of approaches have been proposed so far. One of the most popular approaches is based on lexical chain methods cite. In this method words are represented by lexical chains which are connected like graphs and the target word senses are induced from the context in a bootstrapping way. cite used 6 different senses of a target word in the context of its occurrence. Similarly, cite defined a notion of "topic" ...

(b)

Figure 4: (a) Original related work text of [Qian et al., 2014]; (b) related work text generated using SciRelate

The *Word Sense Induction (WSI)* subsection in Figure 4a discusses various WSI approaches mentioned in the referenced papers (i.e., context feature approaches, vector space model approaches, graph-based approaches, hypergraph approaches, and word sense induction evaluation). In contrast, the *Lexical Chain Extraction* subsection focuses on the role of lexical chain extraction in natural language processing, its applications, and the methods used for extracting lexical chains. The WSI subsection in Figure 4b, generated by our system, covers a broader range of topics, including several WSI and word sense disambiguation (WSD) approaches and related technological advancements. However, several minor topics unrelated to WSI and WSD also appear, such as clustering methods, instance detection in text, senses of target words, topic clustering, and the use of word embeddings to identify similar words—among others. The emergence of these unrelated minor topics indicates that BART tends to summarize papers individually without adequately considering the connections between them. As a result, the generated text lacks coherence, with weak logical flow between consecutive sentences. This issue is exacerbated by the limitations of the source text—in this study, only abstracts were used—which constrains the visibility of interpaper relationships. Prior studies have attempted to address this by extracting additional sections from articles beyond the abstract to better capture conceptual and knowledge connections, thereby enabling the generation of more coherent summaries. Moreover, the integration of additional methods—such as hierarchical encoding–decoding [Shen et al., 2023], discourse-aware summarization [Dong et al., 2021], or knowledge graphs [Qu, 2024, Wang et al., 2022]—into the BART summarization pipeline is crucial to ensure that only relevant information is processed. While we acknowledge these challenges, implementing such approaches requires considerable computational resources, which are beyond the scope of this study.

6 Conclusion and future work

This study presents SciRelate, an innovative framework aimed at facilitating the generation of *Related Work* sections in scientific articles using abstractive multidocument summarization while simultaneously identifying subsection headings using a topic fusion-based approach. The framework integrates document clustering, neural summarization, and keyword extraction into a unified pipeline.

Our experiments yielded several key findings. Initially, we demonstrated that LDA clustering outperforms AHC in generating topic-cohesive clusters that closely approximate the actual number of subsection headings in the NudtRwG corpus. Second, by refining the pretrained BART model, we achieved high ROUGE scores and demonstrated the efficacy of the model in generating coherent related work summaries. Third, we proposed TopicFusion-MP, a keyword extraction approach that integrates LDA, TextRank, and MPNET sentence embeddings and empirically outperforms existing statistical and graph-based methods across multiple experimental scenarios. The method achieved the highest cosine similarity scores, reflecting its superiority in capturing semantic relevance between subsection headings and the generated text.

While the proposed approach demonstrates promising performance, further improvements remain feasible. Future studies may explore the integration of more advanced architectures or external knowledge sources to increase semantic coherence and relevance. Additionally, expanding the source content beyond abstracts (e.g., full texts or citation contexts) could increase coherence and informativeness. Addressing these directions will further advance automated related work generation, bringing us closer to more structured and semantically aware scientific writing assistance.

Statements and Declarations

Conflict of interest The authors of this study declare no conflicts of interest.

Ethical approval Not applicable.

Acknowledgements The authors would like to acknowledge all those who provided valuable feedback and support during the course of this research.

Funding The research is based on work supported by the Taiwan Ministry of Science and Technology under Grant No. NSTC 110-2410-H-006-042-MY3 and NSTC 112-2410-H-006-015-MY2. We would like to thank partially for the research grant supported by the "Higher Education SPROUT Project" and "Center for Innovative FinTech Business Models" of National Cheng Kung University (NCKU), sponsored by the Ministry of Education, Taiwan.

Data availability The datasets used in this study are publicly available. The NudtRwG can be accessed at https://github.com/NudtRwG/NudtRwG-Dataset.

Authors contributions Army Justitia: Conceptualization, Methodology, Investigation, Experiments, Writing Manuscript. Hei-Chia Wang: Conceptualization, Methodology, Supervision, Proofreading. All authors read and approved the final manuscript.

References

[AbuRa'ed et al., 2020] AbuRa'ed, A., Saggion, H., Shvets, A., Bravo, À.: 'Automatic related work section generation: experiments in scientific document abstracting'; Scientometrics, 125, 3 (2020), 3159–3185. https://doi.org/10.1007/s11192-020-03630-2

[Aliyu et al., 2020] Aliyu, M. B., Iqbal, R., James, A., Nkantah, D.: 'SED: An Algorithm for Automatic Identification of Section and Subsection Headings in Text Documents'; IJCSI International Journal of Computer Science, 17, 6 (2020), 40–47. https://doi.org/10.5281/zenodo.4431057

[Ashqar and Mutlu, 2023] Ashqar, G., Mutlu, A.: 'A Comparative Assessment of Various Embeddings for Keyword Extraction'; In Proceeding of 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA). IEEE (2023), 1–6. https://doi.org/10.1109/HORA58378.2023.10156762

[Beltagy et al., 2020] Beltagy, I., Peters, M. E., Cohan, A.: 'Longformer: The Long-Document Transformer'; (2020). Retrieved from http://arxiv.org/abs/2004.05150

[Blei et al., 2003] Blei, D. M., Ng, A. Y., Jordan, M. I.: 'Latent dirichlet allocation'; Journal of Machine Learning Research, 3, Jan (2003), 993–1022.

[Bougouin et al., 2013] Bougouin, A., Boudin, F., Daille, B.: 'TopicRank: Graph-Based Topic Ranking for Keyphrase Extraction'; In International Joint Conference on Natural Language Processing (IJCNLP) (2013), 543–551.

[Campos et al., 2020] Campos, R., Mangaravite, V., Pasquali, A., Jorge, A., Nunes, C., Jatowt, A.: 'YAKE! Keyword extraction from single documents using multiple local features'; Information Sciences, 509 (2020), 257–289. https://doi.org/10.1016/j.ins.2019.09.013

[Chehreghani, 2021] Chehreghani, M. H.: 'Reliable Agglomerative Clustering'; In 2021 International Joint Conference on Neural Networks (IJCNN). IEEE (2021), 1–8. https://doi.org/10.1109/IJCNN52387.2021.9534228

[Chen and Zhuge, 2016] Chen, J., Zhuge, H.: 'Summarizing Related Work through Citations'; International Conference on Semantics, Knowledge and Grids (2016), 54–61. https://doi.org/10.1109/SKG.2016.016

[Chen and Zhuge, 2019] Chen, J., Zhuge, H.: 'Automatic generation of related work through summarizing citations'; Concurrency and Computation: Practice and Experience, 31, 3 (2019), 1–12. https://doi.org/10.1002/cpe.4261

[Chen et al., 2022] Chen, X., Alamro, H., Li, M., Gao, S., Yan, R., Gao, X., Zhang, X.: 'Target-aware Abstractive Related Work Generation with Contrastive Learning'; In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (2022), 373–383. https://doi.org/10.1145/3477495.3532065

[Chen et al., 2021] Chen, X., Alamro, H., Li, M., Gao, S., Zhang, X., Zhao, D., Yan, R.: 'Capturing relations between scientific papers: An abstractive model for related work section generation'; Proceeding of 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021) (2021), 6068–6077. https://doi.org/10.18653/v1/2021.acl-long.473

[Deng et al., 2021] Deng, Z., Zeng, Z., Gu, W., Ji, J., Hua, B.: 'Automatic related work section generation by sentence extraction and reordering'; In CEUR Workshop Proceedings (Vol. 2871) (2021), 101–110.

[Dong et al., 2021] Dong, Y., Mircea, A., Cheung, J. C. K.: 'Discourse-aware unsupervised

summarization of long scientific documents'; In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics (2021), 1089–1102. https://doi.org/10.18653/v1/2021.eacl-main.93

[Ge et al., 2021] Ge, Y., Dinh, L., Liu, X., Su, J., Lu, Z., Wang, A., Diesner, J.: 'BACO: A background knowledge- And content-based framework for citing sentence generation'; In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (2021), 1466–1478. https://doi.org/10.18653/v1/2021.acl-long.116

[Gu and Hahnloser, 2022] Gu, N., Hahnloser, R. H. R.: 'Controllable Citation Text Generation'; In Proceedings of the Fourth Workshop on Scholarly Document Processing (SDP 2024). Bangkok, Thailand: Association for Computational Linguistics (2022). Retrieved from https://aclanthology.org/2024.sdp-1/

[Gulati et al., 2023] Gulati, V., Kumar, D., Popescu, D. E., Hemanth, J. D.: 'Extractive Article Summarization Using Integrated TextRank and BM25+ Algorithm'; Electronics, 12, 2 (2023), 372. https://doi.org/10.3390/electronics12020372

[Guo et al., 2021a] Guo, C., Lu, M., Wei, W.: 'An Improved LDA Topic Modeling Method Based on Partition for Medium and Long Texts'; Annals of Data Science, 8, 2 (2021a), 331–344. https://doi.org/10.1007/s40745-019-00218-3

[Guo et al., 2021b] Guo, M., Ainslie, J., Uthus, D.: 'LongT5: Efficient Text-To-Text Transformer for Long Sequences'; (2021b).

[Harandizadeh et al., 2022] Harandizadeh, B., Priniski, J. H., Morstatter, F.: 'Keyword Assisted Embedded Topic Model'; In Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining (2022), 372–380. https://doi.org/10.1145/3488560.3498518

[He et al., 2016] He, L., Li, W., Zhuge, H.: 'Exploring differential topic models for comparative summarization of scientific papers'; In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers (2016), 1028–1038.

[Hoang and Kan, 2010] Hoang, C. D. V., Kan, M. Y.: 'Towards automated related work summarization'; In Coling 2010: Posters (Vol. 2). Beijing, China (2010), 427–435.

[Hu and Wan, 2014] Hu, Y., Wan, X.: 'Automatic generation of related work sections in scientific papers: An optimization approach'; In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: Association for Computational Linguistics (ACL) (2014), 1624–1633. https://doi.org/10.3115/v1/d14-1170

[Jung et al., 2022] Jung, S. Y., Lin, T. H., Liao, C. H., Yuan, S. M., Sun, C. T.: 'Intent-Controllable Citation Text Generation'; Mathematics, 10, 1763 (2022). https://doi.org/10.3390/math10101763

[Justitia and Wang, 2025] Justitia, A., Wang, H.-C.: 'Investigating Longformer-BART for the Paragraph-Based Generation of Abstractive Multidocument Related Work Summaries in Scientific Articles'; Manuscript under Review (2025).

[Khan et al., 2022] Khan, M. Q., Shahid, A., Uddin, M. I., Roman, M., Alharbi, A., Alosaimi, W., et al.: 'Impact analysis of keyword extraction using contextual word embedding'; PeerJ Computer Science, 8 (2022), 1–16. https://doi.org/10.7717/peerj-cs.967

[La Quatra and Cagliero, 2023] La Quatra, M., Cagliero, L.: 'BART-IT: An Efficient Sequence-to-Sequence Model for Italian Text Summarization'; Future Internet, 15, 1 (2023). https://doi.org/10.3390/fi15010015

[Lemarié et al., 2012] Lemarié, J., Jr, R. F. L., Péry-Woodley, M.-P.: 'Understanding How

Headings Influence Text Processing'; Discours. Revue de Linguistique, Psycholinguistique et Informatique. A Journal of Linguistics, Psycholinguistics and Computational Linguistics, , July 2012 (2012). https://doi.org/10.4000/discours.8600

[Lewis et al., 2020] Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., et al.: 'BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension'; In Proceedings of the 58th annual meeting of the association for computational linguistics. Stroudsburg, USA: Association for Computational Linguistics (2020), 7871–7880. https://doi.org/10.18653/v1/2020.acl-main.703.

[Li, 2021] Li, J.: 'A comparative study of keyword extraction algorithms for English texts'; Journal of Intelligent Systems, 30, 1 (2021), 808–815. https://doi.org/10.1515/jisys-2021-0040

[Li et al., 2024] Li, X., Lee, Y.-H., Ouyang, J.: 'Cited Text Spans for Scientific Citation Text Generation'; In Proceedings of the Fourth Workshop on Scholarly Document Processing (SDP 2024). Bangkok, Thailand: Association for Computational Linguistics (2024), 90–104. Retrieved from https://aclanthology.org/2024.sdp-1.9

[Li et al., 2022] Li, X., Mandal, B., Ouyang, J.: 'CORWA: A Citation-Oriented Related Work Annotation Dataset'; In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Seattle, United States: Association for Computational Linguistics (2022), 5426–5440. https://doi.org/10.18653/v1/2022.naacl-main.397

[Li and Ouyang, 2025] Li, X., Ouyang, J.: 'Explaining Relationships Among Research Papers'; In Proceedings of the 31st International Conference on Computational Linguistics. Abu Dhabi, UAE: Association for Computational Linguistics (2025), 1080–1105. Retrieved from https://aclanthology.org/2025.coling-main.73/

[Liu et al., 2023] Liu, J., Zhang, Q., Shi, C., Naseem, U., Wang, S., Hu, L., Tsang, I.: 'Causal Intervention for Abstractive Related Work Generation'; In Findings of the Association for Computational Linguistics: EMNLP 2023. Stroudsburg, PA, USA: Association for Computational Linguistics (2023), 2148–2159. https://doi.org/10.18653/v1/2023.findings-emnlp.141

[Luu et al., 2021] Luu, K., Wu, X., Koncel-Kedziorski, R., Lo, K., Cachola, I., Smith, N. A.: 'Explaining relationships between scientific documents'; Proceedings of the Conference 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Proceedings of the Conference (ACL-IJCNLP 2021) (2021), 2130–2144. https://doi.org/10.18653/v1/2021.acl-long.166

[Mandal et al., 2024] Mandal, B., Li, X., Ouyang, J.: 'Contextualizing Generated Citation Texts'; In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). Torino, Italia: ELRA and ICCL (2024), 3849–3854. Retrieved from https://aclanthology.org/2024.lrec-main.340/

[Martin-boyle et al., 2009] Martin-boyle, A., Tyagi, A., Hearst, M. A., Kang, D.: 'Shallow Synthesis of Knowledge in GPT-Generated Texts : A Case Study in Automatic Related Work Composition'; (2009).

[Maslim et al., 2024] Maslim, M., Wang, H. C., Putra, C. D., Prabowo, Y. D.: 'A Trustworthy Automated Short-Answer Scoring System Using a New Dataset and Hybrid Transfer Learning Method'; International Journal of Interactive Multimedia and Artificial Intelligence, 8, 5 (2024), 37–45. https://doi.org/10.9781/ijimai.2024.02.003

[Mihalcea and Tarau, 2004] Mihalcea, R., Tarau, P.: 'TextRank: Bringing Order into Texts'; In Proceeding of the 2004 Conference on Empirical Methods in Natural Language Processing (2004), 404–411. https://doi.org/10.1016/0305-0491(73)90144-2

[Monreal et al., 2004] Monreal, C. S., Salom, L. G., Olivares, M. C.: 'What Do Section Headings and Subheadings Tell The Reader? A Study of the Macrostructure of Research Articles in Three Engineering Disciplines'; 25 Años de Lingüística En España. Hitos y Retos (2004), 595–603.

[Qian et al., 2014] Qian, T., Ji, D., Zhang, M., Teng, C., Xia, C.: 'Word sense induction using lexical chain based hypergraph model'; COLING 2014 - 25th International Conference on Computational Linguistics, Proceedings of COLING 2014: Technical Papers, 1, 1 (2014), 1601–1611.

[Qu, 2024] Qu, Y.: 'Leveraging Knowledge-aware Methodologies for Multi-document Summarization'; In Companion Proceedings of the ACM Web Conference (2024), 1206–1209. https://doi.org/10.1145/3589335.3651262

[Rashid et al., 2019] Rashid, J., Shah, S. M. A., Irtaza, A.: 'Fuzzy topic modeling approach for text mining over short text'; Information Processing and Management, 56, 6 (2019), 102060. https://doi.org/10.1016/j.ipm.2019.102060

[Reimers and Gurevych, 2019] Reimers, N., Gurevych, I.: 'Sentence-BERT: Sentence embeddings using siamese BERT-networks'; In Proceedings of the Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, (2019), 3982–3992. https://doi.org/10.18653/v1/d19-1410

[Ridwan and Retnawati, 2021] Ridwan, M. R., Retnawati, H.: 'Application of Cluster Analysis Using Agglomerative Method'; Numerical: Jurnal Matematika Dan Pendidikan Matematika, 5, 1 (2021), 33–48. https://doi.org/10.25217/numerical.v5i1.1538

[Rishu and Kukreja, 2024] Rishu, Kukreja, V.: 'Comic exploration and Insights: Recent trends in LDA-Based recognition studies'; Expert Systems with Applications, 255, July (2024), 124732. https://doi.org/10.1016/j.eswa.2024.124732

[Röder et al., 2015] Röder, M., Both, A., Hinneburg, A.: 'Exploring the Space of Topic Coherence Measures'; In Proceeding of 8th ACM international conference on Web search and data mining (2015), 399–408.

[Rose et al., 2010] Rose, S., Engel, D., Cramer, N., Cowley, W.: 'Automatic Keyword Extraction from Individual Documents'; In Text Mining. Wiley (2010), 1–20. https://doi.org/10.1002/9780470689646.ch1

[Shapira et al., 2021] Shapira, O., Pasunuru, R., Dagan, I., Amsterdamer, Y.: 'Multi-Document Keyphrase Extraction: Dataset, Baselines and Review'; (2021).

[Shen et al., 2023] Shen, C., Cheng, L., Nguyen, X. P., You, Y., Bing, L.: 'A Hierarchical Encoding-Decoding Scheme for Abstractive Multi-document Summarization'; Findings of the Association for Computational Linguistics: EMNLP 2023 (2023), 5872–5887. https://doi.org/10.18653/v1/2023.findings-emnlp.391

[Song et al., 2020] Song, K., Tan, X., Qin, T., Lu, J., Liu, T.: 'MPNet : Masked and Permuted Pre-training for Language Understanding'; In Proceeding of 34th Conference on Neural Information Processing Systems. Vancouver, Canada (2020), 1–14.

[Song et al., 2009] Song, Y., Pan, S., Liu, S., Zhou, M. X., Qian, W.: 'Topic and keyword re-ranking for LDA-based topic modeling'; Proceedings of the 18th ACM Conference on Information and Knowledge Management (2009), 1757–1760. https://doi.org/10.1145/1645953.1646223

[Tullu, 2019] Tullu, M.: 'Writing the title and abstract for a research paper: Being concise, precise, and meticulous is the key'; Saudi Journal of Anaesthesia, 13, 5 (2019), S12–S17. https://doi.org/10.4103/sja.SJA_685_18

[Vijaya et al., 2019] Vijaya, Sharma, S., Batra, N.: 'Comparative Study of Single Linkage, Complete Linkage, and Ward Method of Agglomerative Clustering'; In Proceeding of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects, COMITCon 2019. IEEE (2019), 568–573. https://doi.org/10.1109/COMITCon.2019.8862232

[Wan and Xiao, 2008] Wan, X., Xiao, J.: 'Single Document Keyphrase Extraction using Neighborhood Knowledge'; Proceedings of the National Conference on Artificial Intelligence, 2 (2008), 855–860.

[Wang et al., 2022] Wang, G., Li, W., Lai, E., Jiang, J.: 'KATSum: Knowledge-aware Abstractive Text Summarization';, , 2 (2022). Retrieved from http://arxiv.org/abs/2212.03371

[Wang et al., 2020a] Wang, H. C., Hsiao, W. C., Chang, S. H.: 'Automatic paper writing based on a RNN and the TextRank algorithm'; Applied Soft Computing Journal, 97 (2020a), 106767. https://doi.org/10.1016/j.asoc.2020.106767

[Wang et al., 2020b] Wang, P. C., Li, S. S., Zhou, H. F., Tang, J. T., Wang, T.: 'ToC-RWG: Explore the Combination of Topic Model and Citation Information for Automatic Related Work Generation'; IEEE ACCESS, 8 (2020b), 13043–13055. https://doi.org/10.1109/ACCESS.2019.2959056

[Wang et al., 2019] Wang, P., Li, S., Zhou, H., Tang, J., Wang, T.: 'An Analytical Study on a Benchmark Corpus Constructed for Related Work Generation'; In Natural Language Processing and Chinese Computing: 8th CCF International Conference (Vol. NLPCC 2019). Dunhuang, China: Springer International Publishing (2019), 421–432. https://doi.org/10.1007/978-3-030-32233-5_33

[Wang, 2024] Wang, Y.: 'Research on the TF–IDF algorithm combined with semantics for automatic extraction of keywords from network news texts'; Journal of Intelligent Systems, 33, 1 (2024). https://doi.org/10.1515/jisys-2023-0300

[Wang et al., 2020c] Wang, Y., Liu, X., Gao, Z.: 'Neural related work summarization with a joint context-driven attention mechanism'; In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018 (2020c), 1776–1786. https://doi.org/10.18653/v1/d18-1204

[Xiao et al., 2022] Xiao, W., Beltagy, I., Carenini, G., Cohan, A.: 'PRIMERA: Pyramid-based Masked Sentence Pre-training for Multi-document Summarization'; In Proceedings of the Annual Meeting of the Association for Computational Linguistics (Vol. 1) (2022), 5245–5263. https://doi.org/10.18653/v1/2022.acl-long.360

[Xing et al., 2020] Xing, X., Fan, X., Wan, X.: 'Automatic Generation of Citation Texts in Scholarly Papers: A Pilot Study'; In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics (2020), 6181–6190. https://doi.org/10.18653/v1/2020.acl-main.550

[Zaheer et al., 2020] Zaheer, M., Guruganesh, G., Dubey, A., Ainslie, J., Alberti, C., Ontanon, S., et al.: 'Big bird: Transformers for longer sequences'; Advances in Neural Information Processing Systems, 2020-Decem, NeurIPS (2020).

[Zhang et al., 2020] Zhang, J., Zhao, Y., Saleh, M., Liu, P. J.: 'PEGASUS: Pre-Training with extracted gap-sentences for abstractive summarization'; In Proceedings of the International Conference on Machine Learning (Vol. 119). PLMR (2020), 11328–11339.

[Zhang et al., 2024] Zhang, L., Wang, W., Ma, J., Wen, Y.: 'IWF-TextRank Keyword Extraction Algorithm Modelling'; Applied Sciences, 14, 22 (2024), 10657. https://doi.org/10.3390/app142210657

[Zhuohao et al., 2021] Zhuohao, W., Dong, W., Qing, L.: 'Keyword Extraction from Scientific Research Projects Based on SRP-TF-IDF'; Chinese Journal of Electronics, 30, 4 (2021), 652–657. https://doi.org/10.1049/cje.2021.05.007