

## Text Summarization: A Bibliometric Study and Systematic Literature Review

Meiling Xu<sup>1, 2</sup> , Hayati Abd Rahman<sup>1\*</sup> , Feng Li<sup>1, 2</sup> 

<sup>1</sup> College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Shah Alam 40450, Malaysia

<sup>2</sup> College of Computer and Information Engineering, Hebei Finance University, Baoding 071051, China

Corresponding Author Email: [hayatiar@fskm.uitm.edu.my](mailto:hayatiar@fskm.uitm.edu.my)

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.290538>

### ABSTRACT

**Received:** 5 July 2024

**Revised:** 9 September 2024

**Accepted:** 1 October 2024

**Available online:** 24 October 2024

#### Keywords:

*text summarization; extractive summarization; abstractive summarization; bibliometric analysis*

With the exponential growth of text data from many sources such as internet content, news stories, scientific literature, and legal documents, the significance of automatic text summarization (ATS) technology is expanding in the field of natural language processing (NLP). Conventional manual techniques for summarizing text are both time-consuming and labor-intensive. Moreover, they face challenges in terms of economic viability and feasibility when applied to enormous amounts of data. ATS technology has undergone significant advancements since the 1950s, transitioning from basic rule-based methods to sophisticated systems that utilize machine learning, neural networks, and deep learning. Both extractive and abstractive summarization techniques have achieved substantial gains in this development, leading to progress in text processing technologies. Nevertheless, despite ongoing advancements in ATS, the existing body of literature is still deficient in comprehensive evaluations of its most recent research advancements and forthcoming trends. This deficiency impedes a more profound comprehension of scholarly research and real-world implementations. In order to tackle this issue, this work utilizes bibliometric and content analytic techniques to carry out a thorough systematic evaluation of important literature in the present field of text summarization research, with the goal of delineating contemporary research patterns and technological progress.

## 1. INTRODUCTION

In the 21st century, with the rapid expansion of the internet, the volume of text data has grown exponentially, transforming the internet into a vast repository of information encompassing articles, novels, books, scientific research, and more. In this context, automatic text summarization technology has emerged as a crucial tool for handling large-scale text resources. It facilitates the effective extraction of crucial data from lengthy texts, reducing the necessity to read complete papers separately and greatly streamlining the process of retrieving and obtaining information.

Luhn [1] was one of the pioneering researchers in automatic text summarization, defining a summary as "the process of extracting the most important information from a document." According to his definition, the core of text summarization is to identify and extract key information from large volumes of text. Text summarization is a procedure that condenses extensive documents into brief representations, with the goal of reducing length and conveying the text while maintaining the essential information. A proficient summary should succinctly capture the primary substance of the paper while substantially decreasing its length, furnishing readers with a brief and all-encompassing synopsis [2]. Radev et al. [3] defines a summary as a concise rendition of one or more source texts, with the primary objective of transmitting crucial information from the original content. A typical summary is

usually no longer than half the length of the original material. Gupta and Lehal [4] provided a definition of text summarization as the act of condensing the original text while still maintaining important information. They stressed that the primary objective of text summarization is to preserve the most crucial information by means of compression. Automatic text summarization technology is extensively employed in diverse fields such as news, emails, novels, legal documents, and medical literature. Its purpose is to extract and condense essential information, enabling efficient retrieval and processing of vast amounts of data [5]. Automatic text summarization technology is extensively employed in diverse fields such as news, emails, novels, legal papers, and medical literature. Its purpose is to extract and condense essential information, enabling efficient retrieval and processing of vast amounts of data.

Extractive text summarization technology has made substantial progress since its debut. Preliminary research conducted by Nazari and Mahdavi [6] mostly examined conventional extraction techniques that relied on term frequency and sentence position. In contrast, Kirmani et al. [7] incorporated word embedding models to improve the precision of sentence scoring. Yadav and Chatterjee [8] enhanced extractive summarization methods by utilising graph-based strategies to optimize the representation of connections between sentences. Utilizing deep learning technologies, Suleiman et al. [9] employed these methods for

extractive summarization, resulting in a notable enhancement in summarization effectiveness. On the other hand, research has also explored certain domains. Kanapala et al. [10] specifically examined domain-specific summarization strategies for legal documents, whereas Dutta et al. [11] assessed and contrasted several extraction algorithms employed in microblog summarization. In their study, Jacquenet et al. [12] investigated abstractive deep-learning techniques and discussed the difficulties related to summarization. Mahajani et al. [13] have conducted reviews on both extractive and abstractive summarization approaches. Together, these works have enhanced the theoretical comprehension of extractive text summarization and expanded its practical applications and efficacy.

In the field of abstractive summarization technology, initial research primarily focused on models driven by neural networks, as demonstrated by the studies of Lin and Ng [14] and Tandel et al. [15]. More recent inquiries have further explored abstractive summarization techniques utilizing deep learning. The works of Suleiman and Awajan [16], Zhao et al. [17], and Dong et al. [18] are notable in this field. Additionally, several studies, such as those by Jacquenet et al. [12] and El-Kassas et al. [5], delve into deep learning-based abstractive summarization techniques and explore the challenges associated with them.

In recent years, numerous comprehensive surveys have evaluated traditional abstractive and extractive methods. Notable examples include the works of Gupta and Gupta [19], and Wibawa and Kurniawan [20]. Furthermore, the application of deep learning in abstractive and extractive text summarization has been extensively discussed by Saiyyad and Patil [21] and Saleh et al. [22].

This study aims to systematically review the literature on text summarization and conduct bibliometric and content analysis using the CiteSpace tool. The main research questions addressed in this paper are:

- (1) What are the trends in the number of publications in this field?
- (2) Which authors, publications, and disciplines are prominent in the current research landscape?
- (3) What are the main research hotspots in the field?
- (4) What methods are employed in text summarization technologies?
- (5) What are the advantages and disadvantages of these methods?

The paper is organized as follows: The introduction is followed by a section that explains the methods used for data collection and research. The next section conducts a bibliometric analysis of the literature's characteristics. The fourth section analyses the content of the literature. Finally, the fifth section summarizes the research findings and provides an overview of future research directions.

## 2. DATA COLLECTION AND RESEARCH METHODS

### 2.1 Data collection

The present study utilized the Scopus database as the primary literature retrieval platform. The detailed process of literature search and selection is outlined in Table 1. To ensure the scientific rigor and quality of the selected literature, this study exclusively focused on English-language journal articles indexed in Scopus, excluding conference papers, books, and

monographs. Using the search query ALL=("extractive" OR "abstractive") AND ("text summarization") in the Scopus core collection, a total of 2,061 documents were retrieved. After applying stringent selection criteria, 615 journal articles were identified for inclusion.

**Table 1.** Literature screening process

Literature Search	Scopus
Last retrieval time	Aug 2024
Search topics	("extractive" OR "abstractive") AND ("text summarization")
Number of searches	2061
Document type	Article
Source type	Journal
Final number of literatures	615

### 2.2 Research methods

This study combines bibliometric and content analysis methodologies to systematically explore text summarization technologies and their future development trends. Bibliometrics utilizes mathematical statistics and computational analysis methods to quantitatively evaluate the external attributes of literature, thereby uncovering the structure of knowledge and the dynamics of development within the discipline. Bibliometrics offers a thorough and unbiased analysis of the development and research trends of topics using visualisation methods like publication counts, core author distribution, and keyword grouping. Nevertheless, although this approach successfully emphasizes broad research patterns, it cannot thoroughly analyze the precise details of the literature.

Content analysis, as a qualitative review method, involves systematically organizing, synthesizing, and summarizing the core content of the literature to conduct a thorough examination of the research status. This approach authentically reflects the intrinsic characteristics of the research topic, offering detailed insights into the current state of deep learning-based text summarization techniques. Nevertheless, content analysis has limitations in sample selection and is somewhat subjective. This study uses a mixed research strategy that combines quantitative and qualitative approaches in order to combine the strengths of both approaches. On the one hand, bibliometrics is used to obtain a thorough understanding of research trends in text summarization technologies, while on the other hand, content analysis is used to delve deeper into the field's main issues and future directions. In addition to facilitating a comprehensive grasp of the state of the topic today, this combined research approach offers significant insights for further study and advancement.

## 3. BIBLIOMETRIC ANALYSIS OF LITERATURE

### 3.1 Analysis of publication years

The number of publications is often used as a key indicator of the academic community's interest in a particular research area. By analyzing the temporal distribution of these publications, we can understand changes in research interest over time. Figure 1 illustrates the evolving trend of research interest in the field of text summarization. The data shows a

significant increase in both the volume of literature and citation frequency in this field since 2011. Specifically, the number of publications rose from 12 in 2017 to 136 in 2023, reflecting a year-on-year increase. This growth trend indicates a heightened research activity in the field.

Detailed analysis reveals that prior to 2014, there were relatively few relevant publications. However, from 2015 to 2022, the number of publications increased annually, reaching 136 in 2023. By mid-2024, the number of related publications had already reached 79. These figures not only indicate that research in text summarization continues to maintain high relevance but also demonstrate a notable increase in research interest. Overall, this growth trend reflects the academic community's sustained attention to text summarization technologies and highlights the broad demand for these technologies in practical applications. With the increasing volume of data and advancements in technology, research in text summarization is becoming increasingly active and shows promising prospects for the future.

### 3.2 Analysis of publication fields

Analysis of the 615 articles reveals that research on text summarization is widely applied across multiple disciplines. As shown in Figure 2, publications in the field of computer science account for 40.2% of the total, highlighting its

dominant role in deep learning-based text summarization research.

The following is a decline in the use of deep learning for text summarization across different domains: Chemical Engineering, Energy, Biochemistry, Genetics and Molecular Biology, Medicine, Chemistry, Environmental Science, Neuroscience, Earth and Planetary Sciences, Agricultural and Biological Sciences, Economics, Econometrics and Finance, Health Professions, Business, Management and Accounting, Physics and Astronomy, Multidisciplinary, Computer Science, Engineering, Mathematics, Social Sciences, Materials Science, Decision Sciences, Business, Management and Accounting, and Pharmacology and Pharmaceutics.

### 3.3 Journal analysis of publications

Figure 3 presents the annual publication counts for the top ten sources in the field of text summarization. Analysis of this figure reveals that the top five journals in this domain are Expert Systems with Applications (33 articles), IEEE Access (31 articles), Multimedia Tools and Applications (20 articles), Information Processing and Management (17 articles), and Knowledge-Based Systems (17 articles). Detailed information regarding the Core Area Journal Name, Research Area, Number of Publications, and Impact Factor for 2023-2024 is provided in Table 2.

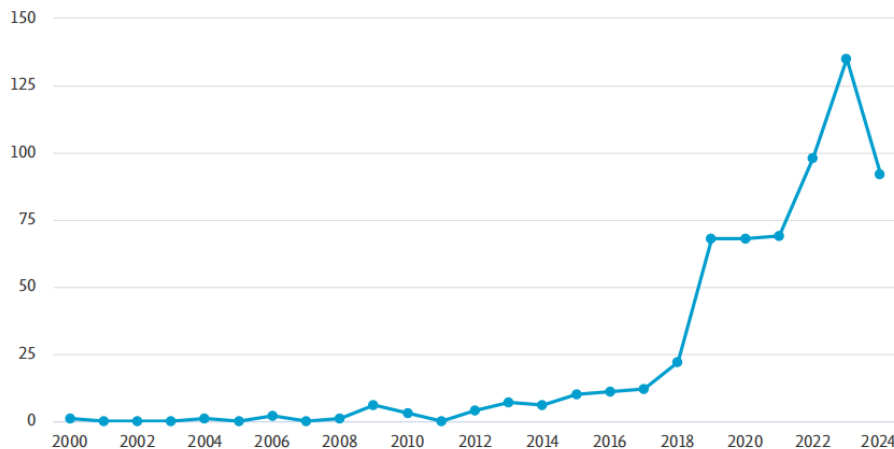


Figure 1. Number of Articles from 2000 to August 2024

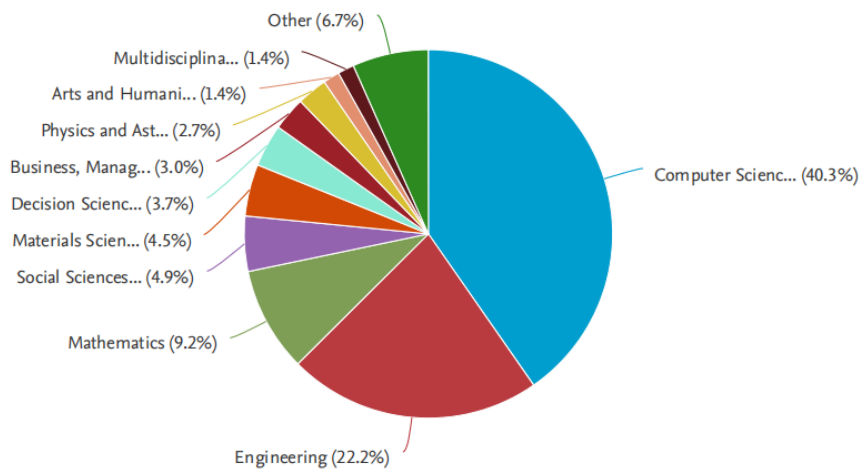


Figure 2. Documents categorized by subject area

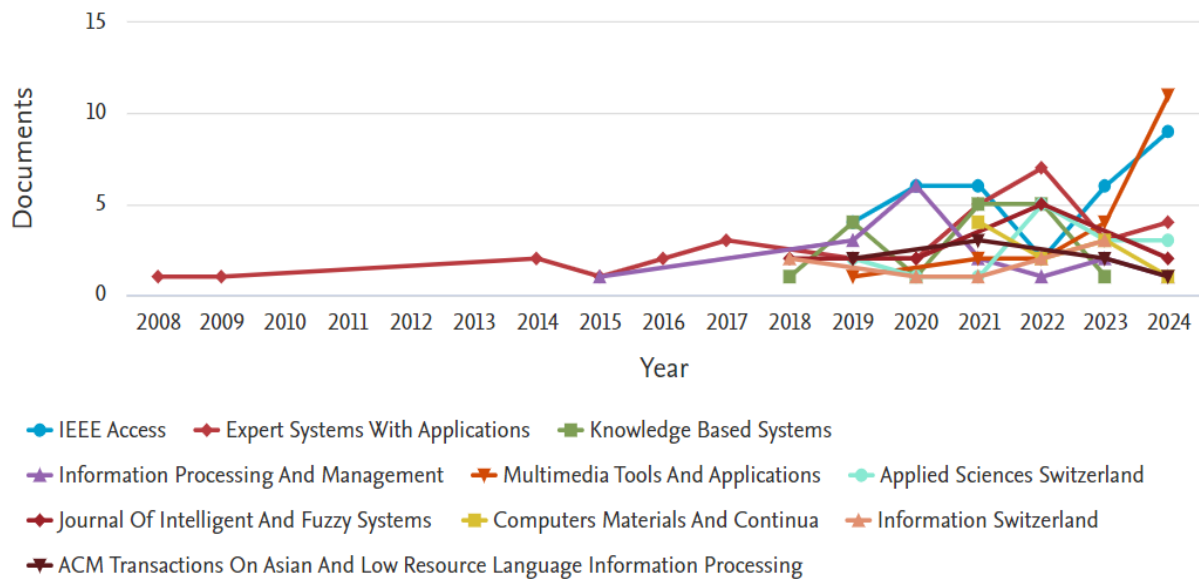


Figure 3. Distribution of journal publications filtered by the top ten sources

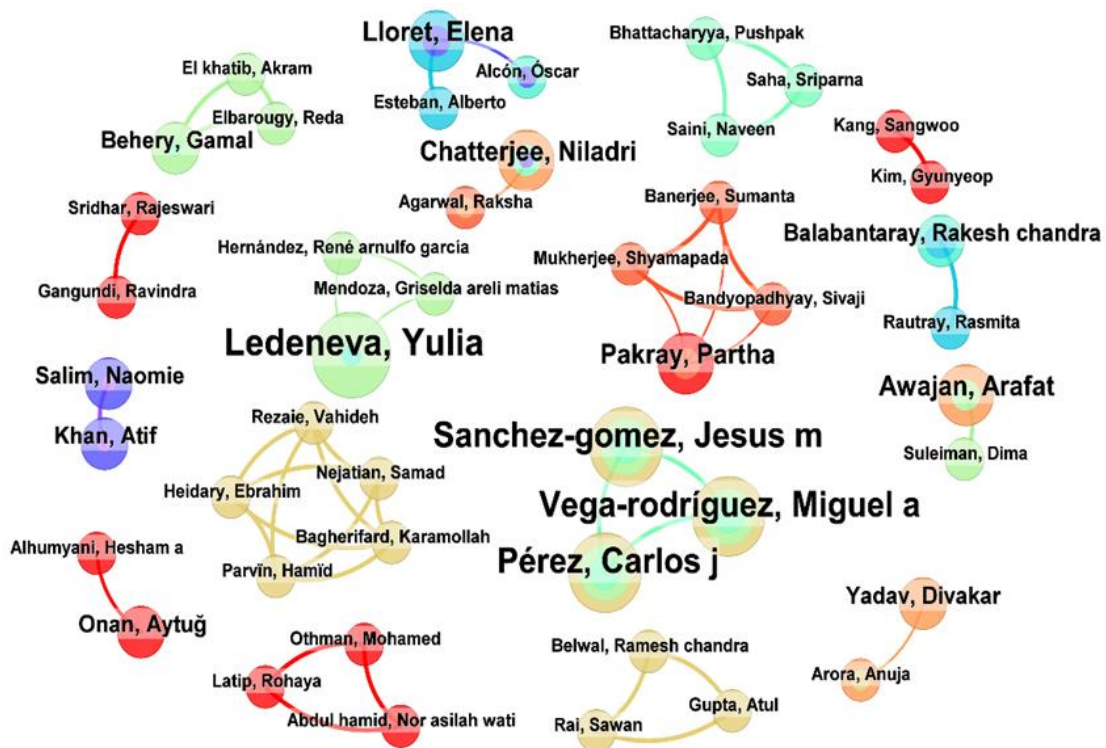


Figure 4. Author co-occurrence analysis

Table 2. Journal publications

No.	Core Area Journal Name	Research Area	Number of Publications	Impact Factor 2023-2024
1	Expert Systems with Applications	Artificial Intelligence	33	6.4
2	IEEE Access	Electrical engineering computer science	31	3.5
3	Multimedia Tools and Applications	Multimedia systems	20	4.2
4	Information Processing and Management	Information retrieval	17	3.6
5	Knowledge Based Systems	Knowledge management	17	5.4
6	Applied Sciences Switzerland	sciences across multiple disciplines	15	2.6
7	Journal Of Intelligent and Fuzzy Systems	Intelligent systems	14	3.1
8	Computers Materials and Continua	Environment and Sustainability	9	4.5
9	Information Switzerland	Informatics, information systems	9	2.2
10	ACM Transactions on Asian and Low Resource Language Information Processing	Natural language processing, computational linguistics	8	1.8

### 3.4 Author co-occurrence analysis

Using tools such as CiteSpace, this study performed an author co-occurrence analysis of 615 publications in the field of text summarization research. CiteSpace software facilitates the identification of collaboration networks among cited authors and enables co-citation analysis. By examining author centrality and citation frequency, we can pinpoint high-impact authors and their key publications in the field, as well as reveal the co-citation relationships among authors and the similarity and interconnectedness of research topics.

According to the co-citation analysis data (see Figure 4), notable authors with a high volume of publications include Yulia Ledeneva, Partha Pakray, Atif Khan, and Rafael Ferreira. These authors exhibit high centrality within the collaboration network, indicating their significant influence and contributions to the field of text summarization. The structure of co-citation relationships also reflects the collaboration patterns among authors and the hierarchical nature of research topics, further confirming the crucial role of these highly cited authors in advancing the field.

## 4. CONTENT-BASED BIBLIOMETRIC ANALYSIS

### 4.1 Keyword co-occurrence analysis

Keywords serve as a concise summary of the content of a document, representing the main topics discussed within the literature and can be used to analyze the themes and hotspots of research. After processing, this study generated a co-occurrence map of keywords in English-language documents and identified the terms with the strongest co-occurrence intensity. The research then analyzed the directions and developmental trends in text summarization studies. Figure 5 illustrates the co-occurrence intensity of keywords in the English literature, where high-frequency keywords are represented at various nodes. The size of each node reflects the

co-occurrence intensity of the corresponding keyword, and the keywords are connected by lines indicating their relationships. Using CiteSpace software, this study conducted a keyword clustering analysis to determine the frequency and centrality of various keywords within the retrieved literature. After removing duplicate content, the top ten English terms ranked by co-occurrence strength include: text processing, extractive summarizations, deep learning, text summarization, natural language processing, graphic methods, semantics, information retrieval, abstractive text summarization, and artificial intelligence.

The results of the keyword co-occurrence analysis shed light on the primary research directions and technological focal points in the field of text summarization. According to the top ten phrases, academics are mostly interested in cutting-edge technologies that are essential to text processing and summary creation, such as deep learning, natural language processing, and information retrieval. Since "deep learning" and "natural language processing" are used so frequently, it is clear that current research focuses on utilizing these technologies to improve the caliber and efficacy of text summarization.

Moreover, the presence of both "extractive summarization" and "abstractive text summarization" highlights ongoing efforts by researchers to explore different summarization strategies, including extracting key information from the original text and generating new forms of expression. The appearance of terms like "text processing" and "graphic methods" further suggests that foundational text processing techniques and graphical methods are continuously evolving and being applied. This keyword co-occurrence also reflects a growing focus on semantics and artificial intelligence, indicating a deepening need for understanding and processing text at more profound levels. Overall, these keywords reveal the diversity and complexity within the text summarization field, illustrating the broad scope of technological innovation and methodological exploration in current research.

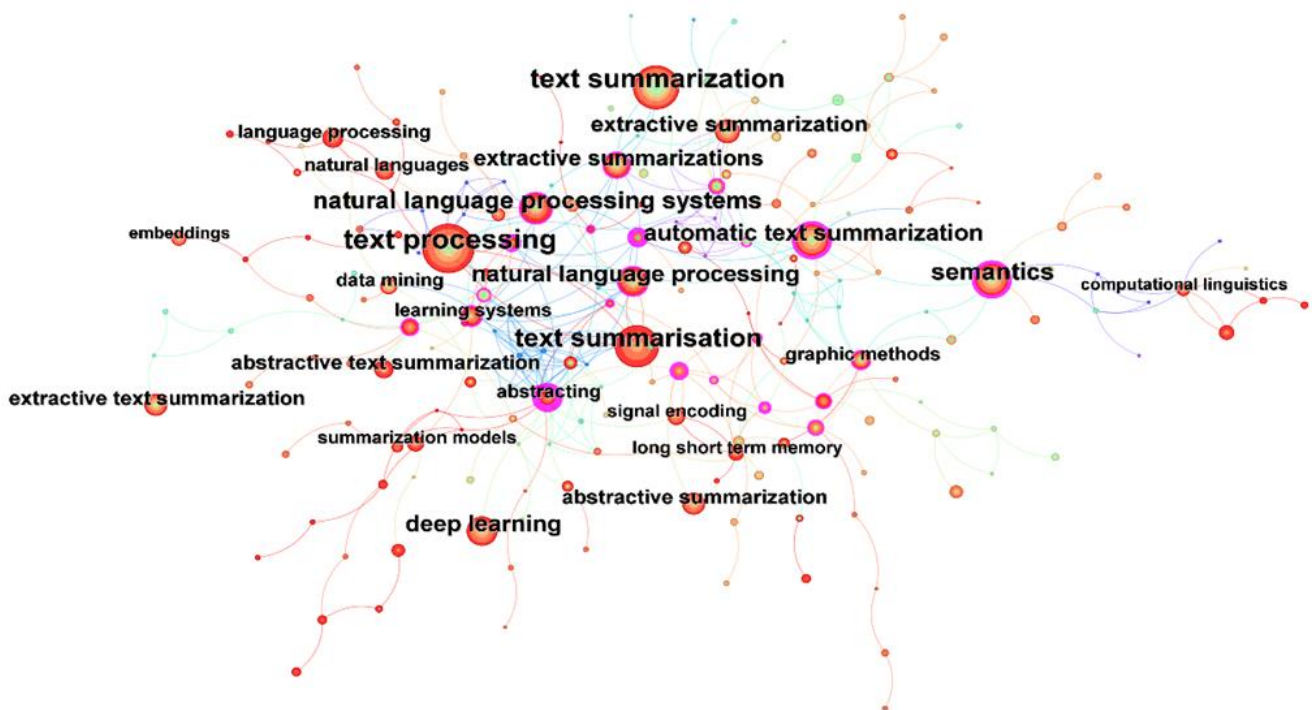
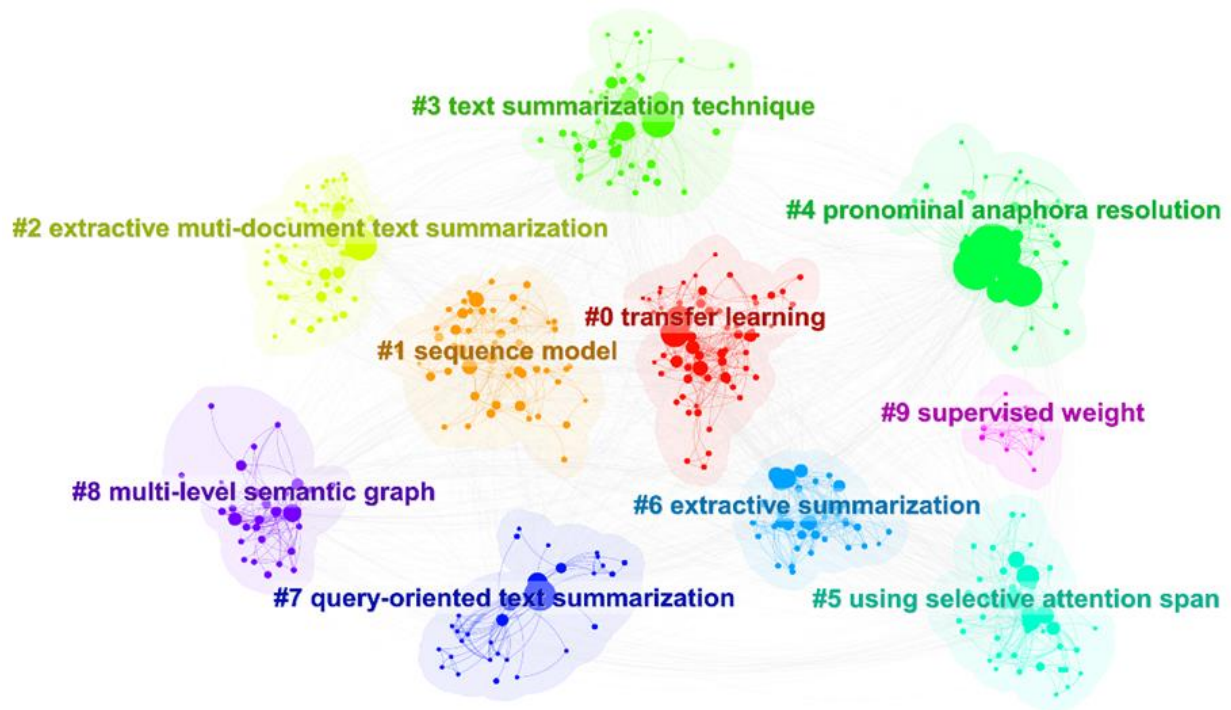


Figure 5. Keyword co-occurrence analysis



**Figure 6.** Keywords clustering analysis

## 4.2 Keyword clustering analysis

The clustering analysis of reference keywords categorized them into 10 distinct groups (Figure 6). This analysis reveals the research trends and hotspots within the field of text summarization, aiding in the identification of currently prominent techniques and methods such as transfer learning, sequence models, and extractive summarization. Additionally, it highlights unresolved challenges and issues in the research, including pronoun coreference resolution and selective attention span. The clustering analysis also helps to uncover intersections between technologies and applications, such as the integration of query-focused text summarization with multi-layered semantic graphs, as well as academic contributions reflected in the literature, like the application of supervised weights. Overall, this analysis provides a systematic understanding of the text summarization field, uncovering key research directions and technological applications, and offering valuable insights for further research planning and resource allocation.

## 4.3 Further analysis based on literature content

### 4.3.1 Extractive text summarization

#### (1) Approach

Figure 7 illustrates the framework of an extractive text summarization system, which is composed of the following key stages:

#### A. Pre-processing

**Tokenization:** The input text is divided into individual words or tokens. **Stop-word Removal:** Commonly used words (e.g., "the," "and," "in") that do not contribute significant meaning are removed. **Stemming/Lemmatization:** Words are reduced to their base or root form to standardize variations of a word. **Feature Extraction:** The text is transformed into a structured format that can be analyzed. Common approaches include:

**N-gram Models:** Sequences of 'n' words are used to capture context.

**Bag-of-Words (BoW):** The text is represented by the frequency of words, disregarding grammar and word order.

**Graph-Based Models:** Sentences or words are represented as nodes in a graph, with edges representing relationships between them [23].

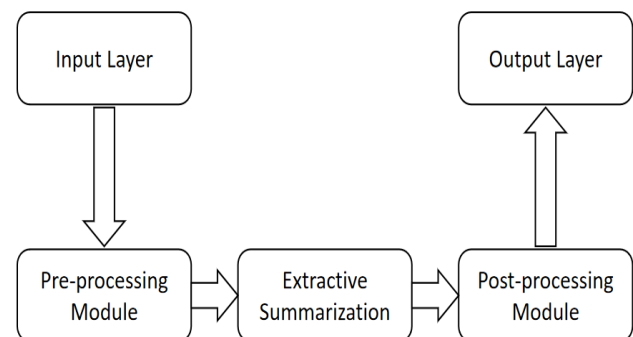
#### B. Extractive summarization

**Scoring Mechanism:** Each sentence in the text is scored based on various features such as frequency, position, or similarity to the main topic.

**Ranking:** Sentences are ranked according to their scores, with higher-ranked sentences considered more important for the summary [24].

**Selection:** The highest-scoring sentences are selected for inclusion in the summary [25].

**Concatenation:** The selected sentences are concatenated to form the final summary, usually preserving the order in which they appeared in the original text.



**Figure 7.** The framework of an extractive text summarization system

### C. Post-processing

Sentence Reordering: The order of sentences may be adjusted to improve readability and coherence.

Pronoun Resolution: Pronouns may be replaced with their corresponding antecedents to avoid ambiguity.

Temporal Expression Normalization: Relative time expressions are converted to actual dates for clarity [4].

This architecture efficiently identifies and extracts the most relevant content from the input text, producing a coherent and concise summary.

#### (2) Methods

Table 3 lists various extractive text summarization techniques, including Statistical-Based, Topic-Based, Graph-Based, Semantic-Based, Machine-Learning-Based, Deep-Learning-Based. It also presents the benefits and weaknesses of each method along with a brief description.

#### (3) Advantages and disadvantages

The extractive summarization approach is often preferred for its speed and simplicity compared to the more complex abstractive methods. This technique directly extracts sentences from the original text, which not only streamlines the summarization process but also enhances accuracy. By preserving the exact terminology and phrasing of the source material, extractive summaries ensure that readers are

presented with content that remains true to the original context and intent. This fidelity to the source is particularly beneficial in domains where precise language is critical, such as legal, scientific, or technical documents. Consequently, the extractive approach minimizes the risk of misinterpretation and maintains the integrity of the original message [26].

Despite its advantages, the extractive summarization approach has several notable drawbacks. Firstly, extractive summaries may contain repetitive sentences or information, leading to redundancy [27]. Secondly, the extracted sentences are often longer than the average sentence length, which can affect the conciseness of the summary. In multi-document scenarios, temporal expressions may conflict because sentences are selected from different documents. Additionally, extracted sentences might lack logical coherence and semantic consistency, primarily due to unnatural connections between sentences, unresolved co-reference relationships, and "dangling" anaphora issues [24]. Finally, important information may be dispersed across different sentences, leading to overlooked conflicting information or incomplete coverage of relevant content. This issue is particularly pronounced in texts with multiple topics and can only be mitigated if the summary is sufficiently detailed [24].

**Table 3.** Extractive text summarization Techniques

Method	Benefits	Weaknesses	Examples
Semantic-Based	Latent Semantic Analysis (LSA) is a language-agnostic technique that produces semantically related statements [28].	The caliber of the produced summary is contingent upon the efficacy of the semantic representation of the source material. Furthermore, the computation of Singular Value Decomposition (SVD) might be laborious [28].	Gambhir and Gupta [28]
Statistical-Based	Reduced processor power and memory requirements. No necessity for additional linguistic proficiency or complex processing, and language-agnostic [29].	Important sentences may be excluded if their scores are lower. Similar sentences might be included due to higher scores.	Gambhir and Gupta [28]
Topic-Based	Summary sentences focus on different topics in the input document(s).	Regardless of their relationship to the primary subjects, sentences with lower scores will not be included in the summary [30]. The quality of the summary is influenced by the themes chosen.	Ma et al. [31]
Graph-Based	Improves coherence and identifies redundant information [24]. It is independent of language [32] and applicable across different domains [33].	Semantically equivalent phrases may be difficult to find in graphs that use the Bag of Words model and similarity metrics [34]. The precision of similarity computation affects the correctness of chosen sentences.	Nasar et al. [32]
Machine-Learning-Based	Requires substantial training data to enhance sentence selection accuracy [24]. Relatively simple regression models can outperform other classifiers in terms of effectiveness [28].	Demands an extensive dataset of manually generated extractive summaries, with each sentence in the original training materials categorized as either "summary" or "non-summary" [24].	Moratanch and Chitrakala [24]
Deep-Learning-Based	The network can be trained to align with human readers' style, and the feature set can be adjusted according to user needs [24].	Manual construction of training data requires significant human effort. Neural networks tend to be slow during both training and testing phases, and it is challenging to understand how the network makes decisions [24].	Yousefi-Azar and Hamey [35]

### 4.3.2 Abstractive text summarization

#### (1) Approach

Figure 8 illustrates the framework of an abstractive text summarization system, which is composed of the following key stages:

##### A. Pre-processing

Tokenization: The input text is divided into individual words or tokens.

Stop-word removal is the process of eliminating frequently used words that don't significantly add meaning, such as "the," "and," and "in."

Stemming/Lemmatization: Words are reduced to their base or root form to standardize word variations.

Stop-word removal is the process of eliminating frequently used words that don't significantly add meaning, such as "the," "and," and "in."

Stemming/Lemmatization: To standardize word variants, words are reduced to their base or root form.

Feature Extraction: The text is transformed into a format suitable for advanced modeling techniques. Common approaches include:

Word Embeddings: Words are represented as dense vectors capturing semantic meanings (e.g., Word2Vec, GloVe).

Contextual Embeddings: Advanced models like BERT or GPT provide dynamic, context-aware representations of words and sentences.

Attention Mechanisms: Focuses on relevant parts of the text to improve the generation of summaries (e.g., in transformer-based models).

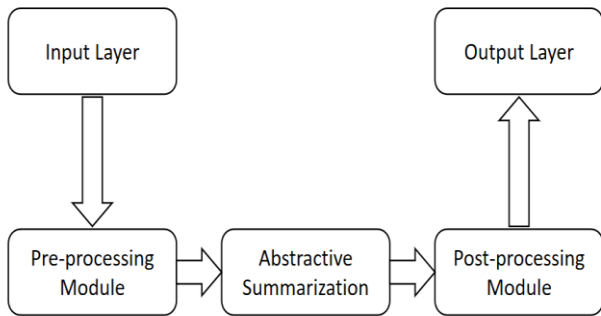


Figure 8. The framework of an abstractive text summarization system

### B. Abstractive Summarization

Sequence-to-Sequence Models: These models, such as LSTM-based or Transformer-based architectures, generate summaries by encoding the input text into a context-aware representation and then decoding it into a summary (e.g., Seq2Seq with Attention).

Pre-trained Language Models: Utilizes large pre-trained models (e.g., GPT-3, T5) to generate human-like summaries by fine-tuning them on summarization tasks.

Reinforcement Learning: Sometimes used to fine-tune the model by rewarding it for generating summaries that meet specific quality metrics.

### C. Post-processing

The generated summary is refined to enhance readability and coherence. This may include:

Grammar and Style Correction: Ensures the summary adheres to grammatical rules and stylistic conventions.

Redundancy Removal: Identifies and removes repetitive information to make the summary more concise.

Fluency Adjustment: Modifies the text to ensure it reads naturally.

Human Evaluation: Human annotators assess the quality of the summaries based on criteria such as relevance, coherence, and readability.

Automated Metrics: The quality of the generated summaries is assessed against reference summaries using metrics such as ROUGE (Recall-Oriented Understudy for Gisting Evaluation).

### (2) Methods

In contemporary Natural Language Processing (NLP) tasks, deep learning algorithms have emerged as the preeminent approach for text summarization. Compared to traditional algorithms, deep learning methods are better at capturing complex semantics and contextual relationships in text, thus offering significant advantages in generating high-quality summaries. This section will focus on introducing the

application of deep learning methods in text summarization, delving into the main frameworks and technical advancements of these methods, without covering other types of algorithms.

Table 4 lists various abstractive text summarization techniques in deep learning, including Seq2Seq Models, Atten techniques, techniques, techniques, techniques, techniques, techniques, techniques, Mechanisms, Transformers, BERT-based Models, GPT-based Models, Pointer-Generator Networks, Hierarchical Transformers. It also presents the benefits and weaknesses of each method along with a brief description.

### (3) Advantages and Disadvantages

It employs more versatile expressions derived from paraphrasing, compression, or fusion to generate superior summaries utilizing distinct terminology not included in the original text [27]. According to Sun et al. [36], the produced summary resembles the manual summary more. When compared to extractive methods, abstractive methods can further reduce the text. This increased condensation results from the ability to further eliminate duplication through the creation of new sentences, which ultimately achieves a strong compression rate [37].

Abstractive text summarization has several notable drawbacks compared to extractive summarization. Firstly, it relies on advanced natural language generation techniques, which make the model training and generation process more complex and computationally intensive. Unlike extractive methods that focus on extracting and rearranging existing content, abstractive summarization involves generating new sentences, posing greater challenges in terms of grammatical accuracy and contextual coherence, which can result in summaries that may be less natural or precise [27]. Additionally, abstractive summarization typically requires extensive annotated data for training, which not only increases the cost of data preparation but also makes the model more susceptible to issues with handling out-of-vocabulary words [38].

### 4.3.3 Hybrid text summarization

#### (1) Approach

Hybrid methodologies include extractive and abstractive techniques. Figure 9 illustrates the framework of a hybrid text summarization. It typically comprises the following phases:

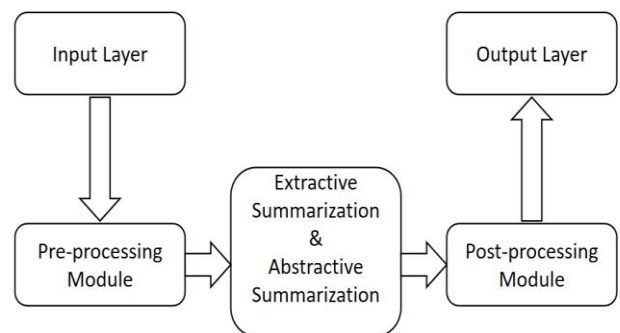


Figure 9. The framework of a hybrid text summarization system

#### A. Pre-processing

Tokenization: The input text is segmented into discrete tokens or words, facilitating subsequent processing.

Stop-word Removal: Frequently occurring, non-informative terms (e.g., "the," "is," and "on") are eliminated to emphasize the more significant information.



Stemming/Lemmatization: Words are truncated to their root or basic form, standardizing differences in word forms.

**B. Extractive Summarization**

Sentence Evaluation: Each sentence in the text is assessed for significance utilizing methodologies such as TF-IDF (Term Frequency-Inverse Document Frequency) or cosine similarity.

Sentence Selection: The highest-ranked sentences are chosen to constitute the essential material of the summary. This guarantees the retention of essential information.

Compression: The selected sentences may be compressed to remove unnecessary details or to merge overlapping information, making the extractive summary more concise.

**Table 4.** Abstractive Text Summarization Techniques in Deep Learning

Method	Benefits	Weaknesses	Examples
Seq2Seq Models	Sequence-to-Sequence models generate summaries by encoding input text and decoding it into a summary. Captures complex semantics. Generate coherent and fluent summaries.	May produce repetitive or irrelevant content. Requires large datasets.	Sutskever [39]
Attention Mechanisms	Enhances Seq2Seq models by allowing the model to focus on different parts of the input text dynamically. Improves contextual understanding. Focuses on important parts of the text.	Can be computationally intensive. May still miss key details.	Bahdanau [40]
Transformers	Uses self-attention mechanisms to process the entire input text in parallel, capturing long-range dependencies. Handles long-range dependencies. Highly parallelizable.	High computational resource requirements Complex model architecture	Vaswani et al. [41]
BERT-based Models	Utilizes BERT for pre-trained contextual embeddings to generate summaries. Captures rich contextual information. Pre-trained on large corpora.	Requires fine-tuning for specific tasks Computationally expensive	Wang et al. [42]
GPT-based Models	Leverages GPT for autoregressive generation of summaries based on large-scale pre-training. Generates highly fluent text. Can produce creative summaries.	Risk of generating non-factual or biased content High computational cost	Radford et al. [43]
Pointer-Generator Networks	Combines the capabilities of both sequence-to-sequence models and pointer networks to handle out-of-vocabulary words. Handles rare words effectively. Provides more accurate summaries.	May still generate irrelevant content Complex architecture	See et al. [44]
Hierarchical Transformers	Utilizes hierarchical models to capture document-level structure and context. Captures long-range dependencies. Effective for document-level summarization	Complex model architecture. High computational requirements.	Liu and Lapata [45]

Feature Extraction: The extractive summary is converted into a structured format for further processing. This might involve:

Word Embeddings: Dense vector representations of words are used to capture semantic meanings (e.g., Word2Vec, GloVe) [46].

Contextual Embeddings: More advanced models like BERT or GPT are employed to provide context-aware word and sentence representations [47].

Attention Mechanisms: Applied to focus on the most relevant parts of the extractive summary when generating the final summary.

**C. Abstractive Summarization**

Sequence-to-Sequence Models: The extractive summary is used as input to a Seq2Seq model, such as a Transformer, which generates a more fluent and human-like summary by encoding the extractive summary and decoding it into a more natural and concise form.

Pre-trained Language Models: Large, pre-trained models (e.g., T5, BART) can be fine-tuned on summarization tasks to enhance the abstractive generation of the summary.

Reinforcement Learning: Fine-tuning may also involve reinforcement learning, where the model is rewarded for producing summaries that are both accurate and readable.

**D. Post-processing**

The combined summary undergoes further refinement to improve readability and coherence. This includes:

Grammar and Style Correction: Adjusting the summary to adhere to grammatical norms and stylistic consistency.

Fluency Adjustment: Ensuring that the summary flows naturally and is easy to read.

Human Evaluation: Human reviewers assess the final summary's quality based on relevance, coherence, fluency, and completeness.

Automated Metrics: Tools like ROUGE and BLEU (Bilingual Evaluation Understudy) are used to quantitatively measure the summary's performance against reference summaries, providing feedback for further model improvements.

**(2) Methods**

Table 5 lists various hybrid text summarization techniques, including Extractive-Abstractive Hybrid Approach, Extractive Pre-processing for Abstractive Summarization, Abstractive Summarization with Extractive Guidance, Hybrid Neural Architectures. It also presents the benefits and weaknesses of each method along with a brief description.

**(3) Advantages and Disadvantages**

Combining the advantages of both extractive and abstractive approaches. The two approaches are complementary, augmenting the overall effectiveness of summarization [25].

The resultant summary demonstrates worse quality than the pure abstractive method, as it depends on extracts rather than the original text. The research community is increasingly

focusing on the extractive ATS method, employing several ways to yield more coherent and relevant summaries, whereas

the abstractive approach is relatively challenging and necessitates significant natural language processing expertise.

**Table 5.** Hybrid Text Summarization Techniques

Method	Benefits	Weaknesses	Examples
Extractive-Abstractive Hybrid Approach	<p>This method first uses extractive techniques to identify important sentences or segments in the text, and then applies abstractive techniques to generate a summary from these selected parts. This approach aims to leverage the strengths of both methods: the accuracy of extractive summarization and the fluency of abstractive summarization.</p> <p>Leverages the precision of extractive methods to ensure key information is included.</p> <p>Improves fluency and readability using abstractive techniques.</p>	<p>The quality of the final summary depends on the effectiveness of both the extractive and abstractive components.</p> <p>May require complex training procedures and fine-tuning.</p>	Chen and Bansal [48]
Extractive Pre-processing for Abstractive Summarization	<p>Extractive techniques are used to identify important sentences or segments, which are then fed into an abstractive model to produce a summary that is both coherent and contextually relevant.</p> <p>Helps in reducing the input size for the abstractive model, potentially improving processing efficiency.</p> <p>Ensures that important information is captured in the extractive phase.</p>	<p>The extractive phase may miss nuanced information that is important for the abstractive phase.</p> <p>Potential for redundancy or information loss during the transition between phases.</p>	Liu and Lapata [49]
Abstractive Summarization with Extractive Guidance	<p>Extractive summarization scores or features are used to guide the abstractive summarizer, helping it focus on the most important parts of the text.</p> <p>Utilizes extractive scores to enhance the focus of the abstractive model, improving the relevance of generated summaries.</p> <p>Can lead to summaries that are both informative and coherent.</p>	<p>The effectiveness of the guidance depends on the quality of the extractive scoring mechanism.</p> <p>May require additional computational resources to integrate and balance the guidance with the abstractive process.</p>	Liu [50]
Hybrid Neural Architectures	<p>This method employs neural network architectures that combine aspects of both extractive and abstractive summarization in a single model. These models are designed to leverage the strengths of both approaches within a unified framework.</p> <p>Integrates Both Approaches: Combines the strengths of both extractive and abstractive methods within a single model, which can improve overall summary quality.</p> <p>Enhanced Understanding: Hierarchical bidirectional transformers capture complex dependencies in documents, enhancing the model's ability to generate more accurate summaries.</p>	<p>Complexity in Design and Training: The model architecture is more complex, which can make design and training more challenging.</p> <p>Resource Intensive: Requires significant computational resources for training and fine-tuning.</p>	Zhang et al. [51]

## 5. RESEARCH CONCLUSIONS AND FUTURE DIRECTIONS

### 5.1 Research conclusions

This study integrates bibliometric and content analysis methods to provide a comprehensive exploration of text summarization technologies and their future development trends. The findings highlight significant trends and advancements within the field and offer valuable insights into future research directions.

**Publication Trends:** Since 2011, there has been a notable increase in the number of publications related to text summarization, reflecting a rise in research interest and activity. This growth is evident in both the volume of literature and citation frequency, underscoring the expanding importance and relevance of the field.

**Journal and Discipline Analysis:** Research on text summarization is primarily concentrated in the field of computer science, with journals such as *Expert Systems with Applications* and *IEEE Access* playing a prominent role. Other journals also contribute to the advancement of text summarization technology. The research spans multiple disciplines, including engineering, mathematics, and social

sciences, demonstrating the broad applicability of summarization techniques.

**Author Collaboration and Influence:** Co-occurrence analysis of authors reveals a relatively simple collaboration network, with key figures such as Yulia Ledeneva, Partha Pakray, Atif Khan, and Rafael Ferreira showing high centrality and influence. This network structure highlights individual contributions and emerging collaboration patterns, emphasizing the critical role of these authors in advancing text summarization research.

**Keyword and Content Analysis:** Co-occurrence analysis of keywords identifies major research areas and hotspots, such as deep learning, extractive summarization, and abstractive summarization. Each method has its unique advantages and challenges. Extractive summarization ensures information completeness by selecting important sentences from the source text but may suffer from redundancy and repetition. Abstractive summarization generates more natural summaries but may face issues with semantic consistency and information omission. Hybrid summarization methods aim to combine the strengths of both approaches but still require further optimization in practical applications.

**Further Analysis Based on Literature Content:** Detailed analysis of the literature reveals the evolution and refinement

of text summarization techniques. Extractive summarization focuses on identifying and extracting key sentences from the source text, ensuring that important information is retained but potentially leading to less coherent summaries. Abstractive summarizing entails the creation of novel sentences that encapsulate the essence of the original text, yielding more coherent summaries but encountering difficulties in preserving semantic fidelity. Hybrid methods attempt to leverage the strengths of both approaches by combining extractive and generative techniques to produce more balanced summaries. Key papers provide insights into the development and practical application of these methods.

## 5.2 Future directions

In future research, text summarization technology can develop in the following directions:

**Enhancing Semantic Consistency in Generated Summaries:** Future research should concentrate on enhancing the ability of generative summarizing techniques to preserve semantic consistency. Researchers can explore more advanced generative models and training methods to mitigate issues such as information omission and semantic inconsistencies in generated content [52].

**Optimizing Hybrid Methods:** Hybrid summarization methods have the potential to combine the strengths of both extractive and generative approaches. Future work can concentrate on refining these hybrid methods to achieve more efficient summarization in practice, while ensuring both information completeness and coherence [53].

**Cross-Language and Multimodal Summarization:** With the growing volume of global data, developing cross-language summarization technologies will become increasingly important. Additionally, summarization techniques that integrate text with other data modalities (such as images or videos) can offer richer information representations [54, 55].

**Addressing Challenges with Long Texts and Complex Content:** Current technologies still face challenges when dealing with long texts and complex content. Future research can focus on improving the ability to process these types of texts to provide more accurate and useful summaries [56].

**Application in Practical Scenarios:** Applying text summarization technology to real-world scenarios, such as legal documents, medical records, and news reports, can offer significant practical value. Future work should focus on effectively integrating the technology into these domains to meet practical needs [57, 58].

## REFERENCES

- [1] Luhn, H.P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2): 159-165. <https://doi.org/10.1147/rd.22.0159>
- [2] Al-Saleh, A., Menai, M.E.B. (2018). Solving multi-document summarization as an orienteering problem. *Algorithms*, 11(7): 96. <https://doi.org/10.3390/a11070096>
- [3] Radev, D., Hovy, E., McKeown, K. (2002). Introduction to the special issue on summarization. *Computational Linguistics*, 28(4): 399-408. <https://doi.org/10.1162/089120102762671927>
- [4] Gupta, V., Lehal, G.S. (2010). A survey of text summarization extractive techniques. *Journal of emerging technologies in web intelligence*, 2(3): 258-268. <https://doi.org/10.4304/jetwi.2.3.258-268>
- [5] El-Kassas, W.S., Salama, C.R., Rafea, A.A., Mohamed, H.K. (2021). Automatic text summarization: A comprehensive survey. *Expert Systems with Applications*, 165: 113679. <https://doi.org/10.1016/j.eswa.2020.113679>
- [6] Nazari, N., Mahdavi, M.A. (2019). A survey on automatic text summarization. *Journal of AI and Data Mining*, 7(1): 121-135. <https://doi.org/10.22044/jadm.2018.6139.1726>
- [7] Kirmani, M., Manzoor Hakak, N., Mohd, M., Mohd, M. (2019). Hybrid text summarization: A survey. In *Soft Computing: Theories and Applications: Proceedings of SoCTA 2017*, pp. 63-73. Springer Singapore. [https://doi.org/10.1007/978-981-13-0589-4\\_7](https://doi.org/10.1007/978-981-13-0589-4_7)
- [8] Yadav, N., Chatterjee, N. (2016). Text summarization using sentiment analysis for DUC data. In *2016 International Conference on Information Technology (ICIT)*, Bhubaneswar, India, pp. 229-234. <https://doi.org/10.1109/ICIT.2016.054>
- [9] Suleiman, D., Awajan, A.A., Al Etaiwi, W. (2019). Arabic text keywords extraction using word2vec. In *2019 2nd International Conference on New Trends in Computing Sciences (ICTCS)*, Amman, Jordan, pp. 1-7. <https://doi.org/10.1109/ICTCS.2019.8923034>
- [10] Kanapala, A., Pal, S., Pamula, R. (2019). Text summarization from legal documents: A survey. *Artificial Intelligence Review*, 51: 371-402. <https://doi.org/10.1007/s10462-017-9566-2>
- [11] Dutta, M., Das, A.K., Mallick, C., Sarkar, A., Das, A.K. (2019). A graph based approach on extractive summarization. In *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2018*, Springer, Singapore, pp. 179-187. [https://doi.org/10.1007/978-981-13-1498-8\\_16](https://doi.org/10.1007/978-981-13-1498-8_16)
- [12] Jacquenet, F., Bernard, M., Largeton, C. (2019). Meeting summarization, a challenge for deep learning. In *International Work-Conference on Artificial Neural Networks*, 11506: 644-655. [https://doi.org/10.1007/978-3-030-20521-8\\_53](https://doi.org/10.1007/978-3-030-20521-8_53)
- [13] Mahajani, A., Pandya, V., Maria, I., Sharma, D. (2019). A comprehensive survey on extractive and abstractive techniques for text summarization. *Ambient Communications and Computer Systems*, 904: 339-351. [https://doi.org/10.1007/978-981-13-5934-7\\_31](https://doi.org/10.1007/978-981-13-5934-7_31)
- [14] Lin, H., Ng, V. (2019). Abstractive summarization: A survey of the state of the art. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1): 9815-9822. <https://doi.org/10.1609/aaai.v33i01.33019815>
- [15] Tandel, J., Mistree, K., Shah, P. (2019). A review on neural network based abstractive text summarization models. In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, Bombay, India, pp. 1-4. <https://doi.org/10.1109/I2CT45611.2019.9033912>
- [16] Suleiman, D., Awajan, A. (2020). Deep learning based abstractive text summarization: Approaches, datasets, evaluation measures, and challenges. *Mathematical Problems in Engineering*, 2020(1): 9365340. <https://doi.org/10.1155/2020/9365340>
- [17] Zhao, J., Liu, M., Gao, L., Jin, Y., Du, L., Zhao, H., Haffari, G. (2020). Summpip: Unsupervised multi-document summarization with sentence graph compression. In *Proceedings of the 43rd International*

- ACM SIGIR Conference on Research and Development in Information Retrieval, New York, United States, pp. 1949-1952. <https://doi.org/10.1145/3397271.3401327>
- [18] Dong, L., Yang, N., Wang, W., Wei, F., Liu, X., Wang, Y., Gao, J.F., Zhou, M., Hon, H.W. (2019). Unified language model pre-training for natural language understanding and generation. *Advances in Neural Information Processing Systems*, 32.
- [19] Gupta, S., Gupta, S.K. (2019). Abstractive summarization: An overview of the state of the art. *Expert Systems with Applications*, 121: 49-65. <https://doi.org/10.1016/j.eswa.2018.12.011>
- [20] Wibawa, A.P., Kurniawan, F. (2024). A survey of text summarization: Techniques, evaluation and challenges. *Natural Language Processing Journal*, 7: 100070. <https://doi.org/10.1016/j.nlp.2024.100070>
- [21] Saiyyad, M.M., Patil, N.N. (2024). Text summarization using deep learning techniques: A Review. *Engineering Proceedings*, 59(1): 194. <https://doi.org/10.3390/engproc2023059194>
- [22] Saleh, M.E., Wazery, Y.M., Ali, A.A. (2024). A systematic literature review of deep learning-based text summarization: Techniques, input representation, training strategies, mechanisms, datasets, evaluation, and challenges. *Expert Systems with Applications*, 252: 124153. <https://doi.org/10.1016/j.eswa.2024.124153>
- [23] Joshi, M., Wang, H., McClean, S. (2018). Dense semantic graph and its application in single document summarisation. *Emerging Ideas on Information Filtering and Retrieval*, 746: 55-67. [https://doi.org/10.1007/978-3-319-68392-8\\_4](https://doi.org/10.1007/978-3-319-68392-8_4)
- [24] Moratanch, N., Chitrakala, S. (2017). A survey on extractive text summarization. In *2017 International Conference on Computer, Communication and Signal Processing (ICCCSP)*, Chennai, India, pp. 1-6. <https://doi.org/10.1109/ICCCSP.2017.7944061>
- [25] Wang, S., Zhao, X., Li, B., Ge, B., Tang, D. (2017). Integrating extractive and abstractive models for long text summarization. In *2017 IEEE International Congress on Big Data (BigData congress)*, Honolulu, HI, USA, pp. 305-312. <https://doi.org/10.1109/BigDataCongress.2017.46>
- [26] Tandel, A., Modi, B., Gupta, P., Wagle, S., Khedkar, S. (2016). Multi-document text summarization-a survey. In *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)*, Ernakulam, India, pp. 331-334. <https://doi.org/10.1109/SAPIENCE.2016.7684115>
- [27] Hou, L., Hu, P., Bei, C. (2018). Abstractive document summarization via neural model with joint attention. In *Natural Language Processing and Chinese Computing: 6th CCF International Conference, NLPCC 2017, Dalian, China, 10619: 329-338*. [https://doi.org/10.1007/978-3-319-73618-1\\_28](https://doi.org/10.1007/978-3-319-73618-1_28)
- [28] Gambhir, M., Gupta, V. (2017). Recent automatic text summarization techniques: A survey. *Artificial Intelligence Review*, 47(1): 1-66. <https://doi.org/10.1007/s10462-016-9475-9>
- [29] Ko, Y., Seo, J. (2008). An effective sentence-extraction technique using contextual information and statistical approaches for text summarization. *Pattern Recognition Letters*, 29(9): 1366-1371. <https://doi.org/10.1016/j.patrec.2008.02.008>
- [30] Wang, Y., Ma, J. (2013). A comprehensive method for text summarization based on latent semantic analysis. In *Natural Language Processing and Chinese Computing: Second CCF Conference, NLPCC 2013, Chongqing, China, pp. 394-401*. [https://doi.org/10.1007/978-3-642-41644-6\\_38](https://doi.org/10.1007/978-3-642-41644-6_38)
- [31] Ma, T., Wang, H., Zhao, Y., Tian, Y., Al-Nabhan, N. (2020). Topic-based automatic summarization algorithm for Chinese short text. *Mathematical Biosciences and Engineering*, 17(4): 3582-3600. <https://doi.org/10.3934/mbe.2020202>
- [32] Nasar, Z., Jaffry, S.W., Malik, M.K. (2019). Textual keyword extraction and summarization: State-of-the-art. *Information Processing & Management*, 56(6): 102088. <https://doi.org/10.1016/j.ipm.2019.102088>
- [33] Nallapati, R., Zhai, F., Zhou, B. (2017). Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1). <https://doi.org/10.1609/aaai.v31i1.10958>
- [34] Khan, A., Salim, N., Farman, H., Khan, M., Jan, B., Ahmad, A., Ahmad, I., Paul, A. (2018). Abstractive text summarization based on improved semantic graph approach. *International Journal of Parallel Programming*, 46: 992-1016. <https://doi.org/10.1007/s10766-018-0560-3>
- [35] Yousefi-Azar, M., Hamey, L. (2017). Text summarization using unsupervised deep learning. *Expert Systems with Applications*, 68: 93-105. <https://doi.org/10.1016/j.eswa.2016.10.017>
- [36] Sun, R., Wang, Z., Ren, Y., Ji, D. (2016). Query-biased multi-document abstractive summarization via submodular maximization using event guidance. In *Web-Age Information Management: 17th International Conference, WAIM 2016, Nanchang, China, pp. 310-322*. [https://doi.org/10.1007/978-3-319-39937-9\\_24](https://doi.org/10.1007/978-3-319-39937-9_24)
- [37] Sakhare, D.Y., Kumar, R., Janmeda, S. (2018). Development of embedded platform for sanskrit grammar-based document summarization. In *Speech and Language Processing for Human-Machine Communications: Proceedings of CSI 2015*, Springer Singapore, pp. 41-50. [https://doi.org/10.1007/978-981-10-6626-9\\_5](https://doi.org/10.1007/978-981-10-6626-9_5)
- [38] Bhat, I.K., Mohd, M., Hashmy, R. (2018). Sumitup: A hybrid single-document text summarizer. In *Soft Computing: Theories and Applications. Advances in Intelligent Systems and Computing*, 583: 619-634. [https://doi.org/10.1007/978-981-10-5687-1\\_56](https://doi.org/10.1007/978-981-10-5687-1_56)
- [39] Sutskever, I. (2014). Sequence to sequence learning with neural networks. *arXiv Preprint arXiv:1409.3215*.
- [40] Bahdanau, D. (2014). Neural machine translation by jointly learning to align and translate. *arXiv Preprint arXiv:1409.0473*. <https://doi.org/10.48550/arXiv.1409.0473>
- [41] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I. (2017). Attention is all you need. In *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA.
- [42] Wang, Q.C., Liu, P.Y., Zhu, Z.F., Yin, H.X., Zhang, Q.Y., Zhang, L.D. (2019). A text abstraction summary model based on BERT word embedding and reinforcement learning. *Applied Sciences*, 9(21): 4701. <https://doi.org/10.3390/app9214701>
- [43] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D.,

- Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8): 9.
- [44] See, A., Liu, P.J., Manning, C.D. (2017). Get to the point: Summarization with pointer-generator networks. *arXiv Preprint* arXiv:1704.04368. <https://doi.org/10.48550/arXiv.1704.04368>
- [45] Liu, Y., Lapata, M. (2019). Hierarchical transformers for multi-document summarization. *arXiv Preprint* arXiv:1905.13164. <https://doi.org/10.48550/arXiv.1905.13164>
- [46] Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26.
- [47] Devlin, J., Chang, M.W., Lee, W., Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv Preprint* arXiv:1810.04805. <https://doi.org/10.48550/arXiv.1810.04805>
- [48] Chen, Y.C., Bansal, M. (2018). Fast abstractive summarization with reinforce-selected sentence rewriting. *arXiv Preprint* arXiv:1805.11080. <https://doi.org/10.48550/arXiv.1805.11080>
- [49] Liu, Y., Lapata, M. (2019). Text summarization with pretrained encoders. *arXiv Preprint* arXiv:1908.08345. <https://doi.org/10.48550/arXiv.1908.08345>
- [50] Liu, Y. (2019). Fine-tune BERT for extractive summarization. *arXiv Preprint* arXiv:1903.10318. <https://doi.org/10.48550/arXiv.1903.10318>
- [51] Zhang, X., Wei, F., Zhou, M. (2019). HIBERT: Document level pre-training of hierarchical bidirectional transformers for document summarization. *arXiv Preprint* arXiv:1905.06566. <https://doi.org/10.48550/arXiv.1905.06566>
- [52] Wang, Y.M., Zhang, J.D., Yang, Z.Y., Wang, B., Jin, J.Y., Liu, Y.T. (2024). Improving extractive summarization with semantic enhancement through topic-injection based BERT model. *Information Processing & Management*, 61(3): 103677. <https://doi.org/10.1016/j.ipm.2024.103677>
- [53] Jia, J., Liang, W., Liang, Y. (2023). A review of hybrid and ensemble in deep learning for natural language processing. *arXiv Preprint* arXiv:2312.05589. <https://doi.org/10.48550/arXiv.2312.05589>
- [54] Li, J., Chen, J., Chen, H., Zhao, D., Yan, R. (2024). Multilingual Generation in Abstractive Summarization: A Comparative Study. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, Torino, Italia, pp. 11827-11837.
- [55] Wang, J.D., Chang, D., Meng, F.Q., Qu, G. (2024). A comprehensive survey and prospect of cross-lingual summarization method research. *Journal of Network Intelligence Taiwan Ubiquitous Information*, 9(1): 384-412.
- [56] Ivgi, M., Shaham, U., Berant, J. (2023). Efficient long-text understanding with short-text models. *Transactions of the Association for Computational Linguistics*, 11: 284-299. [https://doi.org/10.1162/tacl\\_a\\_00547](https://doi.org/10.1162/tacl_a_00547)
- [57] Feijo, D.D.V., Moreira, V.P. (2023). Improving abstractive summarization of legal rulings through textual entailment. *Artificial Intelligence and Law*, 31(1): 91-113. <https://doi.org/10.1007/s10506-021-09305-4>
- [58] Van Veen, D., Van Uden, C., Blankemeier, L., et al. (2024). Adapted large language models can outperform medical experts in clinical text summarization. *Nature Medicine*, 30(4): 1134-1142. <https://doi.org/10.1038/s41591-024-02855-5>