Natural Language Processing Journal 7 (2024) 100070

Contents lists available at ScienceDirect

Natural Language Processing Journal

journal homepage: www.elsevier.com/locate/nlp

A survey of text summarization: Techniques, evaluation and challenges

Supriyono [a,c,*], Aji Prasetya Wibawa [a], Suyono [b], Fachrul Kurniawan [c]

[a] Department of Electrical Engineering and Informatics, Faculty of Engineering, Universitas Negeri Malang, Jl. Semarang no. 5, Malang 65145, Indonesia
[b] Department of Indonesian Literature, Faculty of Letters, Universitas Negeri Malang, Jl. Semarang no. 5, Malang 65145, Indonesia
[c] Informatics Engineering, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia

ARTICLE INFO

Keywords:
PRISMA
Natural language processing
Syntactic
Semantic
Text summarization
Transfer learning

ABSTRACT

This paper explores the complex field of text summarization in Natural Language Processing (NLP), with particular attention to the development and importance of semantic understanding. Text summarization is a crucial component of natural language processing (NLP), which helps to translate large amounts of textual data into clear and understandable representations. As the story progresses, it demonstrates the dynamic transition from simple syntactic structures to sophisticated models with semantic comprehension. In order to effectively summarize, syntactic, semantic, and pragmatic concerns become crucial, highlighting the necessity of capturing not only grammar but also the context and underlying meaning. It examines the wide range of summarization models, from conventional extractive techniques to state-of-the-art tools like pre-trained models. Applications are found in many different fields, demonstrating how versatile summarizing techniques are. Semantic drift and domain-specific knowledge remain obstacles, despite progress. In the future, the study predicts developments like artificial intelligence integration and transfer learning, which motivates academics to investigate these prospects for advancement. The approach, which is based on the PRISMA framework, emphasizes a methodical and open literature review. The work attempts to further natural language processing (NLP) and text summarization by combining various research findings and suggesting future research directions in this dynamic subject.

1. Introduction

Embarking on an exploration of text summarization within the realm of Natural Language Processing (NLP), this study endeavors to unravel the complexities inherent in distilling meaningful insights from extensive textual data. NLP, as a field, seeks to facilitate effective communication between humans and computers by comprehending and processing natural language. Text summarization emerges as a critical application, acting as a bridge between copious amounts of textual information and the need for concise, informative representations (Abdeljaber et al., 2022; Abidin et al., 2020; ALJa'am et al., 2006).

The evolution of text summarization approaches stands as a dynamic narrative, reflecting significant strides over time. From initial methods rooted in syntactic structures to the integration of sophisticated models with semantic understanding, the journey underscores a continual pursuit of more effective and nuanced summarization techniques (Jung et al., 2021; Zhao et al., 2019; Yuan et al., 2021). An understanding of this evolution is vital in comprehending the diverse methodologies employed to tackle the challenges of summarizing intricate textual information (Deroy et al., 2023; Dhawale et al., 2020).

Semantic, syntactic, and pragmatic considerations form the core of effective text summarization (Singh and Deepak, 2021; Sinha et al., 2018). While syntactic structures focus on grammar and word arrangement, semantic understanding delves into the meaning of words and their relationships. Pragmatic aspects consider the context and intention behind language usage. The interaction of these elements shapes the development of summarization techniques, enhancing the extraction of relevant and meaningful information from diverse textual sources (Mohamed et al., 2023; Zhang et al., 2020).

The landscape of text summarization is adorned with a myriad of models and techniques, reflecting ongoing exploration and innovation. From traditional extractive methods that compile existing sentences to abstractive techniques that generate novel summaries, diversity showcases the continuous evolution of text summarization. Pre-trained models, graph-based methods, and neural network architectures stand at the forefront of technological influences, reshaping the dynamics of summarization (Batista et al., 2015; Dave and Jaswal, 2015; Malagi and Radhakrishnan, 2020).

Applications of text summarization span diverse domains, including news articles, scientific literature, business reports, and legal documents (Parmar et al., 2019; Nazar et al., 2016; Moiyadi and Pawar,

* Corresponding author at: Informatics Engineering, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia.
E-mail addresses: supriyono.2305349@students.um.ac.id, priyono@ti.uin-malang.ac.id (Supriyono), aji.prasetya.ft@um.ac.id (A.P. Wibawa), suyono.fs@um.ac.id (Suyono), fachrulk@ti.uin-malang.ac.id (F. Kurniawan).

https://doi.org/10.1016/j.nlp.2024.100070
Received 10 October 2023; Received in revised form 25 March 2024; Accepted 1 April 2024

2949-7191/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

2017). The adaptability of summarization techniques allows for customization based on specific application needs (Rananavare and Reddy, 2017; Mazzieri et al., 2010; Mastronardo and Tamburini, 2019; Mani, 2001). Whether condensing lengthy news articles or distilling key insights from complex research papers, text summarization proves instrumental in enhancing information accessibility and comprehension (Rudinac et al., 2018; Sadiq et al., 2013).

However, despite its advancements, text summarization faces challenges and limitations. Issues such as the loss of crucial information, semantic drift in longer summaries, domain-specific knowledge requirements, and the intricacies of handling multi-document summarization pose ongoing hurdles. Addressing these challenges is integral to refining existing techniques and paving the way for more effective summarization systems.

Looking towards the future, text summarization is poised for exciting trends and opportunities. Transfer learning, artificial intelligence integration, the convergence of NLP, and the exploration of multimodal summarization present promising avenues for further development (Shiva Prakash et al., 2019). Embracing these trends holds the potential to elevate the capabilities and adaptability of text summarization approaches.

Therefore, we meticulously defined the scope for the collection, review, and discussion of the surveyed papers, as well as our research agenda, during the idealization phase by addressing the following research questions: We took a deliberate approach that helped us build a strong framework for our study. This framework guided our exploration of how text summarization methods have changed over time, the complex relationships between semantics and syntax, the different models and techniques used, the evaluation metrics used, and the applications and use cases in different situations. Additionally, our research agenda scrutinized the persistent challenges and limitations faced by existing summarization methods, paving the way for the formulation of future trends and opportunities in the field. This methodological clarity not only enhanced the precision of our study but also facilitated a comprehensive analysis that aligns closely with the intricacies of text summarization within the broader landscape of NLP.

This meticulous approach involved addressing pivotal research questions that would guide the trajectory of our study. One such question, central to our exploration, was defined as follows:

RQ1: How has the evolution of text summarization approaches unfolded over time?

RQ2: How are semantics and syntax used in text summarization?

RQ3: How are the techniques and models in text summarization?

RQ4: How are the evaluation metrics?

RQ5: How are the applications and use cases of text summarization?

RQ6: What are the Challenges and Limitations?

RQ7: What is the trend and opportunity?

The rationale for conducting a systematic literature review (SLR) on text summarization is firmly grounded in the imperative to comprehensively evaluate and synthesize the wealth of knowledge within this swiftly evolving domain. Such a review holds significance for multiple reasons. Firstly, it serves as a robust means to amalgamate diverse research findings and advancements in text summarization methodologies and techniques. By systematically dissecting the existing literature, researchers can discern prevailing trends, emergent technologies, and areas of consensus or contention, thus offering a panoramic view of the field's current landscape. Secondly, an SLR provides a structured framework for assessing the quality and methodological rigor of existing studies in text summarization. Through meticulous procedures, including defining inclusion and exclusion criteria and methodically appraising study designs and methodologies, researchers can gauge the reliability and validity of the evidence base, thereby enhancing confidence in the field's empirical foundations.

Moreover, the systematic review process facilitates the identification of gaps, inconsistencies, and avenues for future research within the domain of text summarization. By synthesizing findings from disparate studies, researchers can pinpoint areas ripe for further investigation, thus charting the course for future research endeavors and advancing the collective understanding of text summarization techniques and applications. Additionally, a systematic literature review serves as a crucial resource for practitioners, policymakers, and stakeholders keen on harnessing text summarization technologies across various domains and applications. An SLR helps people make better decisions and create evidence-based methods for designing, implementing, and evaluating text summarization systems by highlighting the most important ideas and research from previous studies. This makes the systems more useful and significant in a wide range of situations. In essence, the systematic examination of literature on text summarization not only enriches scholarly discourse but also informs practical applications, ultimately propelling the field forward in its quest for innovation and effectiveness.

The culmination of this research endeavor contributes to a broader understanding of text summarization. By investigating the evolution, challenges, and prospects, this study provides insights that inform advancements in NLP and text summarization. Through the exploration of research questions and the synthesis of knowledge, this research seeks to foster innovation and contribute to the ongoing discourse within the field.

2. Research methods: Prisma

In the pursuit of conducting a robust and transparent systematic review, the research methods employed play a pivotal role in shaping the integrity and reliability of the study. In this context, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework emerges as a foundational approach, providing a standardized set of guidelines for the systematic review process. Developed to enhance the clarity and completeness of reporting in systematic reviews and meta-analyses, PRISMA serves as a widely recognized and respected strategy. This systematic review, spanning from August 2023 to January 2024, adheres to the principles outlined by PRISMA, as described by Liberati (2009). By adopting the PRISMA technique, this study ensures a methodologically rigorous and transparent examination of the relevant literature, aiming to contribute valuable insights to the field under investigation.

2.1. Research eligibility, plan, and selection

The planning phase of a research study is paramount, holding immense potential for development and impact on the overall research process. This phase involves the systematic formulation of a research plan, encompassing key elements such as research design, methodology, and ethical considerations. A well-constructed research plan serves as the roadmap for the entire study, guiding researchers in their data collection, analysis, and interpretation endeavors. It provides a solid foundation that ensures the study is methodologically sound, aligned with its research objectives, and conducted ethically.

One of the distinctive features of the planning phase is its adaptability. Researchers can refine and adjust their methodologies as the study progresses, allowing for responsiveness to unforeseen challenges and dynamic changes in the research landscape. This adaptability not only enhances the study's resilience but also positions it to address emerging insights and contribute more effectively to the academic discourse.

The planning phase is also characterized by its commitment to methodological rigor. Researchers meticulously outline their sampling strategies, data collection methods, and variables, ensuring that the study's findings are reliable and valid. This commitment to methodological rigor enhances the credibility of the research, contributing to the robustness and trustworthiness of the study's outcomes.

In the context of research development utilizing the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework, the planning phase gains added significance. PRISMA provides guidelines for conducting systematic reviews, influencing how researchers plan and execute their studies. It outlines steps for literature review, eligibility criteria, and data extraction methods, emphasizing transparency and reproducibility in research planning. Integrating PRISMA into the planning phase ensures a systematic and standardized approach, enhancing the quality and impact of systematic reviews and meta-analyses.

A review's validity, applicability, and comprehensiveness must all be evaluated in relation to its eligibility requirements. The exclusion criteria were intended to be progressive, which is important to note. If the article fails to meet the first exclusion criterion, for instance, it is immediately excluded without any extra verification of the exclusion criteria. This meticulous adherence to exclusion criteria ensures the rigor of the systematic review, contributing to the reliability and validity of the review's findings.

Conditions for inclusion (C)

a C1: Use English with terminology such as "Text Summarization", "Natural Language Processing", or other terms that appear in the title, abstract, or keywords of the article.
b C2: Research Articles or Conference Proceedings and Interpretation of Research Results.
c C3: Paper written in English.

Restriction on inclusion (R)

a R1: No document access.
b R2: When conducting non-English language research.
c R3: This work has no summary text.
d R4: Foreign language, second language, or other language instruction is not included in this research.
e R5: Producing framework recommendations, the research is theoretical.

Using the search string that was developed based on C1 in the Scopus database, the first step of the search method involves getting data out of the database. High-quality papers in the domains of computer science and engineering can be found in this database, which is acknowledged as an important and trustworthy source. The two primary components of this SLR, "Text Summarization" AND "Natural Language Processing", are represented by a straightforward string at the beginning.

Synonyms and related terms were included, nevertheless, to ensure a more thorough search. This defines the search string "Text Summarization" OR "Natural Language Processing" OR "NLP" OR "Scope Text Summarization" OR "NLP Text Summarization" and ("NLP" OR "NLP Summarization" OR "Text Mining" OR "Summary"). The article's publication year was restricted to the period from 2000 to 2023. The papers that were retrieved were loaded into Mendeley in BibTex format. This allowed for the customization, exportation to a spreadsheet, and deletion of duplicate papers.

One of the most important steps in making sure the included studies in systematic literature reviews (SLRs) are high-quality and relevant is the research selection procedure. First, relevant publications are retrieved using pre-established search parameters. Each retrieved document is then put through a rigorous evaluation process in which the keywords, abstract, and title are checked against predetermined eligibility criteria. The evaluation procedure is guaranteed to be impartial and consistent, thanks to this defined methodology. When an article's abstract does not immediately lead to an agreement about whether to include it, the paper is held for additional review during the full-text analysis phase. The reliability and validity of the SLR findings have improved since only studies that satisfy the preset criteria are included in the final review, thanks to this rigorous evaluation process.

2.2. Gathering the data

The data collection process within a systematic literature review (SLR) is a deliberate and methodical undertaking designed to comprehensively identify and retrieve pertinent material from the existing body of literature. The overarching goal is to address specific research questions or objectives, shaping the foundation for meaningful insights and evidence-based conclusions. This meticulous approach begins with the precise formulation of research questions or objectives, which act as a guiding roadmap for the entire review procedure, including the systematic gathering of data.

To ensure the thoroughness and relevance of the gathered data, the process involves the establishment of inclusion and exclusion criteria. These criteria, often encompassing study design, publishing type, language, and relevance to the research topic, serve as filters to select studies that meet predetermined features. Researchers then tailor their search tactics for each database, identifying relevant sources, repositories, and databases to maximize the retrieval of pertinent literature. This phase ensures a comprehensive exploration of the available scholarly material.

Once the search method is applied, a pool of potentially relevant research is generated. This compilation of studies undergoes a rigorous screening process based on predetermined criteria, which includes assessing titles, abstracts, and full texts. The screening phase is crucial in narrowing down the selection to studies that align closely with the established criteria. Following the screening, the included studies undergo a meticulous data extraction process, where key details such as study characteristics, methodology, participants, interventions, outcomes, and results are systematically documented.

Quality evaluation becomes paramount at this stage, aiming to verify the validity and reliability of the included studies and assess the overall strength of the evidence provided. This critical appraisal ensures that the studies chosen for inclusion meet high standards of methodological rigor, contributing to the robustness of the systematic review's findings. By synthesizing and analyzing the gathered data, researchers are equipped to effectively address the research questions or objectives, providing a comprehensive overview of the existing literature on the chosen topic.

Throughout this multifaceted process, rigorous data management and documentation practices are upheld. These practices not only contribute to the transparency and reproducibility of the review findings but also support the growth of knowledge in the relevant field. In the end, a thorough and meticulous data collection process enhances the systematic literature review, making it a valuable tool for further scholarly research and evidence-based decision-making.

In Fig. 1, the meticulous evaluation of condensed texts from the Scopus database is visually depicted, showcasing the rigorous process applied before selecting papers for further examination. This chart serves as a visual representation, elucidating the systematic steps taken to identify and choose pertinent papers that form the basis of analysis in the current study. The figure provides a clear overview of the selection criteria and the strategic approach employed to ensure the inclusion of relevant and high-quality literature in the research investigation.

A. Identification

The identification phase of our systematic literature review was initiated with a thorough and comprehensive search strategy, encompassing various databases, repositories, and pertinent literature sources. This meticulous exploration yielded a substantial total of 3344 papers initially identified as potential sources for inclusion in the review. This exhaustive approach was crucial to ensuring the inclusion of relevant literature pertaining to our research questions.

Moving to the screening stage, the identified papers underwent an initial review based on predefined inclusion and exclusion criteria. Titles and abstracts were scrutinized to assess their relevance, resulting in a refined set of records for further evaluation. This stage aimed

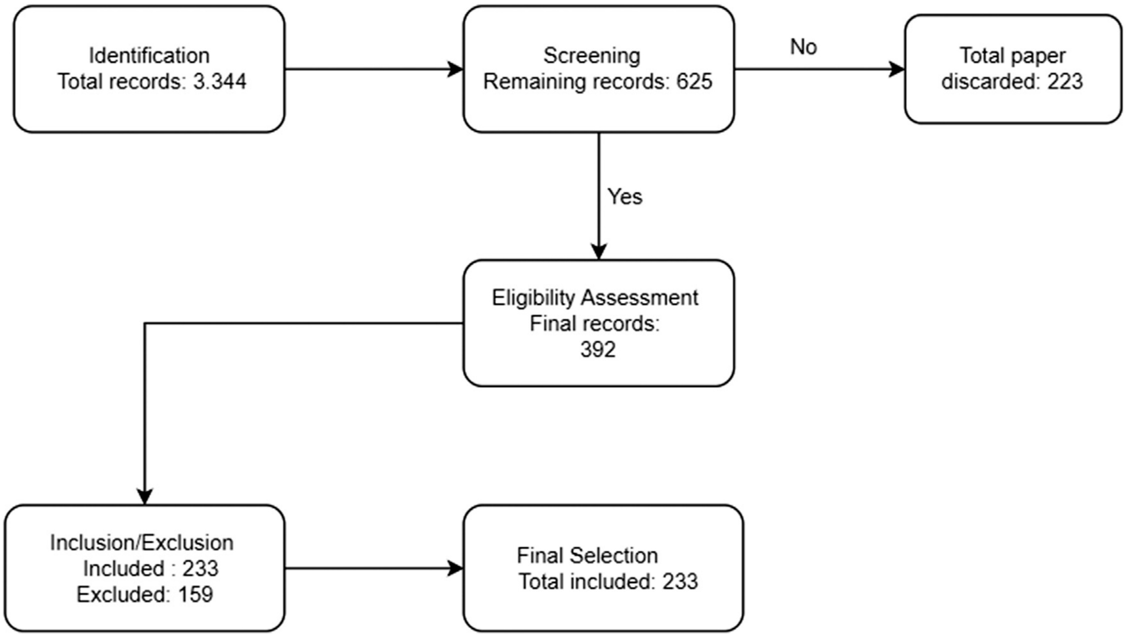

Fig. 1. Study selection flow diagram.

to filter out studies that did not align with the research objectives, ultimately contributing to a more focused selection process.

The subsequent eligibility assessment involved a detailed examination of the full texts of the selected studies. This in-depth analysis aimed to determine whether each study met the predetermined criteria for inclusion in the systematic review. Rigorous assessment criteria were applied to ensure the selection of studies with robust methodologies and relevant findings.

Upon completion of the eligibility assessment, the inclusion and exclusion stage formally determined the final set of studies for the review. The number of included studies (C) and the reasons for excluding others (R) were carefully documented. This transparency in the decision-making process adds credibility and reliability to the systematic review's findings.

The culmination of these stages resulted in the final selection of studies for the systematic literature review. The flow diagram indicates the total number of studies included, highlighting the systematic and transparent nature of the study selection process. This comprehensive approach ensures that the chosen studies meet high standards of relevance, quality, and methodological rigor, setting the stage for a robust analysis and synthesis of the literature to address our research questions.

B. Screening

In the initial stage of study selection, known as identification, an extensive search was undertaken across relevant databases and literature repositories to retrieve potential studies. The total number of records identified during this phase amounted to a substantial figure of 625, representing a diverse array of literature available for consideration in the systematic review. This wide pool of sources formed the basis for the subsequent stages of the study selection process.

The second phase involved the screening of these identified records, wherein a meticulous review of titles and abstracts was conducted. The goal was to assess the relevance of each record to the research questions and predetermined inclusion criteria. After this comprehensive screening, a refined set of records, totaling 625, emerged as the remaining candidates for further evaluation. This screening process played a pivotal role in narrowing down the selection to those records demonstrating initial alignment with the focus of the systematic review.

Moving forward, the third stage comprised a detailed eligibility assessment of the full texts of the identified records. This involved a more thorough examination to determine whether each study met the specific inclusion criteria. Following this meticulous evaluation, the final count of records that met the eligibility criteria was determined, shaping the direction for the subsequent steps in the study selection process.

In the subsequent phase, the inclusion and exclusion criteria were rigorously applied to each study, resulting in a classification of studies as either included or excluded. The reasons for exclusion were carefully documented, ensuring transparency and clarity in the decision-making process. This stage marked a critical juncture where the selection process was fine-tuned to encompass only those studies that met the predetermined criteria.

As the study selection process nears completion, the final selection stage involves presenting the total number of studies that successfully met all inclusion criteria. This culmination serves as a testament to the systematic and rigorous approach employed in identifying, screening, and evaluating studies for inclusion in the systematic review. The resulting selection represents a subset of studies deemed most pertinent to addressing the research questions and objectives of the systematic literature review.

C. Eligibility Assessment

The eligibility assessment phase within the systematic review process serves as a critical juncture where the methodological rigor of studies is meticulously scrutinized to determine their inclusion or exclusion. This phase typically follows the initial screening of studies based on titles and abstracts. During eligibility assessment, researchers delve into the full texts of the identified studies, carefully applying predefined inclusion and exclusion criteria. These criteria often encompass study design, relevance to research questions, and methodological quality. In our specific case, after this detailed examination, a total of 392 studies emerged as meeting the stringent eligibility criteria.

The eligibility assessment aims to ensure that the selected studies align closely with the research objectives and possess the necessary depth and quality for inclusion in the systematic review. Studies that successfully navigate this evaluative phase demonstrate their potential to contribute substantial insights to the overarching research questions, thereby enhancing the robustness of the systematic review.

Upon the completion of the eligibility assessment, the selected 392 studies form a refined pool that has withstood the scrutiny of methodological and topical relevance. This curated selection now serves as the foundation for the subsequent stages of the systematic review, including data extraction, quality evaluation, and synthesis. The comprehensive understanding gained from the eligible studies contributes to the richness and depth of the systematic review's findings, ensuring that the ensuing analysis is grounded in high-quality and pertinent research.

In essence, the eligibility assessment phase is a pivotal step that safeguards the integrity of the systematic review, allowing only studies of the utmost relevance and methodological soundness to progress. This judicious selection process enhances the reliability of the research findings, setting the stage for a comprehensive and insightful analysis of the chosen research topic.

D. Inclusion/Exclusion

In the critical phase of inclusion and exclusion during the study selection process, a comprehensive evaluation was undertaken to determine the eligibility of the identified studies. This stage serves as a pivotal point in ensuring that only studies meeting the predefined criteria are considered for inclusion in the systematic review. The process commenced with the identification of a total of 392 records through extensive searches across various databases and sources.

Following the identification phase, a rigorous screening process was implemented, evaluating titles and abstracts to assess their alignment with the research questions. This initial screening narrowed down the pool to the remaining records. Subsequently, a meticulous eligibility assessment was conducted, involving a detailed examination of the full texts of the selected studies. This thorough review led to the inclusion of 233 studies that closely aligned with the established criteria, showcasing their relevance to the research objectives.

Concurrently, 159 studies were excluded during this phase based on factors such as not meeting inclusion criteria, insufficient data, or methodological limitations. Each exclusion decision was meticulously documented, providing a transparent account of the rigorous standards applied to ensure the integrity and reliability of the final selection. This comprehensive approach underscores the commitment to maintaining a high standard of quality throughout the systematic review process.

The included studies, totaling 233, now form the foundation for the subsequent phases of analysis and synthesis, contributing valuable insights to the systematic review. These studies have been carefully vetted to ensure they meet the established criteria and hold the potential to enhance the understanding of the chosen research topic. The exclusion of 159 studies, while a necessary part of the process, reflects the dedication to upholding stringent standards and selecting studies that contribute meaningfully to the overarching research objectives. Overall, the inclusion and exclusion phase stands as a testament to the thoroughness and precision applied in shaping the composition of the systematic review.

E. Final Selection

The final selection phase of our systematic literature review represents the culmination of a carefully orchestrated process designed to identify and include studies that align with the research questions and objectives. Commencing with a comprehensive identification of potential studies, the subsequent screening and eligibility assessment phases were instrumental in ensuring the inclusion of studies that meet predetermined criteria.

Following a meticulous evaluation of the full texts of the selected studies, the final selection phase concluded with the inclusion of a total of 233 studies. Each study underwent a stringent assessment, considering factors such as study design, relevance, and methodological rigor. The resulting set of studies reflects a curated body of literature that is not only substantial in quantity but also distinguished by its quality and alignment with the focus of our systematic literature review.

The inclusion of 233 studies marks a significant milestone in the research process. This carefully chosen collection of studies serves as the foundation for the subsequent phases of analysis and synthesis. The diverse perspectives and insights offered by these studies promise a rich and comprehensive exploration of the chosen topic.

As we transition from the final selection to the analysis phase, the significance of this curated body of literature becomes apparent. It represents a valuable resource that will be thoroughly examined, dissected, and synthesized to draw meaningful conclusions and contribute to the broader understanding of the research area. The robustness of our final selection ensures that the forthcoming analysis will be built upon a foundation of high-quality, relevant studies.

In summary, the final selection of 233 studies underscores the rigor and methodological soundness applied throughout the study selection process. This curated set positions our systematic literature review for an in-depth exploration, setting the stage for a comprehensive analysis that promises to yield valuable insights and advancements in the chosen field of study.

3. Evolution of text summarization approaches

The evolution of text summarization approaches has traversed distinct methodologies, ranging from rule-based systems to statistical models and machine learning algorithms. Each method brings its own unique characteristics, encompassing strengths and weaknesses that have shaped the landscape of text summarization. As the field progresses, understanding the intricacies and nuances of these approaches becomes imperative for developing more efficient and effective summarization systems.

3.1. Rule-based approaches

Rule-based text summarization approaches have been foundational in the development of summarization systems, employing predefined sets of linguistic rules and patterns to distill the essence of textual content. One notable case of rule-based summarization is the extraction method, where sentences containing specific keywords or phrases deemed significant are extracted for inclusion in the summary (Vishwakarma et al., 2021). For instance, in summarizing news articles about a recent political event, the extraction method might prioritize sentences containing the names of key political figures or locations central to the event (Baralis et al., 2014; Janaki Raman and Meenakshi, 2021). This approach ensures that the summary captures essential details directly related to the subject matter.

Another prevalent rule-based technique is the sentence scoring method, which assigns scores to sentences based on various criteria such as sentence length, frequency of keywords, and the presence of specific linguistic features (Allahyari et al., 2017). Consider a scenario where a rule-based summarizer is tasked with summarizing research papers in the field of artificial intelligence. In this case, the sentence scoring method may assign higher scores to sentences containing technical terms or novel concepts, aiming to encapsulate the core contributions of the research within the summary.

Despite their utility, rule-based approaches such as RAKE (Rapid Automatic Keyword Extraction) and TextRank have limitations, particularly in handling ambiguity and capturing nuanced meanings embedded in the text (Thushara et al., 2019; Chettri and Kr, 2017). For instance, when summarizing opinion pieces or literary works rich in metaphorical language, rule-based systems may struggle to discern the underlying sentiment or thematic elements. This inherent rigidity underscores the need for more adaptive approaches, such as statistical and machine learning-based methods, which can better navigate the complexities of natural language and produce summaries that resonate with human understanding (Saggion and Lapalme, 2000; Saggion and Poibeau, 2013).

3.2. Statistical approaches

Statistical approaches to text summarization have played a significant role in distilling relevant information from voluminous text data. One prominent method is the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm, widely used to identify key terms in a document. TF-IDF calculates a weight for each term based on its frequency in the document compared to its occurrence across the entire corpus (Sung et al., 2016; Tambe et al., 2023). This statistical metric emphasizes terms that are unique to a specific document, helping to prioritize essential information for inclusion in the summary. For instance, in summarizing news articles, TF-IDF can effectively identify and highlight the most salient terms, contributing to the generation of informative and succinct summaries (Cranganu-Cretu et al., 2002; Taylor et al., 2015).

Another statistical approach involves sentence scoring, where sentences are evaluated and ranked based on various statistical criteria. The sumbasic algorithm is a notable example that employs statistical measures such as word frequency and positional importance to iteratively select and remove sentences, creating a concise summary (Teufel, 2018; Thakkar et al., 2010). This method has been particularly effective in scenarios where the objective is to condense text while preserving its essential meaning. In the context of legal documents, for instance, where precision is paramount, sumbasic can offer a streamlined summary that captures the critical points (Nguyen and Dang, 2018; PadmaLahari et al., 2014).

While statistical approaches may lack the semantic sophistication of some newer methods, their simplicity and efficiency make them valuable in certain contexts. In situations where computational resources are limited or when a quick and straightforward summarization is required, statistical approaches remain relevant (Amato et al., 2017a,b; Andhale and Bewoor, 2016). Their ability to distill essential information through quantifiable measures showcases the adaptability and effectiveness of statistical techniques in the diverse landscape of text summarization (Bhola et al., 2022; Alami et al., 2021).

3.3. Machine learning approaches

Machine learning approaches to text summarization represent a paradigm shift in how summaries are generated, relying on algorithms that learn patterns and relationships directly from data rather than being explicitly programmed with rules or heuristics (Thirumoorthy and Britto, 2023; Thomas et al., 2022). One prominent machine learning technique used in text summarization is supervised learning, where models are trained on a labeled dataset consisting of input documents and corresponding human-generated summaries (Gianey and Choudhary, 2017). For instance, in a news summarization task, a supervised learning model might be trained on a dataset of news articles paired with headline summaries, learning to identify the most relevant information to include in the summary based on the input text (Kumar, 2019).

Another machine learning approach widely employed in text summarization is unsupervised learning, where models are trained on unlabeled data without explicit human supervision (Dalianis and Hassel, 2022; Dedhia et al., 2020). One example of unsupervised learning in summarization is the use of clustering algorithms to group similar sentences or documents together based on their semantic content (Alias, 2021). By identifying clusters of related information, unsupervised learning models can generate summaries that capture the diversity of perspectives and topics present in the input text (Verma and Tyagi, 2020; Wu et al., 2021; Martin, 2021). For instance, in summarizing a collection of research papers, an unsupervised learning model might cluster papers based on shared themes or methodologies, producing summaries that highlight the key findings and contributions within each cluster.

Machine learning approaches offer several advantages over traditional rule-based and statistical methods, including the ability to learn from diverse datasets and adapt to different domains and languages (Chali and Mahmud, 2021; Keneshloo et al., 2019). By leveraging neural network architectures such as recurrent neural networks (RNNs) and transformers, machine learning models can capture long-range dependencies and semantic relationships within the text, enabling them to generate more coherent and contextually relevant summaries (Cheng et al., 2020; Xu and Ma, 2023). However, machine learning approaches also pose challenges, such as the need for large amounts of annotated data and computational resources for training complex models (Verma et al., 2019). Moreover, ensuring the ethical and responsible use of machine learning in text summarization remains an ongoing area of research, particularly with regard to issues of bias and fairness in automated content curation and dissemination (Kouris et al., 2019; Kruse et al., 2023; Kulkarni, 2013). Despite these challenges, machine learning approaches continue to drive innovation in text summarization, offering promising avenues for generating high-quality summaries that meet the evolving needs of users in an increasingly data-driven world (Kohakade and Jadhav, 2020; Dr. Vidyagouri and BibiSadiqa, 2022).

3.4. The comparison of several text summarization methods

Table 1 shows the comparison of three text summarization methods: rule-based, statistical, and machine learning approaches, along with their respective pros and cons:

In summary, each text summarization method has its own set of strengths and weaknesses, and the choice of method depends on factors such as the specific requirements of the task, the availability of annotated data, and computational resources. While rule-based approaches offer transparency and simplicity, statistical and machine learning methods provide greater flexibility and performance at the cost of increased complexity and resource requirements (Lee and Lee, 2017; Lehmam, 2010; Zhang et al., 2013; Li and Li, 2018).

Machine learning approaches stand out as having the most significant potential for the development of text summarization systems due to their adaptability, performance, and ongoing innovation. Unlike rule-based and statistical methods, machine learning models, particularly deep learning architectures like transformers, can learn intricate patterns and semantic relationships directly from data, enabling them to summarize text across diverse languages, domains, and genres with minimal manual intervention (Pramudita et al., 2022; Pramoda Devi, 2020). Their remarkable adaptability and generalization capabilities make them well-suited for handling the increasing volume and complexity of textual data across various applications. Moreover, machine learning approaches consistently achieve state-of-the-art results in text summarization tasks, driven by advancements in pre-trained language models, self-supervised learning, and multi-task learning techniques (Prakhar et al., 2020).

Furthermore, the availability of resources and infrastructure, including open-source libraries, pre-trained models, and cloud computing platforms, has democratized the adoption of machine learning in text summarization (Li et al., 2023, 2020; Liao, 2021). Developers and organizations can leverage existing resources to build and deploy machine learning-based summarization systems with ease. Large-scale datasets, such as PubMed and arXiv, provide valuable training data for robust summarization models, facilitating research and experimentation at scale (Humera Khanam and Sravani, 2016). As researchers continue to explore new methodologies and techniques, machine learning approaches are poised to lead the evolution of text summarization, driving innovation and progress in natural language processing and computational linguistics (Kirmani and Shukla, 2022).

4. Semantics and syntactic in text summarization

The development of text summarization techniques has prominently highlighted the crucial role that semantics plays, which includes both syntactic and semantic methods. The dynamic landscape of summarization techniques has undergone a discernible shift from traditional

Table 1
Indicator of text summarization methods.

Aspect	Rule-based	Statistical	Machine learning
Approach	Rigid, predefined linguistic rules	Utilizes statistical models and algorithms	Learns patterns and relationships from data, often with deep learning
Transparency	High	Moderate	Moderate to low
Adaptability	Limited	Limited	High
Complexity Handling	Struggles with complexity and nuance	May struggle with complex language structures	Excels at handling complex relationships and diverse content
Application	Suitable for structured data and consistent news articles	Effective for extractive summarization	Effective for both extractive and abstractive summarization
Computational Resources	Low	Low to Moderate	Moderate to High
Training Data	Specific rules are less data dependent	Requires statistical analysis of data	Requires large datasets for training, especially in deep learning

syntactic approaches to a more nuanced integration of semantic understanding (Hamid and Tarau, 2014; Hassel, 2004). With semantics taking a central role in text summarization, a plethora of diverse methods have surfaced, each distinguished by own unique characteristics, strengths, and weaknesses. This path of evolution shows that people are always trying to make summarization systems better and more efficient (Dong et al., 2021). This shows how important it is to fully understand semantics in order to get meaningful and concise content from textual data (Hong and Chang, 2023; Hovy and Lin, 1996; Jezek and Steinberger, 2008; Jiang et al., 2021).

4.1. Semantic methods

The development of text summarization techniques has brought to light the critical role that semantics plays, including semantic methods. These semantic techniques help to clarify meaning and shape a more complex and contextually rich method of summarizing textual data. Traditional syntactic approaches often relied on grammatical structures and syntactical relationships to condense text (Abdi et al., 2015; Chatterjee and Agarwal, 2023). However, this paradigm shifted with the increased recognition of the significance of semantics in understanding the meaning and context of the text. Semantic methods, which focus on the meaning of words and their relationships, have become integral to modern text summarization (Kipp, 2008; Kaszas et al., 2018).

In semantic text summarization, one notable case is the utilization of Natural Language Processing (NLP) techniques, which enable machines to comprehend the contextual meaning of words and phrases. For instance, algorithms can analyze the semantic relationships between words to identify key concepts and extract essential information for summarization (Haggag, 2013). Semantic approaches also involve the use of semantic graphs, where entities and their relationships are represented, allowing algorithms to navigate and comprehend the intricate semantic structure of the text (MohammedBadry et al., 2013; Lu et al., 2020; Magesh and Ramya, 2020).

In practical terms, consider the summarization of news articles. Semantic methods can discern the main topics, key events, and relationships between entities mentioned in the text (Kipp, 2008; Kireyev, 2008; Israel et al., 2015). By understanding the semantics of the content, a summarization algorithm can generate concise summaries that capture the essential information without losing context or meaning. This illustrates how semantics plays a pivotal role in enhancing the precision and relevance of text summarization.

As the field progresses, ongoing research aims to address the challenges and limitations associated with semantic approaches, such as ambiguity and context-dependent meanings (Mohamed, 2016; Kireyev, 2008). The continuous exploration of semantic techniques underscores their significance in advancing the capabilities of text summarization systems, offering a promising avenue for further refinement and innovation in the field (Wan and Dale, 2001).

4.2. Syntactic approaches

The simultaneous influences of syntactic and semantic methods have shaped the evolution of text summarization approaches. Syntactic approaches traditionally focused on the grammatical structure of sentences, relying on the arrangement of words and their relationships to identify key elements for summarization (Sakhare and Kumar, 2014). For instance, syntax-driven algorithms might prioritize sentences with certain grammatical structures, such as those containing main clauses, to extract essential information. However, the limitations of syntactic methods became apparent as they struggled to capture the nuanced meaning and context within diverse texts (Tymoshenko and Moschitti, 2019).

In contrast, semantic approaches to text summarization have gained prominence by emphasizing the meaning of words and their contextual relationships. A notable case of semantic integration is the utilization of word embeddings, where words are represented as vectors in a semantic space (Liu, 2009; Loukachevitch, 1998). This enables algorithms to capture semantic similarities and differences, allowing for a more nuanced understanding of the underlying meaning of the text. Semantic methods facilitate the identification of key concepts and their connections, providing a more comprehensive basis for generating accurate and contextually rich summaries (Pei-ying, 2009; Pokhrel and Adhikari, 2023; Purushotham Reddy et al., 2021).

A practical example showcasing the synergy of syntactic and semantic approaches is the summarization of legal documents. Syntactic analysis may be employed to identify the structural elements of legal sentences, such as clauses and legal terminology (McLellan et al., 2001). Simultaneously, semantic methods can enhance the summarization process by recognizing the intricate relationships between legal concepts and ensuring that the summarized content reflects not only the syntactic structure but also the nuanced semantic meaning of the legal text. This collaborative utilization of syntactic and semantic techniques underscores their complementary roles in achieving more robust and contextually informed text summarization (Mutlu et al., 2019).

As text summarization methodologies continue to evolve, a balanced integration of both syntactic and semantic approaches remains crucial. When structural insights from syntactic analysis are combined with a semantic understanding of meaning, it is possible for summarization systems to get a more accurate and nuanced picture of the main ideas in different text sources (Zhu et al., 2022).

4.3. The comparison between syntactic and semantic methods

Table 2 shows that the comparison between syntactic and semantic methods in text summarization reveals distinctive approaches with their respective strengths and weaknesses.

Table 2. provides a concise overview of the primary distinctions between syntactic and semantic methods, highlighting their respective strengths, weaknesses, and typical applications in natural language

Table 2
Comparison of syntactic and semantic methods.

Aspect	Syntactic methods	Semantic methods
Focus	Grammatical structure and sentence formation	Word meaning, context, and nuanced semantics
Analysis Techniques	Part-of-speech tagging, dependency parsing	Word embeddings and semantic role labeling
Strengths	Ensure structural coherence	Capture subtle nuances and context
	Useful for grammatical correctness	Facilitate tasks like sentiment analysis
Weaknesses	May lack complete meaning understanding	Can struggle with ambiguity and context
	Less adept at handling nuanced semantics	Computational intensity may be a challenge
Applications	Machine translation, syntax-based tasks	Sentiment analysis, information retrieval
Technique	Dependency Parsing	Word Embeddings
Use Case	Preserving grammatical accuracy in sentences	Understanding context and sentiment in text

processing. The synergy between syntactic and semantic methods in text summarization creates a powerful combination, leveraging their distinct strengths. Syntactic methods contribute by providing structural coherence, ensuring well-organized and logically sequenced summaries. In contrast, semantic methods play a crucial role in deepening the understanding of meaning, facilitating a more nuanced and contextually rich summarization. The complementarity of these approaches results in a comprehensive and effective summarization model (Nenkova, 2005; Niu et al., 2019).

The choice between syntactic and semantic methods often depends on the specific demands of the application (Okurowski et al., 2000; Ou, 2009). Tasks prioritizing grammatical structure, like legal or technical documentation, may favor syntactic methods for their proficiency in capturing intricate sentence structures. Conversely, in contexts requiring a profound comprehension of meaning, such as sentiment analysis or content understanding, semantic methods take precedence for their ability to grasp nuanced semantics and relationships between words (Mirani and Sasi, 2017; Meier and Mujika, 2022).

Ongoing advancements in text summarization involve a concerted effort to explore hybrid approaches that seamlessly integrate both syntactic and semantic elements. This research aims to harness synergies between methods while addressing individual limitations. By creating hybrid models, researchers seek to strike a balance, optimizing the benefits of both syntactic and semantic approaches for sophisticated and adaptable summarization systems (Rane and Govilkar, 2019; Raundale and Shekhar, 2021). This pursuit of hybrid methodologies underscores a commitment to refining and advancing text summarization techniques to meet the evolving demands of diverse textual data and applications (Rautray et al., 2018; Reddy and Guha, 2023). In summary, the comparison highlights the importance of a nuanced understanding of the task, where the effective combination of syntactic and semantic methods leads to more robust and contextually rich summarization systems (Quishpi et al., 2020; Rajasekaran and R. Varalakshmi, 2018).

5. Techniques and models in text summarization

The fusion of various techniques and models has sparked a transformative evolution of the text summarization landscape. From the foundational use of word embeddings to the sophistication of graph-based models, neural network models, and pre-trained models, the field has witnessed a remarkable progression (Verma et al., 2022; Zala et al., 2023). Each of these approaches has its own unique characteristics, leveraging distinctive strengths and grappling with inherent weaknesses. The synergy of these techniques reflects a dynamic journey in the quest for more effective and nuanced text summarization methods. Exploring and improving methods such as word embeddings, graph-based models, neural network models, and pre-trained models reveals the dynamic history of text summarization, contributing to the ever-growing set of summarization techniques (Zhang et al., 2014).

5.1. Word embeddings

Word embeddings have become a cornerstone in the realm of text summarization, fundamentally altering how words are represented and understood within the context of summarization models (Gu et al., 2019; Jain et al., 2017). This transformative approach involves representing words as dense vectors in a continuous vector space, allowing for the capture of semantic relationships and contextual nuances. The integration of word embeddings marks a paradigm shift, enabling more sophisticated and context-aware summarization models (Liu et al., 2019; Mohd et al., 2020; Singh et al., 2018). Notably, popular techniques such as Word2Vec, GloVe, and FastText exemplify the practical application of word embeddings in advancing the field of text summarization (Haider et al., 2020).

A pioneering technique, Word2Vec employs a neural network architecture to learn word representations based on co-occurrence patterns in extensive corpora. The resulting word vectors encapsulate semantic relationships, empowering summarization models to comprehend contextual meaning. In the context of summarizing news articles, for instance, Word2Vec's ability to capture the semantic similarity between words aids in identifying key concepts and generating more coherent and contextually relevant summaries (Wang and Kuo, 2020).

GloVe, or Global Vectors for Word Representation, stands as another influential word embedding technique. By constructing a global word-context matrix and utilizing matrix factorization, GloVe learns word vectors that contribute to a nuanced understanding of words within their broader context (Wang et al., 2013, 2023). In the domain of text summarization, GloVe enhances the summarization process by facilitating the generation of more contextually rich and coherent summaries (Muthiah, 2019).

FastText, an extension of Word2Vec, takes word embeddings a step further by introducing sub-word information into word representations. This proves particularly advantageous in summarization tasks where words may exhibit morphological variations with prefixes or suffixes carrying significant meaning. FastText enhances the summarization model's capacity to comprehend such variations, ultimately leading to the generation of more informative and precise summaries (Mu et al., 2023; Mridha et al., 2021).

The adoption of word embeddings in text summarization represents a departure from traditional methods, providing a more nuanced representation of words and their relationships (Zhang et al., 2013). These embeddings greatly improve the effectiveness of summarization models, capturing semantic meaning and enhancing the coherence and utility of the generated summaries (Kumari and Singh, 2021).

5.2. Graph-based models

Graph-based models in text summarization approaches offer a distinctive methodology, harnessing the intrinsic relationships and structures embedded within a document to distill essential information and create succinct summaries (AL-Khassawneh, 2020; Sornil and Gree-ut, 2006; Zhou et al., 2010). These models adopt a graphical representation of the text, where nodes signify sentences or words, and edges depict the connections between them. One prominent illustration is the TextRank algorithm, which constructs a sentence graph and gauges the similarity between sentences based on co-occurrence patterns. By assigning scores to sentences within the graph, TextRank identifies the most important ones, forming a coherent and contextually rich summary (Zhou et al., 2023).

A noteworthy case is the application of graph-based models in the field of news summarization. In this context, the relationships between sentences play a crucial role in capturing the key events and their interconnections (Bi et al., 2021). Graph-based models, such as TextRank, excel at identifying the most salient sentences that encapsulate the core information, enabling the generation of concise and informative news summaries (Jafarinejad, 2023; Mahajani et al., 2019) (Taeho Jo, 2019). The graphical representation allows these models to discern the significance of each sentence within the broader context of the news article, providing a comprehensive overview.

Another notable example in the realm of graph-based summarization is the LexRank algorithm. LexRank extends the graph-based approach by incorporating eigenvector centrality, which enhances the assessment of sentence importance. For instance, in legal document summarization, where both structural coherence and nuanced meaning are crucial, LexRank's integration of eigenvector centrality proves beneficial. The algorithm does a good job of figuring out how legal ideas relate to each other and choosing sentences that are both semantically and structurally important. This produces summaries that accurately show the complicated legal content (Sharma and Deep, 2014; Sheik and Nirmala, 2021).

Graph-based models underscore their versatility in capturing relationships and structures within diverse domains, emphasizing their application in text summarization (Tsuchiya et al., 2009; Vassiliou et al., 2021). Whether applied to news articles or legal documents, these models showcase their efficacy in distilling meaningful information and generating contextually rich summaries that cater to specific application requirements.

5.3. Neural network models

Revolutionizing text summarization approaches, neural network models showcase their prowess in comprehending and generating concise summaries through the application of artificial intelligence (Chitty-Venkata et al., 2022; Dalal and Malik, 2013). These models often use complicated architectures like recurrent neural networks (RNNs) and transformers to help them find complex patterns and relationships in textual data (Yao et al., 2015). One compelling case is the utilization of encoder–decoder frameworks, where the encoder processes input text and the decoder generates corresponding summaries. For instance, the transformer-based model BERT (Bidirectional Encoder Representations from Transformers) has distinguished itself by effectively understanding contextual information bidirectionally, contributing to more contextually aware summaries (Fang et al., 2023). The flexibility and adaptability of these neural network architectures make them pivotal in the realm of text summarization.

The success of neural network models in text summarization becomes more evident through concrete examples. The GPT (Generative Pre-trained Transformer) series is an illustrative case, wherein models are pre-trained on extensive corpora and fine-tuned for specific summarization tasks (Yang et al., 2023; Bhaskar et al., 2023; Qi and Zhang, 2023; Chen et al., 2023). This pre-training enables the models to grasp intricate language nuances, making them adept at producing coherent and contextually rich summaries. These examples highlight the capacity of neural network models to not only learn hierarchical representations of language but also to adapt to specific summarization requirements, showcasing their effectiveness in diverse applications (Karpagam et al., 2020; Ke et al., 2022).

Despite challenges like computational intensity and the need for substantial training data, neural network models play a crucial role in advancing text summarization (Kaikhah, 2004). Their nuanced understanding of linguistic structures contributes significantly to shaping the landscape of summarization techniques. The ongoing evolution of these models promises further refinement and innovation, aiming for more accurate and contextually relevant summarization techniques (Kumari et al., 2022).

5.4. Pre-trained models

Pre-trained models have become pivotal in revolutionizing text summarization approaches, showcasing their adaptability and prowess in capturing intricate linguistic patterns. One notable exemplar is BERT (Bidirectional Encoder Representations from Transformers). Originally designed for various natural language processing tasks, BERT's pre-training on extensive datasets enables it to grasp contextualized word representations (Parums, 2021). This ability proves particularly advantageous in text summarization, as BERT excels at discerning nuanced dependencies and complexities within a given document. For instance, when applied to news articles, BERT can extract key information and generate concise summaries that reflect a deeper understanding of the context (Rajalakshmi et al., 2023).

Another standout in the realm of pre-trained models is GPT-3 (Generative Pre-trained Transformer 3). Renowned for its state-of-the-art capabilities in language understanding and generation, GPT-3's pre-training on diverse datasets empowers it to comprehend the semantics of input text. In the context of text summarization, GPT-3 can produce human-like summaries by synthesizing information and maintaining coherence. For instance, when tasked with summarizing lengthy scientific papers, GPT-3 showcases its ability to distill complex information into succinct and comprehensible summaries (Abualigah et al., 2020; Sharma and Sharma, 2023; Yadav et al., 2022).

The adoption of pre-trained models in text summarization underlines the transformative impact of leveraging large-scale pre-training on diverse datasets (Patel and Mangaokar, 2020). These models, like BERT and GPT-3, serve as exemplars of the evolving landscape of natural language processing. As they continue to evolve, pre-trained models are likely to play an increasingly crucial role in advancing the efficiency and effectiveness of text summarization systems across various domains and applications (Bin Mohd Amin and Piaralal, 2020; Van Lierde and Chow, 2019).

5.5. The comparison of text processing

Table 3 illustrates the comparison between word embeddings, graph-based models, neural network models, and pre-trained models in the field of text processing, offering insights into their unique characteristics, applications, and effectiveness. The visual representation in the table provides a comprehensive overview of the strengths and applications of each approach, aiding in the understanding of their respective contributions to text processing.

The effectiveness of each model hinges on a nuanced interplay of factors such as task requirements, dataset characteristics, and specific desired outcomes. It necessitates a thoughtful and tailored approach, taking into account the inherent nature of the task, the intricacies of the dataset, and the available computational resources. A comprehensive understanding of these elements is crucial to making informed decisions when selecting, fine-tuning, or developing models that align with the unique demands of the task at hand. This underscores the importance of a judicious evaluation process, where considerations extend beyond algorithmic complexities to include the practical constraints and objectives associated with the given computational environment and the goals of the task (Bhatia and Jaiswal, 2016; Gaikwad and Mahender, 2016).

6. Evaluation metrics

The evaluation of text summarization approaches has undergone a dynamic evolution, progressing from traditional metrics to semantics-focused metrics and incorporating human evaluations. Each phase in this evolution brings forth distinct characteristics, marked by both strengths and weaknesses inherent to the methodologies employed. Traditional metrics laid the foundation for evaluation, relying on quantitative measures to assess the effectiveness of summarization methods.

Table 3
The comparison of text processing.

Aspect	Word embeddings	Graph-based models	Neural network models	Pre-trained models
Representation	Continuous vector space representation of words.	Utilizes graph structures to represent relationships.	Captures sequential dependencies in data.	Pre-learns contextual information from large corpora.
Strengths	Preserves semantic relationships.	Excellent for summarization and key information extraction.	Effective in capturing sequential dependencies.	Versatility across various NLP tasks.
Weaknesses	May struggle with contextual nuances	Challenges with large documents and complexity	The computational intensity may be high.	Resource-intensive during training and deployment
Applications	Sentiment analysis and document clustering.	Text summarization and keyphrase extraction	Machine translation and text generation	Question answering, text completion
Techniques	Word2Vec, GloVe	TextRank, LexRank	Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM)	BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer)
Contextual Understanding	Limited contextual understanding	Global context is captured through graph structure	Sequential dependencies learned over longer spans	In-depth contextual understanding is due to pre-training

As the field advanced, semantics-focused metrics emerged, reflecting a deeper understanding of the nuanced interplay of meaning within summaries. Additionally, human evaluations introduced the essential element of subjective judgment, aligning evaluation processes more closely with the end-users' perspectives. In this multifaceted landscape, each evaluation method contributes unique insights, underscoring the need for a comprehensive and nuanced approach to assess the efficacy of text summarization techniques.

6.1. Traditional metrics

The evaluation of text summarization approaches has witnessed a significant influence from traditional metrics, playing a vital role in quantifying the performance and effectiveness of automatic summarization systems (Upton and Guillot, 2023). Among these metrics, the ROUGE (Recall-Oriented Understudy for Gisting Evaluation) framework has emerged as a cornerstone. ROUGE comprises various submetrics, each designed to measure different facets of summary quality. For instance, ROUGE-N evaluates n-gram overlap, emphasizing the recurrence of contiguous sequences of n items, typically words, between generated summaries and reference documents. Additionally, ROUGE-L measures the longest common subsequence, underscoring the significance of shared words in the evaluation process. By focusing on recall and lexical overlap, ROUGE provides researchers and practitioners with a systematic and quantitative means to objectively assess the resemblance between generated summaries and reference documents (Sri and Dutta, 2021; Shinde et al., 2021).

BLEU (Bilingual Evaluation Understudy) metrics, which were first created for machine translation, have also been applied to text summarization tests, which has expanded the use of quantitative assessment. BLEU concentrates on the precision of n-grams, evaluating how accurately the generated summary aligns with the reference documents. The integration of both ROUGE and BLEU metrics offers a comprehensive framework that enables researchers to systematically evaluate and compare different summarization models. This quantitative lens not only helps us understand the pros and cons of different approaches better, but it also guides the continuous improvement of summarization systems by showing us details about things like recall, precision, and lexical overlap (Alam et al., 2003).

The significance of traditional metrics in text summarization evaluations lies not only in their ability to objectively measure performance but also in their role as benchmarks for comparing different algorithms. These metrics serve as valuable tools for researchers and practitioners, providing a standardized and quantitative foundation for the evaluation and advancement of text summarization models. As the field continues to evolve, the insights derived from traditional metrics remain instrumental in enhancing the coherence and effectiveness of automatic summarization systems (Lloret, 2008; Vinzelberg et al., 2023).

6.2. Semantics-focused metrics

Semantics-focused metrics play a crucial role in advancing the evaluation methodologies for text summarization approaches (Ahmad, 2014; Cao and Zhuge, 2019; Cardoso and Pardo, 2016; Choon-Ching and Selamat, 2008). Traditional metrics, while valuable, often fall short of capturing the nuanced meaning and contextual intricacies of language. Because of this problem, semantics-focused metrics like ROUGE-N have become strong ways to check the quality of summaries by looking at the semantic information stored in n-grams (Ganesh et al., 2022). Emphasizing the overlap of meaningful linguistic units, ROUGE-N provides a more insightful assessment aligned with the inherent goal of summarization to distill essential information while preserving the semantic richness of the original text. This approach helps ensure a comprehensive evaluation that captures the essence of summarization effectiveness.

In addition to ROUGE-N, the METEOR metric stands out as another prominent semantics-focused evaluation tool in text summarization. METEOR goes beyond simple n-gram matching by incorporating synonyms and stemming into its analysis. This nuanced approach allows METEOR to capture semantic similarities between reference summaries and generated summaries more effectively (Gao et al., 2020; Guadalupe Ramos et al., 2019; Gupta et al., 2016). Taking into account stemming and synonyms improves the metric's ability to spot expressions that mean the same thing, which leads to a more thorough evaluation of summarization outputs. So, metrics that focus on semantics like METEOR give a more detailed picture of the semantic quality of the summaries that are generated (Gupta et al., 2023; Goldstein et al., 2000; Foong et al., 2014). This gives us useful information about how accurate the language is and how well it fits the context of the summarization process (Bagalkotkar et al., 2013).

Human evaluations represent a complementary dimension in assessing the effectiveness of text summarization approaches, bridging the gap between automated metrics and the subjective nature of human language understanding (Gupta and Patel, 2020). While automated metrics like ROUGE-N and METEOR bring objectivity and efficiency to the evaluation process, human evaluations provide a qualitative layer by incorporating human judgment. In such evaluations, human assessors can holistically evaluate summaries based on factors like coherence, informativeness, and overall comprehension. By combining semantics-focused metrics with human evaluations, a more comprehensive and balanced understanding of the strengths and weaknesses of text summarization approaches can be achieved, ultimately contributing to the continual refinement and advancement of summarization techniques (MohammedBadry et al., 2013).

6.3. Human evaluations

Human evaluations play a pivotal role in the comprehensive assessment of text summarization approaches, offering a distinctive perspective beyond traditional metrics and semantics-focused measures

(McLellan et al., 2001; Ou, 2009; Pei-ying, 2009). In contrast to automated evaluation methods, human assessments bring an element of subjective judgment that reflects the intricacies of language comprehension and interpretation. One widely employed approach in human evaluations is pairwise comparison, where assessors compare two summaries and rank them based on their effectiveness in conveying the main information or meeting specific criteria. Additionally, Likert scales provide a numerical rating system, enabling evaluators to quantify aspects such as coherence, informativeness, and overall quality. These human-centric evaluations capture the subtleties of language, readability, and natural flow that automated metrics might not fully grasp, contributing to a more holistic understanding of the true effectiveness of summarization systems.

To illustrate the significance of human evaluations, consider a scenario where two summarization models achieve similar scores on traditional metrics like ROUGE and BLEU, indicating comparable performance (Sharifi et al., 2010). However, in a human evaluation, assessors might identify nuances such as one summary being more coherent, linguistically natural, or contextually appropriate, thus revealing the limitations of solely relying on automated metrics. Human evaluations add depth by considering factors like creativity, capturing the essence of the original text, and maintaining readability elements that contribute to the overall user experience. Integrating human judgment into the evaluation process acknowledges the inherent subjectivity of language and ensures a more nuanced understanding of the strengths and weaknesses of text summarization approaches (Siddiqui et al., 2021; Sharma and Sharma, 2022).

Examples of human evaluations abound in summarization research. Researchers often conduct experiments where human assessors rank summaries based on criteria specific to the task, such as capturing key information, maintaining readability, or preserving the intended sentiment. These assessments are crucial for refining summarization models, addressing their shortcomings, and tailoring them to meet the diverse demands of users who seek not just information retention but also a coherent and engaging presentation. Overall, the inclusion of human evaluations in the evaluation framework contributes to a more authentic and user-focused perspective on the performance of text summarization approaches (Jadon and Pareek, 2016; Indra et al., 2020; Jeng et al., 2017).

6.4. The comparison of evaluation metrics

Table 4 provides a comprehensive comparison of three evaluation metrics for text summarization approaches: Traditional Metrics, Semantics-focused Metrics, and Human Evaluations. This comparative analysis offers insights into the performance of summarization models across diverse criteria, encompassing both conventional quantitative measures and assessments aligned with semantic understanding and human perception.

This table provides an overview of the characteristics, strengths, weaknesses, and applications of traditional metrics, semantics-focused metrics, and human evaluations in the context of evaluating text summarization approaches. Each evaluation method contributes unique insights, and a combination of these metrics offers a more comprehensive understanding of the efficacy of summarization techniques.

7. Applications and use cases

The applications and use cases of text summarization have witnessed a dynamic evolution, catering to diverse domains and professional contexts. From the concise distillation of breaking news in articles to the extraction of pivotal insights from intricate scientific literature, the versatility of text summarization has become increasingly evident. Business reports benefit from the streamlined communication of crucial information, facilitating efficient decision-making, while legal documents undergo succinct condensation for enhanced analysis. Beyond these mainstream applications, text summarization has found its niche in specialized domains such as medical records, technical manuals, and industry-specific documents, where tailored summarization approaches address unique challenges. In navigating this landscape, each method exhibits distinct characteristics, boasting strengths and grappling with inherent weaknesses, emphasizing the necessity for a nuanced understanding of the diverse applications and their respective demands within the evolving field of text summarization.

7.1. News articles

In the dynamic landscape of news article summarization, a variety of approaches have emerged to address the challenge of distilling essential information from voluminous texts. Extractive methods, exemplified by algorithms like TextRank, prioritize maintaining the original wording and coherence of news content (Batista et al., 2015; Rananavare and Reddy, 2017). By constructing a graph representation and assigning importance scores based on factors like word frequency and position, TextRank adeptly identifies pivotal sentences that encapsulate the core ideas of the article. This approach is particularly valuable when fidelity to the source text is essential, ensuring that the generated summary reflects the nuanced details of the news story.

On the other end of the spectrum, abstractive summarization techniques, demonstrated by pre-trained models like BERT and GPT, introduce a creative dimension to summarization (Dalal and Malik, 2013; Dedhia et al., 2020). These models, armed with a contextual understanding of language, generate entirely new sentences that convey the essence of the news. By rephrasing and paraphrasing content, abstractive summarization excels at presenting concise yet informative summaries. This approach proves especially useful when brevity and a fresh articulation of key information are paramount, offering a more dynamic alternative to traditional extractive methods.

Hybrid approaches, such as BERTSUM, merge elements of both extractive and abstractive summarization, presenting a nuanced solution to the complexities of news article summarization (Israel et al., 2015). These hybrid models try to use the best parts of both paradigms by first using extractive methods to find important sentences and then using abstractive methods to make the summary even better. Additionally, the importance of domain-specific models, like the Sumy library, is underscored, emphasizing the significance of training summarization models on datasets tailored to the distinct language and context of news articles (Yadav et al., 2021). This wide range of extractive, abstractive, and hybrid methods, along with domain-specific solutions, shows how flexible and useful text summarization is for dealing with news stories in a wide range of situations and needs (Yong-Kwang Kim, 2006; Yang et al., 2019).

7.2. Scientific literature

Summarizing scientific literature represents a crucial endeavor in the realm of knowledge dissemination, where the complexity of research papers often demands advanced text summarization techniques. One prevalent approach involves extractive summarization, where algorithms like TextRank and LexRank meticulously analyze the structure of scientific articles. These algorithms assign importance scores to sentences based on factors such as word frequency and semantic relationships, effectively identifying and consolidating the most crucial passages from the source material. This method is particularly advantageous for preserving the precision and technical accuracy inherent in scientific content.

In parallel, abstractive summarization techniques, leveraging pre-trained language models like BERT or GPT, offer an innovative means of summarizing scientific literature (Abdeljaber et al., 2022; Andhale and Bewoor, 2016). By generating concise summaries that involve understanding and rephrasing the content, these models contribute to a more reader-friendly and condensed representation of complex

Table 4
The evaluation metrics for text summarization approaches.

Aspect	Traditional metrics	Semantics-focused metrics	Human evaluations
Focus	Quantitative assessment of summary features.	Emphasis on capturing semantic content and nuances.	Subjective evaluation by human assessors.
Methodology	Automated measurements (e.g., ROUGE, BLEU).	Analyzes semantic overlap, considering meaning in summaries.	Human judgment is assessed through pairwise comparisons and Likert scales
Strengths	Objective, efficient, and easy to automate.	Captures semantic richness, considers context and meaning.	Reflect nuances, creativity, and the overall user experience.
Weaknesses	May not fully grasp semantic intricacies.	May not cover all aspects of semantic meaning.	Subject to individual biases, resource-intensive.
Applications	Widely used as benchmarks for model comparison.	Provides a refined perspective on semantic quality.	Offers a user-focused view, critical for real-world impact.
Examples evaluation	ROUGE, BLEU.	ROUGE-N, METEOR.	Pairwise comparisons and Likert scales in experimental setups.
Utilization Examples	Systematic comparison and benchmarking of models.	Assessing semantic accuracy and contextual appropriateness.	Fine-tuning models based on user preferences and needs.

research. Using domain-specific models that have been trained on datasets of scientific literature also improves the summarization process by capturing the specific language and context of scientific discourse. These models play a pivotal role in producing summaries that strike a balance between accuracy and accessibility.

Scientific literature often presents intricate concepts and employs specialized language, necessitating a nuanced approach to summarization (Humera Khanam and Sravani, 2016; Foong et al., 2014). Pre-trained models with a focus on contextual understanding, such as BERT, contribute significantly to achieving more coherent and informative summaries. These models adeptly comprehend the nuanced language used in scientific articles, ensuring that the generated summaries reflect the depth and significance of the original research. As text summarization techniques continue to evolve, the adaptation of these methods to scientific literature underscores their crucial role in conveying complex findings to researchers, students, and professionals within the scientific community.

7.3. Business reports

Summarizing business reports is an essential undertaking in the corporate landscape, where the dense and multifaceted nature of such documents demands specialized text summarization approaches (Hegdepatil and Davuluri, 2021). Extractive methods, including algorithms like TextRank and LexRank, excel at pinpointing and consolidating pivotal information directly from business reports (Jing and Wan, 2023; Mutlu et al., 2019). These algorithms analyze the structural characteristics of the document, assigning importance scores based on relevance and prominence. Extractive summarization proves invaluable when precision and fidelity to the original text are paramount, ensuring accurate representation of financial data, market trends, and strategic recommendations.

In contrast, abstractive summarization techniques, often leveraging pre-trained models like BERT or GPT, provide a more dynamic approach by generating new sentences that capture essential information concisely and comprehensibly (Faizal and Abraham, 2022). These models possess a nuanced understanding of the intricate language used in business reports, enabling the creation of summaries that are not only succinct but also reader-friendly. Abstractive methods become particularly beneficial when the goal is to present clear and accessible representations of complex business insights (Niu et al., 2019).

Also, domain-specific summarization models that are trained on datasets with lots of business-related content help make summarization more accurate by understanding the specific language and context of the business domain (Hegdepatil and Davuluri, 2021). This ensures that the generated summaries not only convey key information but also maintain industry-specific nuances crucial for decision-makers (Motilal Lodhi et al., 2022; Moher et al., 2015). The continuous refinement of summarization techniques tailored to business reports underscores their pivotal role in providing swift and actionable insights to professionals and decision-makers navigating the dynamic corporate landscape (Motilal Lodhi et al., 2022).

7.4. Legal documents

Legal document summarization involves navigating the complexities of intricate texts, such as contracts and court opinions, and distilling essential information for efficient comprehension (Jain et al., 2021). Extractive summarization methods, like TextRank, prove instrumental in maintaining the precision and fidelity of legal details (Kanapala et al., 2019; Jain et al., 2021). In a scenario involving a complex contract, extractive summarization ensures that key passages are directly extracted, preserving the specificity of legal language critical for understanding contractual obligations. This approach caters to legal professionals by swiftly providing access to nuanced details, aiding in decision-making processes (Kanapala et al., 2019; Karmaker and Hossen, 2019).

Complementing extractive methods and abstractive summarization techniques, powered by pre-trained language models like BERT or GPT, enhances clarity and accessibility in legal document summarization. For instance, when dealing with intricate court decisions, abstractive summarization generates concise yet reader-friendly summaries (Feijo et al., 2023; Ferreira et al., 2014; Haas, 2013). This dynamic approach captures essential legal information, maintaining accuracy while facilitating a more accessible understanding of the case. Abstractive methods become valuable tools in legal research, enabling legal professionals to efficiently comprehend complex legal concepts and make informed decisions (Vaissnave and Deepalakshmi, 2023).

Domain-specific expertise plays a pivotal role in legal summarization, with models trained on legal datasets catering to the specialized language and terminology within legal texts. These models ensure that summaries accurately convey nuanced meanings, especially in cases involving specific legal domains like intellectual property law (Vaissnave and Deepalakshmi, 2023). By addressing the unique language and context of legal documents, domain-specific models contribute to the efficiency of legal research and decision-making, providing summaries that encapsulate intricate legal nuances essential for legal professionals (Thirumoorthy and Britto, 2023; Biswas et al., 2018).

Table 5
The comparison of applications and use cases.

Aspects	News articles	Scientific literature	Business reports	Legal documents	Other niche domains
Use Case	Summarize breaking news, events, or stories	Condense research papers and articles	Extract key insights, trends, and data	Provide concise legal case summaries	Summarize specific domain knowledge
Objectives	Provide quick overviews for readers	Facilitate quick understanding for researchers	Aid decision-making and strategic planning	Assist legal professionals in extracting key information	Distill essential information in specialized areas
Challenges	Handling evolving news stories and updates	Grappling with technical jargon and complex concepts	Navigating through diverse business data	Addressing legal language complexities	Adapting to varying domain-specific terminologies
Techniques	Emphasis on recent information, diversity	Emphasis on technical details and context	Focus on financial figures and trends	Incorporate legal terminology and precedents	Adapt to specialized vocabularies
Benefits	Quick access to news highlights	Rapid comprehension of research findings	Efficient decision-making processes	Streamlined legal case reviews	Enhanced understanding in niche areas
Examples application	Summarizing news articles from major outlets	Summarizing scientific research papers	Condensing quarterly financial reports	Extracting key legal points from case documents	Summarizing domain-specific documents

7.5. Other niche domains

Summarizing content in other niche domains involves tailoring approaches to the specific characteristics of each domain (Antony et al., 2023). In the case of technical manuals, precision and clarity in conveying procedural information are paramount. Extractive summarization methods, exemplified by algorithms like TextRank, excel at identifying and condensing step-by-step processes. For instance, in a technical manual for machinery operation, extractive summarization can pinpoint key instructions, safety protocols, and operational steps, facilitating efficient comprehension for users navigating complex machinery documentation (Benharrak et al., 2022; Purushotham Reddy et al., 2021; Vale et al., 2020).

In the realm of medical records, contextual understanding is crucial for accurate summarization. Abstractive summarization techniques, leveraging pre-trained models like BERT, prove instrumental in generating summaries that capture the nuanced medical context and treatment plans. Consider a scenario involving a patient's medical history; abstractive summarization ensures that the summary maintains essential medical details while presenting a clearer overview for healthcare professionals. This approach enhances accessibility and comprehension, facilitating quicker decision-making based on the summarized medical information (Umadevi et al., 2018).

Summarizing patents represents a unique challenge due to the hybrid nature of legal and technical content (Okurowski et al., 2000; Lasya Sriranga et al., 2020). Hybrid approaches that combine legal precision with technical summarization are essential for efficiently capturing both aspects. Extractive methods can identify legal claims and technical specifications, while abstractive techniques ensure clarity and accessibility in the summary. In the case of patent summarization, this hybrid approach effectively captures both the legal protection elements and the technical innovations described in the document (Deng et al., 2022; Benharrak et al., 2022). This shows that summarization techniques can be easily adjusted to meet the complex needs of patents, where legal and technical aspects are closely linked (Saggion and Lapalme, 2000).

7.6. The comparison of differences in use cases for text summarization

Table 5 provides a comprehensive overview of the distinctions in the application and use cases of text summarization, delineating specific approaches tailored for news articles, scientific literature, business reports, legal documents, and other specialized domains (Salman, 2023). This comparison sheds light on the complex needs and wide range of functions of text summarization methods that are needed to deal with the problems that come up with different types of content and the specific needs of each industry.

Text summarization techniques address the distinctive requirements and challenges inherent in diverse applications, underscoring the significance of tailored approaches that suit specific use cases and domains. The adaptability of summarization methods ensures their effectiveness in providing concise and relevant information across a wide array of contexts.

8. Challenges and limitations

The terrain of text summarization is complex, with both obstacles and restrictions that affect the efficacy of various strategies. Some of these problems include losing important information through text imputation, the possibility of semantic drift in longer summaries, the need for domain-specific knowledge, the difficulty of summarizing multiple documents, and the ethical issues that come up during the process. Each approach to text summarization brings forth its own distinctive set of characteristics, encompassing strengths and weaknesses, reflecting the complex interplay between the evolving methods and the hurdles they encounter in extracting meaningful and concise content from diverse textual data.

8.1. Loss of crucial information: Teks imputation

The challenge of losing crucial information in text summarization, often exacerbated by text imputation, manifests in various scenarios with notable implications. In news article summarization, for instance, where timely and accurate reporting is paramount, the use of text imputation techniques may lead to the omission of critical details, distorting the representation of events. If a key element is inaccurately imputed or omitted, the resulting summary may misinform readers and compromise the integrity of the summarization process. Striking a balance between conciseness and information preservation is imperative to mitigate the risk of vital details being lost during the summarization process.

One pertinent example of text imputation challenges arises in the context of scientific literature summarization. Research papers often contain complex terminology and specialized information. If text imputation inaccurately fills in missing details, the summary may convey misleading findings or misinterpret the nuanced content of the original document. In this case, the loss of crucial information due to imputation can have far-reaching consequences, impacting the understanding and dissemination of scientific knowledge.

Furthermore, the domain of legal document summarization faces challenges related to text imputation. Legal texts often involve language, and nuanced details are crucial for accurate interpretation. Text imputation errors in legal summarization could lead to misrepresentations of legal arguments or case outcomes, potentially influencing legal decisions. The complexities of legal language and the high stakes involved underscore the need for robust text summarization approaches that minimize the loss of critical information through judicious handling of text imputation challenges. Addressing these challenges requires a nuanced approach to text imputation that prioritizes accuracy and fidelity to the original content.

8.2. *Semantic drift in longer summaries*

Semantic drift in longer summaries poses a challenge to maintaining the fidelity of information as the length increases. This phenomenon is particularly evident in scientific literature summarization, where the intricate details of research papers can lead to deviations from the primary focus. For instance, a lengthy scientific article on genetics may explore diverse aspects such as methodologies, experimental results, and ethical considerations. However, in summarizing such content, an extended summary may inadvertently introduce unrelated details, causing a drift in semantics and diluting the precision of the original focus on genetics.

In legal document summarization, semantic drift can manifest when lengthy legal briefs or judgments are condensed into summaries. The complexity of legal language and the inclusion of various legal precedents and arguments can pose a challenge to maintaining the intended meaning. A well-intentioned summary that grows in length may start incorporating peripheral legal concepts, resulting in a semantic drift that could potentially mislead readers relying on the summary for a quick understanding of the legal case.

Addressing semantic drift is crucial in news article summarization as well. For instance, a detailed news article covering a current event may delve into multiple aspects, including background information, expert opinions, and historical context. A lengthy summary might inadvertently introduce unrelated details or deviate from the central theme, causing a semantic shift that affects the clarity and accuracy of the summarized information. These examples underscore the importance of carefully managing semantic drift in longer summaries across various domains to ensure that the condensed content remains faithful to the original context and maintains its intended meaning.

8.3. *Domain-specific knowledge requirements*

Text summarization faces a formidable challenge in meeting the specific knowledge requirements of various domains, contributing to the nuanced nature of information extraction. In the medical field, summarizing research papers and clinical reports requires an in-depth understanding of intricate medical concepts, treatment modalities, and the ever-evolving landscape of healthcare. For example, a summarization system must grapple with the complexities of drug interactions, disease mechanisms, and the latest medical advancements to produce summaries that accurately convey critical information. The absence of domain-specific knowledge in medical summarization could lead to inaccuracies and potentially compromise the utility of the generated summaries.

Legal document summarization encounters a distinct set of challenges, demanding expertise in legal jargon, case law, and statutory intricacies. The summarization of legal briefs, court decisions, or contracts necessitates a keen awareness of legal precedents and the historical context of cases. A system without the requisite domain-specific knowledge might struggle to navigate the complexities of legal language, risking misinterpretations that could have significant consequences. In the financial domain, summarizing reports demands a comprehensive understanding of financial terms, accounting principles, and industry-specific metrics. Summarization tools should adeptly interpret balance sheets, income statements, and cash flow reports to distill key financial insights accurately. The absence of financial domain expertise may lead to oversights in crucial financial data, diminishing the reliability and utility of the generated summaries.

In the end, using domain-specific knowledge becomes an important part of solving problems and making sure that text summarization is accurate and useful in a wide range of specialized areas. The adaptability of summarization systems to these specific knowledge requirements is essential for maintaining the integrity and accuracy of the generated summaries across medical, legal, financial, and other specialized domains.

8.4. *Handling of multi-document summarization*

Multi-document summarization presents a multifaceted challenge in the realm of natural language processing, requiring a sophisticated approach to distill relevant information from numerous source documents. The intricacy lies in the necessity of intelligently fusing diverse content, accommodating varied perspectives, and handling potential redundancies. One of the primary challenges involves determining the relevance of each document, factoring in considerations such as document length, thematic alignment, and the presence of critical terms. The summarization system must effectively weigh the significance of each document to create a cohesive summary that avoids repetition and provides a comprehensive overview.

Cross-document coreference adds another level of difficulty, as it requires resolving pronouns, references, and named entities that appear in multiple documents so that the summary stays logical. Achieving a consistent tone and style throughout the summary is paramount, given the diversity in writing styles and tones present in the source documents. Scalability becomes a key consideration, as the system must efficiently process and analyze information from a multitude of documents, accounting for variations in document lengths. Also, coming up with good evaluation criteria that are specific to the details of summarizing multiple documents is still hard. This is because the task is always changing, and we need strong evaluation criteria.

To overcome these challenges, the integration of advanced natural language processing techniques and machine learning algorithms is imperative. This ensures the system's ability to navigate through the complexities of multi-document summarization, providing informative and coherent summaries that capture essential content from diverse source documents. As research and development in this field progress, the refinement of methodologies and the incorporation of domain-specific knowledge will contribute to further enhancing the accuracy and effectiveness of multi-document summarization systems.

8.5. *Ethical considerations*

Ethical considerations in text summarization are fundamental to navigating the complex landscape of information extraction and representation. One significant ethical challenge lies in the potential reinforcement of biases within the training data, which may lead to summaries that reflect and perpetuate societal prejudices. Addressing this concern requires a proactive approach to identify and mitigate biases, promoting fairness and inclusivity in the generated summaries. Ethical text summarization endeavors to create systems that actively counteract biases and contribute to more equitable representations of diverse perspectives and topics.

Privacy concerns also loom large in the ethical landscape of text summarization. As systems distill information from potentially sensitive or private documents, it becomes imperative to ensure compliance with data protection regulations. Balancing the need to summarize relevant information with the imperative to respect individuals' privacy rights is at the core of ethical text summarization practices. Striking this balance necessitates a nuanced understanding of privacy principles, transparency in data handling, and an unwavering commitment to safeguarding user data.

Furthermore, the ethical deployment of text summarization involves considerations of transparency and explainability. Users and stakeholders are rightfully concerned about the opacity of algorithms, and ethical practices dictate a commitment to making the functionality of summarization models more transparent. This involves providing clear and understandable explanations for the decisions made by the system, fostering trust and accountability in the use of text summarization technology. Ethical text summarization strives for transparency to empower users with a deeper understanding of the technology and its implications.

9. Future trends and opportunities

The future landscape of text summarization is poised for transformative advancements, presenting a host of opportunities to redefine how information is distilled and presented. Several significant trends that promise to have a significant impact on the field are present in this evolution. Transfer learning and zero-shot learning stand out as pioneering approaches, enabling models to leverage pre-existing knowledge and adapt to new domains, thereby enhancing the adaptability and efficiency of text summarization. The infusion of artificial intelligence (AI) introduces a paradigm shift, empowering systems with advanced contextual understanding and the ability to capture intricate relationships within textual content. Concurrently, the integration of Natural Language Processing (NLP) techniques is set to refine the linguistic comprehension of summarization models, allowing for more nuanced and human-like summaries. Furthermore, the convergence of text with other modalities, such as images and audio, heralds the era of multimodal summarization, promising a holistic approach to information extraction. This introduction explores the anticipated future trends and opportunities, encapsulating the potential of transfer learning, AI, NLP, and multimodal approaches to revolutionize the landscape of text summarization.

9.1. Transfer learning and zero-shot learning

In the realm of text summarization, the potential for development is notably pronounced in both transfer learning and zero-shot learning approaches. Transfer learning offers a robust solution when a substantial amount of labeled data is available in a related domain. The strategy involves training models on a diverse dataset and fine-tuning them for specific domains, streamlining the adaptation process. The potential for further development in transfer learning includes refining pre-trained models, optimizing knowledge transfer efficiency, and addressing challenges associated with shifts between different domains. This approach proves particularly effective when overarching patterns can be harnessed across diverse subjects.

Conversely, zero-shot learning holds considerable promise for scenarios where labeled data is scarce or entirely absent in the target domain. Its adaptability to entirely new and unseen domains position it as a valuable asset in the evolving landscape of information summarization. The potential for development in zero-shot learning encompasses refining models' capabilities to generalize and interpret content from diverse and evolving subjects. By focusing on improving generalization capabilities and addressing challenges related to unfamiliar terminologies, zero-shot learning demonstrates its potential to handle a variety of topics without explicit training data.

Looking ahead, the synthesis of transfer learning and zero-shot learning stands out as a compelling avenue for the future of text summarization development. A hybrid approach that harnesses the strengths of both paradigms could lead to more versatile and powerful summarization models. This strategy envisions models that efficiently leverage existing knowledge while maintaining adaptability to entirely new and diverse domains. Future research may focus on refining these hybrid models, addressing potential limitations, and establishing frameworks that dynamically select between transfer and zero-shot learning based on the nuances of the target domain. Such integrated approaches could pave the way for a transformative era in text summarization, where efficiency and adaptability seamlessly coalesce across a spectrum of domains and subjects.

9.2. Artificial intelligence in text summarization

The future trend and opportunity of artificial intelligence (AI) in text summarization hold significant promise, representing a paradigm shift in how information is distilled and presented. As AI continues to advance, its integration into text summarization models is expected to bring about enhanced contextual understanding, allowing for more nuanced and coherent summaries. AI-driven models, particularly those based on advanced machine learning algorithms and deep learning architectures, are poised to elevate the accuracy and sophistication of summarization outputs. The opportunity lies in the potential to develop systems that not only capture surface-level information but also delve into the underlying meaning and relationships within the text, leading to more contextually rich and insightful summaries.

The integration of AI in text summarization is expected to go beyond traditional rule-based approaches, allowing models to dynamically adapt to different writing styles, genres, and domains. Natural Language Processing (NLP) techniques, a subset of AI, play a pivotal role in refining linguistic comprehension within summarization systems. Future trends suggest that AI-driven summarization models will increasingly excel in understanding idiomatic expressions, cultural nuances, and complex sentence structures. This presents an opportunity to create more versatile and adaptive summarization systems that cater to a diverse range of content types and user preferences.

Also, as AI keeps getting better at summarizing text, more research will likely be done into hybrid models that combine AI-driven methods with other methods, like graph-based models or word embeddings. These hybrids aim to capitalize on the strengths of AI while mitigating potential limitations, providing a holistic solution for more effective information extraction. To sum up, the future trends and opportunities of AI in text summarization show a huge step forward in making more advanced, context-aware, and flexible summarization systems that can adapt to the changing needs of different textual data and applications.

9.3. The integration of natural language processing

The future trajectory of text summarization presents exciting opportunities through its integration with other Natural Language Processing (NLP) tasks, notably question answering and sentiment analysis. The synergy between text summarization and question answering is poised to redefine information retrieval, where succinct summaries generated by summarization models become valuable resources for addressing specific queries. This integration streamlines the process of extracting relevant information from documents, fostering a more seamless transition from understanding the broader context to addressing precise user questions. The evolving trend may involve further refining the interoperability between summarization and question–answering models, leading to more efficient and targeted information extraction.

Another promising avenue is the integration of text summarization with sentiment analysis, which provides a richer understanding of the subjective elements within textual content. As summarization models distill key information, incorporating sentiment analysis enables the interpretation of the emotional tone expressed in the original text. This integration proves valuable in contexts where grasping not only the factual details but also the underlying sentiments is essential. Future developments may delve into more advanced techniques that go beyond summarizing factual content, encompassing the nuanced emotional aspects, thereby contributing to a more comprehensive interpretation of the sentiment within the text.

The overarching vision for the future entails the enhancement of cross-task synergies, with text summarization seamlessly collaborating with various NLP tasks. This may lead to the creation of unified models proficient in summarization, question answering, and sentiment analysis concurrently. The development of such integrated systems represents a significant leap forward, promising more sophisticated, versatile, and contextually aware NLP models. The aim is to cater to diverse user needs, enabling a more holistic and natural interaction between users and AI systems in comprehending and extracting valuable insights from textual data.

9.4. Multimodal summarization

The future of text summarization is undergoing a transformative shift with the advent of multimodal summarization techniques, paving the way for more comprehensive and nuanced content distillation. One notable trend is the seamless integration of textual and visual information, enabling summarization models to capture a more holistic understanding of documents that include both text and images. This integration ensures that the summarization process not only considers the textual context but also incorporates visual elements, resulting in summaries that encapsulate the entire narrative of multimedia-rich content, such as news articles featuring both textual information and accompanying images.

As multimedia content continues to proliferate, the future holds promising opportunities for advancements in audio–visual summarization. Podcasts, videos, and other audio–visual formats present a challenge and an opportunity for summarization models to evolve and effectively distill key information from diverse modalities. Future trends suggest the development of summarization systems capable of analyzing spoken content, interpreting visual cues, and providing concise summaries that encapsulate both auditory and visual elements. This trajectory aligns with the growing demand for summarization technologies to adapt to the evolving nature of digital content consumption.

Moreover, multimodal summarization not only broadens the types of information that can be summarized but also enhances contextual understanding. By considering multiple modalities, summarization models can discern intricate relationships between textual and visual elements, resulting in more nuanced and contextually rich summaries. This trend emphasizes the importance of capturing the complete narrative and meaning from diverse content sources, fostering the development of summarization systems that are not only versatile across modalities but also adept at providing a comprehensive overview of the intricate relationships within the content. As technology advances, research in multimodal summarization is likely to explore innovative techniques to harness the synergies between different modalities, leading to more sophisticated and adaptable summarization systems poised for diverse applications.

10. Conclusion

In summary, the persistent emphasis on the importance of semantic understanding in current and future text summarization techniques underscores its foundational role in unlocking the true potential of summarization systems. Semantic understanding allows models to transcend mere syntactic structures, delving into the intricate layers of meaning within textual content. This emphasis becomes even more critical in the contemporary landscape, where the sheer diversity and complexity of textual data demands a more sophisticated approach to summarization. Recognizing and prioritizing semantic comprehension ensures that summarization techniques not only generate concise summaries but also capture the nuanced context and deeper insights embedded in the text.

Encouraging continued research in the field is paramount to overcoming existing challenges and driving innovation in text summarization. Existing hurdles, such as handling ambiguity, context disambiguation, and navigating diverse linguistic nuances, require ongoing attention and exploration. Researchers are urged to delve into advanced natural language processing techniques, semantic modeling, and novel approaches that synergize multiple linguistic dimensions. This encouragement aims to foster a culture of curiosity and exploration where scholars push the boundaries of what is achievable in text summarization, contributing to the refinement and evolution of current methodologies.

Moreover, the call for ongoing research aligns with the evolving nature of textual data and the dynamic requirements of diverse applications. As information sources continue to diversify and evolve, text summarization techniques must adapt to provide accurate and contextually rich summaries across various domains. Researchers are encouraged to explore hybrid models that seamlessly integrate semantic understanding with other advanced techniques, creating more versatile and adaptive summarization systems.

The journey towards advancing text summarization techniques is a collective endeavor that involves collaboration, knowledge sharing, and a commitment to overcoming the inherent challenges. By fostering a community-driven research culture, the field can collectively address semantic complexities, enhance contextual awareness, and explore innovative solutions that resonate with the evolving demands of users and the ever-expanding scope of textual information. In conclusion, the reiteration of the importance of semantic understanding serves not only as a guiding principle but as an invitation for researchers to embark on a continuous quest for excellence in the field of text summarization.

CRediT authorship contribution statement

Supriyono: Writing – original draft, Data curation, Conceptualization. **Aji Prasetya Wibawa:** Supervision, Investigation, Formal analysis. **Suyono:** Writing – review & editing, Methodology, Conceptualization. **Fachrul Kurniawan:** Writing – review & editing, Visualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Supriyono reports administrative support was provided by State University of Malang. Aji Prasetya Wibawa reports administrative support was provided by State University of Malang. Co-author is my supervisor on the Doctoral Programme in Electrical Engineering and Informatics at Malang State University

References

Abdeljaber, H.A.M., Ahmad, S., Alharbi, A., Kumar, S., 2022. XAI-based reinforcement learning approach for text summarization of social IoT-based content. Secur. Commun. Netw. 2022, 1–12. http://dx.doi.org/10.1155/2022/7516832.

Abdi, A., Idris, N., Alguliev, R.M., Aliguliyev, R.M., 2015. Automatic summarization assessment through a combination of semantic and syntactic information for intelligent educational systems. Inf. Process. Manage. 51 (4), 340–358. http://dx.doi.org/10.1016/j.ipm.2015.02.001.

Abidin, A.Z.Z., Murdianingsih, Y., Suryadi, U.T., Setiyadi, D., 2020. Text summarizing system of English subjects and text mining subjects for computer science students. J. Crit. Rev. 7 (5), 730–742. http://dx.doi.org/10.31838/jcr.07.05.150.

Abualigah, L., Bashabsheh, M.Q., Alabool, H., Shehab, M., 2020. Text summarization: A brief review. Stud. Comput. Intell. 874 (January), 1–15. http://dx.doi.org/10.1007/978-3-030-34614-0_1.

Ahmad, A., 2014. Text summarization based on several natural language techniques. https://api.semanticscholar.org/CorpusID:56391076.

AL-Khassawneh, Y.A., 2020. The use of semantic role labelling with triangle-graph based text summarization. Int. J. Emerg. Trends Eng. Res. 8 (4), 1162–1169. http://dx.doi.org/10.30534/ijeter/2020/34842020.

Alam, H., Kumar, A., Nakamura, M., Rahman, F., Tarnikova, Y., Wilcox, Che, 2003. Structured and unstructured document summarization:design of a commercial summarizer using Lexical chains. In: Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings. Vol. 1, pp. 1147–1152. http://dx.doi.org/10.1109/ICDAR.2003.1227836.

Alami, N., Mallahi, M. El, Amakdouf, H., Qjidaa, H., 2021. Hybrid method for text summarization based on statistical and semantic treatment. Multimedia Tools Appl. 80 (13), 19567–19600. http://dx.doi.org/10.1007/s11042-021-10613-9.

Alias, S., 2021. Unsupervised text feature extraction for academic chatbot using constrained FP-growth. ASM Sci. J. 14, 1–11. http://dx.doi.org/10.32802/asmscj.2020.576.

ALJa'am, J.M., Jaoua, A.M., Hasnah, A.M., Hassan, F., Mohamed, H., Mosaid, T., Saleh, H., Abdullah, F., Cherif, H., 2006. Text summarization based on conceptual data classification. Int. J. Inf. Technol. Web Eng. 1 (4), 22–36. http://dx.doi.org/10.4018/jitwe.2006100102.

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S.D.E.B.J., Kochut, K., 2017. Text summarization techniques: A brief survey. Int. J. Adv. Comput. Sci. Appl. 8 (10), http://dx.doi.org/10.14569/ijacsa.2017.081052.

Amato, F., D'Acierno, A., Colace, F., Moscato, V., Penta, A., Picariello, A., 2017a. Semantic Summarization of News from Heterogeneous Sources. pp. 305–314. http://dx.doi.org/10.1007/978-3-319-49109-7_29.

Amato, F., Moscato, V., Picariello, A., Sperlí, G., D'Acierno, A., Penta, A., 2017b. Semantic summarization of web news. In: Encyclopedia with Semantic Computing and Robotic Intelligence. Vol. 01, (01), http://dx.doi.org/10.1142/S2425038416300068, 1630006.

Andhale, N., Bewoor, L.A., 2016. An overview of text summarization techniques. In: 2016 International Conference on Computing Communication Control and Automation. ICCUBEA, pp. 1–7. http://dx.doi.org/10.1109/ICCUBEA.2016.7860024.

Antony, D., Abhishek, S., Singh, S., Kodagali, S., Darapaneni, N., Rao, M., Paduri, A.R., BG, S., 2023. A survey of advanced methods for efficient text summarization. In: 2023 IEEE 13th Annual Computing and Communication Workshop and Conference. CCWC, pp. 0962–0968. http://dx.doi.org/10.1109/CCWC57344.2023.10099322.

Bagalkotkar, A., Kandelwal, A., Pandey, S., Kamath, S.S., 2013. A novel technique for efficient text document summarization as a service. In: 2013 Third International Conference on Advances in Computing and Communications. pp. 50–53. http://dx.doi.org/10.1109/ICACC.2013.17.

Baralis, E., Cagliero, L., Jabeen, S., Fiori, A., Shah, S., 2014. Combining Semantics and Social Knowledge for News Article Summarization. pp. 209–230. http://dx.doi.org/10.4018/978-1-4666-6086-1.ch012.

Batista, J., Ferreira, R., Tomaz, H., Ferreira, R., Dueire Lins, R., Simske, S., Silva, G., Riss, M., 2015. A quantitative and qualitative assessment of automatic text summarization systems. In: Proceedings of the 2015 ACM Symposium on Document Engineering. pp. 65–68. http://dx.doi.org/10.1145/2682571.2797081.

Benharrak, K., Lehmann, F., Dang, H., Buschek, D., 2022. SummaryLens – A smartphone app for exploring interactive use of automated text summarization in everyday life. In: 27th International Conference on Intelligent User Interfaces. pp. 93–96. http://dx.doi.org/10.1145/3490100.3516471.

Bhaskar, A., Fabbri, A., Durrett, G., 2023. Prompted opinion summarization with GPT-3.5. In: Findings of the Association for Computational Linguistics: ACL 2023. pp. 9282–9300. http://dx.doi.org/10.18653/v1/2023.findings-acl.591.

Bhatia, N., Jaiswal, A., 2016. Automatic text summarization and it's methods - a review. In: 2016 6th International Conference - Cloud System and Big Data Engineering (Confluence). pp. 65–72. http://dx.doi.org/10.1109/CONFLUENCE.2016.7508049.

Bhola, A., Mullapudi, J., Kollipara, S., Sanaka, T., 2022. Text summarization based on ranking techniques. In: 2022 5th International Conference on Contemporary Computing and Informatics. IC3I, pp. 1463–1467. http://dx.doi.org/10.1109/IC3I56241.2022.10072962.

Bi, Q., Li, H., Lu, K., Yang, H., 2021. Augmented abstractive summarization with document-level semantic graph. J. Data Sci. 45, 0–464. http://dx.doi.org/10.6339/21-JDS1012.

Bin Mohd Amin, M.R., Piaralal, S.K., 2020. Antecedents and outcomes of service recovery satisfaction: Perspectives on open and distance learning in Malaysia. Int. J. Bus. Innov. Res. 21 (1), 56–78. http://dx.doi.org/10.1504/IJBIR.2020.104034.

Biswas, S., Rautray, R., Dash, R., Dash, R., 2018. Text summarization: A review. In: 2018 2nd International Conference on Data Science and Business Analytics. ICDSBA, pp. 231–235. http://dx.doi.org/10.1109/ICDSBA.2018.00048.

Cao, M., Zhuge, H., 2019. Automatic evaluation of text summarization based on semantic link network. In: 2019 15th International Conference on Semantics, Knowledge and Grids. SKG, pp. 107–114. http://dx.doi.org/10.1109/SKG49510.2019.00026.

Cardoso, P.C.F., Pardo, T.A.S., 2016. Multi-document summarization using semantic discourse models. In: Procesamiento Del Lenguaje Natural. https://api.semanticscholar.org/CorpusID:52596330.

Chali, Y., Mahmud, A., 2021. Query-based summarization using reinforcement learning and transformer model. In: 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence. ICTAI, pp. 129–136. http://dx.doi.org/10.1109/ICTAI52525.2021.00026.

Chatterjee, N., Agarwal, R., 2023. Studying the effect of syntactic simplification on text summarization. IETE Techn. Rev. (Inst. Electron. Telecommun. Eng., India) 40 (2), 155–166. http://dx.doi.org/10.1080/02564602.2022.2055670.

Chen, H., Zhang, Z., Huang, S., Hu, J., Ni, W., Liu, J., 2023. Textcnn-based ensemble learning model for Japanese text multi-classification. Comput. Electr. Eng. 109, 108751. http://dx.doi.org/10.1016/j.compeleceng.2023.108751.

Cheng, J., Zhang, F., Guo, X., 2020. A syntax-augmented and headline-aware neural text summarization method. IEEE Access 8, 218360–218371. http://dx.doi.org/10.1109/ACCESS.2020.3042886.

Chettri, R., Kr, U., 2017. Automatic text summarization. Int. J. Comput. Appl. 161 (1), 5–7. http://dx.doi.org/10.5120/ijca2017912326.

Chitty-Venkata, K.T., Emani, M., Vishwanath, V., Somani, A.K., 2022. Neural architecture search for transformers: A survey. IEEE Access 10, 108374–108412. http://dx.doi.org/10.1109/ACCESS.2022.3212767.

Choon-Ching, N., Selamat, A., 2008. Text summarization review. https://api.semanticscholar.org/CorpusID:16400901.

Cranganu-Cretu, B., Chen, Z., Uchimoto, T., Miya, K., 2002. Automatic text summarizing based on sentence extraction: A statistical approach. Int. J. Appl. Electromagn. Mech. 13 (1–4), 19–23. http://dx.doi.org/10.3233/JAE-2002-513.

Dalal, V., Malik, L., 2013. A survey of extractive and abstractive text summarization techniques. In: 2013 6th International Conference on Emerging Trends in Engineering and Technology. pp. 109–110. http://dx.doi.org/10.1109/ICETET.2013.31.

Dalianis, H., Hassel, M., 2022. From SweSum to ScandSum- Automatic text summarization for the Scandinavian languages. pp. 1–7.

Dave, H., Jaswal, S., 2015. Multiple text document summarization system using hybrid summarization technique. In: 2015 1st International Conference on Next Generation Computing Technologies. NGCT, pp. 804–808. http://dx.doi.org/10.1109/NGCT.2015.7375231.

Dedhia, P.R., Pachgade, H.P., Malani, A.P., Raul, N., Naik, M., 2020. Study on abstractive text summarization techniques. In: 2020 International Conference on Emerging Trends in Information Technology and Engineering. Ic-ETITE, pp. 1–8. http://dx.doi.org/10.1109/ic-ETITE47903.2020.087.

Deng, H., Ergu, D., Liu, F., Cai, Y., Ma, B., 2022. Text sentiment analysis of fusion model based on attention mechanism. Procedia Comput. Sci. 199, 741–748. http://dx.doi.org/10.1016/j.procs.2022.01.092.

Deroy, A., Ghosh, K., Ghosh, S., 2023. Ensemble methods for improving extractive summarization of legal case judgements. Artif. Intell. Law http://dx.doi.org/10.1007/s10506-023-09349-8.

Dhawale, A.D., Kulkarni, S.B., Kumbhakarna, V., 2020. Survey of Progressive Era of Text Summarization for Indian and Foreign Languages Using Natural Language Processing. pp. 654–662. http://dx.doi.org/10.1007/978-3-030-38040-3_74.

Dong, L., Satpute, M.N., Wu, W., Du, D.-Z., 2021. Two-phase multidocument summarization through content-attention-based subtopic detection. IEEE Trans. Comput. Soc. Syst. 8 (6), 1379–1392. http://dx.doi.org/10.1109/TCSS.2021.3079206.

Dr. Vidyagouri, B.H., BibiSadiqa, M.D., 2022. Text summarization using machine learning algorithm. Int. J. Scient. Res. Comput. Sci., Eng. Inf. Technol. 16, 7–173. http://dx.doi.org/10.32628/CSEIT228421.

Faizal, B., Abraham, S., 2022. NLP based automated business report summarization. In: 2022 International Conference on Innovative Trends in Information Technology. ICITIIT 2022, pp. 1–4. http://dx.doi.org/10.1109/ICITIIT54346.2022.9744151.

Fang, Z., He, Y., Procter, R., 2023. BERTTM: Leveraging contextualized word embeddings from pre-trained language models for neural topic modeling. arXiv, abs/2305.0. https://api.semanticscholar.org/CorpusID:258714724.

Feijo, D., De, V., Moreira, V.P., 2023. Improving abstractive summarization of legal rulings through textual entailment. Artif. Intell. Law 31 (1), 91–113. http://dx.doi.org/10.1007/s10506-021-09305-4.

Ferreira, R., Freitas, F., Cabral, L., De, S., Lins, R.D., Lima, R., Franca, G., Simske, S.J., Favaro, L., 2014. A context based text summarization system. In: 2014 11th IAPR International Workshop on Document Analysis Systems. pp. 66–70. http://dx.doi.org/10.1109/DAS.2014.19.

Foong, O.-M., Yong, S.-P., Lee, A.-L., 2014. Text Summarization in Android Mobile Devices. pp. 571–578. http://dx.doi.org/10.1007/978-981-4585-18-7_64.

Gaikwad, D.K., Mahender, C.N., 2016. A review paper on text summarization. Int. J. Adv. Res. Comput. Commun. Eng. (IJARCCE) 5 (3), 154–160. http://dx.doi.org/10.17148/IJARCCE.2016.5340.

Ganesh, A., Jaya, A., Sunitha, C., 2022. An overview of semantic based document summarization in different languages. ECS Trans. 107 (1), 6007–6017. http://dx.doi.org/10.1149/10701.6007ecst.

Gao, Y., Xu, Y., Huang, H., Liu, Q., Wei, L., Liu, L., 2020. Jointly learning topics in sentence embedding for document summarization. IEEE Trans. Knowl. Data Eng. 32 (4), 688–699. http://dx.doi.org/10.1109/TKDE.2019.2892430.

Gianey, H.K., Choudhary, R., 2017. Comprehensive review on supervised machine learning algorithms. In: Proceedings - 2017 International Conference on Machine Learning and Data Science. MLDS 2017 2018-Janua, pp. 38–43. http://dx.doi.org/10.1109/MLDS.2017.11.

Goldstein, J., Mittal, V., Carbonell, J., Kantrowitz, M., 2000. Multi-document summarization by sentence extraction. In: NAACL-ANLP 2000 Workshop on Automatic Summarization. Vol. 4, pp. 40–48. http://dx.doi.org/10.3115/1117575.1117580.

Gu, Y., Gui, X., Shen, Y., Liao, D., 2019. An empirical evaluation on word embeddings across reading comprehension. In: 2019 IEEE 11th International Conference on Advanced Infocomm Technology. ICAIT, pp. 157–161. http://dx.doi.org/10.1109/ICAIT.2019.8935932.

Guadalupe Ramos, J., Navarro-Alatorre, I., Flores Becerra, G., Flores-Sánchez, O., 2019. A formal technique for text summarization from web pages by using latent semantic analysis. Res. Comput. Sci. 148 (3), 11–22. http://dx.doi.org/10.13053/rcs-148-3-1.

Gupta, H., Kottwani, A., Gogia, S., Chaudhari, S., 2016. Text analysis and information retrieval of text data. In: 2016 International Conference on Wireless Communications, Signal Processing and Networking. WiSPNET, pp. 788–792. http://dx.doi.org/10.1109/WiSPNET.2016.7566241.

Gupta, H., Patel, M., 2020. Study of extractive text summarizer using the elmo embedding. In: 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud). I-SMAC, pp. 829–834. http://dx.doi.org/10.1109/I-SMAC49090.2020.9243610.

Gupta, S., Sharaff, A., Nagwani, N.K., 2023. Frequent item-set mining and clustering based ranked biomedical text summarization. J. Supercomput. 79 (1), 139–159. http://dx.doi.org/10.1007/s11227-022-04578-1.

Haas, C., 2013. Writing Technology. Routledge, http://dx.doi.org/10.4324/9780203811238.

Haggag, M.H., 2013. Semantic text summarization based on syntactic patterns. Int. J. Inf. Retrieval Res. 3 (4), 18–34. http://dx.doi.org/10.4018/ijirr.2013100102.

Haider, M.M., Hossin, M.A., Mahi, H.R., Arif, H., 2020. Automatic text summarization using gensim Word2Vec and K-means clustering algorithm. In: 2020 IEEE Region 10 Symposium. TENSYMP, pp. 283–286. http://dx.doi.org/10.1109/TENSYMP50017.2020.9230670.

Hamid, F., Tarau, P., 2014. Text summarization as an assistive technology. In: Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments. pp. 1–4. http://dx.doi.org/10.1145/2674396.2674440.

Hassel, M., 2004. Summaries and the Process of Summarization from Evaluation of Automatic Text Summarization -a Practical Implementation. https://api.semanticscholar.org/CorpusID:17363819.

Hegdepatil, P., Davuluri, K., 2021. Business intelligence based novel marketing strategy approach using automatic speech recognition and text summarization. In: 2021 2nd International Conference on Computing and Data Science. CDS, pp. 595–602. http://dx.doi.org/10.1109/CDS52072.2021.00108.

Hong, Y.X., Chang, C.H., 2023. Improving colloquial case legal judgment prediction via abstractive text summarization. Comput. Law Secur. Rev. 51, 105863. http://dx.doi.org/10.1016/j.clsr.2023.105863.

Hovy, E., Lin, C.-Y., 1996. Automated Text Summarization and the SUMMARIST System. Vol. 197, http://dx.doi.org/10.3115/1119089.1119121.

Humera Khanam, M., Sravani, S., 2016. Text summarization for telugu document. Inform. Sci. 18 (6), 25–28. http://dx.doi.org/10.9790/0661-1806052528.

Indra, Z., Jusman, Y., Winarso, D., 2020. Text summarization application for Indonesian Twitter document by using top-N feature selection algorithm. In: 2020 1st International Conference on Information Technology, Advanced Mechanical and Electrical Engineering. ICITAMEE, pp. 238–243. http://dx.doi.org/10.1109/ICITAMEE50454.2020.9398413.

Israel, Q., Han, H., Song, I.-Y., 2015. Semantic analysis for focused multi-document summarization (fMDS) of text. In: Proceedings of the 30th Annual ACM Symposium on Applied Computing. pp. 339–344. http://dx.doi.org/10.1145/2695664.2695672.

Jadon, M.K., Pareek, A., 2016. A Method for Automatic Text Summarization Using Consensus of Multiple Similarity Measures and Ranking Techniques. ICON. https://api.semanticscholar.org/CorpusID:6310788.

Jafarinejad, F., 2023. Synset2Node: A new synset embedding based upon graph embeddings. Intell. Syst. Appl. 17, 200159. http://dx.doi.org/10.1016/j.iswa.2022.200159.

Jain, A., Bhatia, D., Thakur, M.K., 2017. Extractive text summarization using word vector embedding. In: 2017 International Conference on Machine Learning and Data Science. MLDS, pp. 51–55. http://dx.doi.org/10.1109/MLDS.2017.12.

Jain, D., Borah, M.D., Biswas, A., 2021. Summarization of legal documents: Where are we now and the way forward. Comp. Sci. Rev. 40, 100388. http://dx.doi.org/10.1016/j.cosrev.2021.100388.

Janaki Raman, K., Meenakshi, K., 2021. Automatic Text Summarization of Article (NEWS) Using Lexical Chains and WordNet—A Review. pp. 271–282. http://dx.doi.org/10.1007/978-981-15-5329-5_26.

Jeng, Y.-L., Huang, Y.-M., Huang, T.-C., 2017. Using Summarization Technology for Supporting Problem-Based Learning. pp. 123–131. http://dx.doi.org/10.1007/978-3-319-52836-6_15.

Jezek, K., Steinberger, J., 2008. Automatic summarizing: (The state of the art 2007 and new challenges). In: Proceedings of Znalosti. pp. 1–12.

Jiang, J., Zhang, H., Dai, C., Zhao, Q., Feng, H., Ji, Z., Ganchev, I., 2021. Enhancements of attention-based bidirectional LSTM for hybrid automatic text summarization. IEEE Access 9, 123660–123671. http://dx.doi.org/10.1109/ACCESS.2021.3110143.

Jing, R., Wan, F., 2023. Research on automatic text summarization based on extractive method. In: 2023 International Conference on Computer Graphics and Image Processing. CGIP, pp. 72–77. http://dx.doi.org/10.1109/CGIP58526.2023.00022.

Jung, C., Yoon, W.C., Datta, R., Jung, S., 2021. Knowledge base driven automatic text summarization using multi-objective optimization. Int. J. Adv. Comput. Sci. Appl. 12 (8), 836–849. http://dx.doi.org/10.14569/IJACSA.2021.0120895.

Kaikhah, K., 2004. Automatic text summarization with neural networks. In: 2004 2nd International IEEE Conference Intelligent Systems - Proceedings. Vol. 1, pp. 40–44. http://dx.doi.org/10.1109/is.2004.1344634.

Kanapala, A., Pal, S., Pamula, R., 2019. Text summarization from legal documents: a survey. Artif. Intell. Rev. 51 (3), 371–402. http://dx.doi.org/10.1007/s10462-017-9566-2.

Karmaker, P.C., Hossen, M.S., 2019. Performance analysis of frequency and graph theoretic based text summarization. In: 2019 International Conference on Electrical, Computer and Communication Engineering. ECCE, pp. 1–5. http://dx.doi.org/10.1109/ECACE.2019.8679452.

Karpagam, K., Saradha, A., Manikandan, K., Madusudanan, K., 2020. Enhancement of single document text summarization using reinforcement learning with non-deterministic rewards. Int. J. Inf. Technol. Comput. Sci. 12 (4), 19–27. http://dx.doi.org/10.5815/ijitcs.2020.04.03.

Kaszas, V., Tundik, M.A., Szaszak, G., 2018. A semantic space approach for automatic summarization of documents. In: 2018 9th IEEE International Conference on Cognitive Infocommunications. CogInfoCom, pp. 000153–000158. http://dx.doi.org/10.1109/CogInfoCom.2018.8639891.

Ke, X., Khan, A., Bonchi, F., 2022. Multi-relation graph summarization. ACM Trans. Knowl. Discov. Data 16 (5), 1–30. http://dx.doi.org/10.1145/3494561.

Keneshloo, Y., Shi, T., Ramakrishnan, N., Reddy, C.K., 2019. Deep reinforcement learning for sequence-to-sequence models. IEEE Trans. Neural Netw. Learn. Syst. 1–21. http://dx.doi.org/10.1109/TNNLS.2019.2929141.

Kipp, D., 2008. Shallow semantics for topic-oriented multi-document automatic text summarization. https://api.semanticscholar.org/CorpusID:65193337.

Kireyev, K., 2008. Using latent semantic analysis for extractive summarization. In: Theory and Applications of Categories. https://api.semanticscholar.org/CorpusID:20304397.

Kirmani, M., Shukla, A.K., 2022. Systematic review of methods used in text summarization. In: 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering. ICACITE 2022, pp. 1048–1052. http://dx.doi.org/10.1109/ICACITE53722.2022.9823831.

Kohakade, P., Jadhav, S., 2020. Text Summarization Framework Using Advanced Machine Learning Algorithms. https://api.semanticscholar.org/CorpusID:215995289.

Kouris, P., Alexandridis, G., Stafylopatis, A., 2019. Abstractive text summarization based on deep learning and semantic content generalization. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 5082–5092. http://dx.doi.org/10.18653/v1/P19-1501.

Kruse, O., Rapp, C., Anson, C.M., Benetos, K., Cotos, E., Devitt, A., 2023. Digital Writing Technologies in Higher Education. http://dx.doi.org/10.1007/978-3-031-36033-6.

Kulkarni, A.A., 2013. An automatic Text Summarization using feature terms for relevance measure. IOSR J. Comput. Eng. 9 (3), 62–66. http://dx.doi.org/10.9790/0661-0936266.

Kumar, S., 2019. Machine learning (supervised). In: International Series in Operations Research and Management Science. Vol. 264, pp. 507–568. http://dx.doi.org/10.1007/978-3-319-68837-4_16.

Kumari, A., Raviraj, A., Murari, A.K., Mukherjee, I., 2022. Real life implementation of text summarization technique. In: 2022 9th International Conference on Computing for Sustainable Global Development. INDIACom, pp. 69–73. http://dx.doi.org/10.23919/INDIACom54597.2022.9763270.

Kumari, N., Singh, P., 2021. Text Summarization and Its Types. pp. 368–378. http://dx.doi.org/10.4018/978-1-7998-4730-4.ch017.

Lasya Sriranga, G., Likitha, P., Meghana, B., Jayanthi, N., 2020. Efficient text summarizer using point to generator technique. Int. J. Eng. Appl. Sci. Technol. 5 (1), 488–492. http://dx.doi.org/10.33564/ijeast.2020.v05i01.086.

Lee, G.H., Lee, K.J., 2017. Automatic text summarization using reinforcement learning with embedding features. In: Proceedings of the 8th International Joint Conference on Natural Language Processing. Vol. 2, pp. 193–197, https://api.semanticscholar.org/CorpusID:3135474.

Lehmam, A., 2010. Essential summarizer: innovative automatic text summarization software in twenty languages. In: RIAO Conference. https://api.semanticscholar.org/CorpusID:22689867.

Li, G., Li, J., 2018. Research on sentiment classification for tang poetry based on TF-IDF and FP-growth. In: 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference. IAEAC, pp. 630–634. http://dx.doi.org/10.1109/IAEAC.2018.8577715.

Li, H., Peng, Q., Mou, X., Wang, Y., Zeng, Z., Bashir, M.F., 2023. Abstractive financial news summarization via transformer-BiLSTM encoder and graph attention-based decoder. IEEE/ACM Trans. Audio, Speech, Lang. Process. 31, 3190–3205. http://dx.doi.org/10.1109/TASLP.2023.3304473.

Li, Z., Peng, Z., Tang, S., Zhang, C., Ma, H., 2020. Text summarization method based on double attention pointer network. IEEE Access 8, 11279–11288. http://dx.doi.org/10.1109/ACCESS.2020.2965575.

Liao, D., 2021. Sentence embeddings using supervised contrastive learning. arXiv, abs/2106.0. https://api.semanticscholar.org/CorpusID:235377036.

Liberati, A., 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. Ann. Intern. Med. 151 (4), http://dx.doi.org/10.7326/0003-4819-151-4-200908180-00136, W.

Liu, S., 2009. Experiences with and reflections on text summarization tools. Int. J. Comput. Intell. Syst. 2 (3), 202–218. http://dx.doi.org/10.1080/18756891.2009.9727654.

Liu, Y., Che, W., Wang, Y., Zheng, B., Qin, B., Liu, T., 2019. Deep contextualized word embeddings for universal dependency parsing. ACM Trans. Asian Low-Resource Lang. Inf. Process. (TALLIP) 19, 1–17, https://api.semanticscholar.org/CorpusID:199009034.

Lloret, E., 2008. Text summarization : an overview.

Loukachevitch, N., 1998. Text summarization based on thematic representation of texts. Inform. Sci. 6, 8–76.

Lu, F., Cong, P., Huang, X., 2020. Utilizing textual information in knowledge graph embedding: A survey of methods and applications. IEEE Access 1. http://dx.doi.org/10.1109/ACCESS.2020.2995074.

Magesh, N.S., Ramya, T., 2020. Improving text summarization using latent semantic analysis. Softw. Eng. Technol. 12, 25–30, https://api.semanticscholar.org/CorpusID:215924812.

Mahajani, A., Pandya, V., Maria, I., Sharma, D., 2019. Ranking-Based Sentence Retrieval for Text Summarization. pp. 465–474. http://dx.doi.org/10.1007/978-981-13-2414-7_43.

Malagi, S.S., Radhakrishnan, R., 2020. An overview of automatic text summarization techniques. Inform. Sci. 8 (15), 99–105, https://www.ijert.org/.

Mani, I., 2001. Recent developments in text summarization. In: Proceedings of the Tenth International Conference on Information and Knowledge Management. pp. 529–531. http://dx.doi.org/10.1145/502585.502677.

Martin, L., 2021. Automatic Sentence Simplification Using Controllable and Unsupervised Methods [Sorbonne Université]. Computation and Language, Sorbonne Université, (Issue 2021SORUS265). https://theses.hal.science/tel-03543971.

Mastronardo, C., Tamburini, F., 2019. Enhancing a text summarization system with ELMo. In: Italian Conference on Computational Linguistics. https://api.semanticscholar.org/CorpusID:204899995.

Mazzieri, M., Topi, S., Dragoni, A.F., Vallesi, G., 2010. Modeling contextualized textual knowledge as a Long-Term Working Memory. In: The European Symposium on Artificial Neural Networks. https://api.semanticscholar.org/CorpusID:18045194.

McLellan, P., Tombros, A., Jose, J.M., Ounis, I., Whitehead, M., 2001. Evaluating summarisation technologies: A task oriented approach. In: New Developments in Digital Libraries. https://api.semanticscholar.org/CorpusID:6317110.

Meier, R., Mujika, A., 2022. Open-ended reinforcement learning with neural reward functions. In: Advances in Neural Information Processing Systems, vol. 35, https://api.semanticscholar.org/CorpusID:246904679.

Mirani, T.B., Sasi, S., 2017. Two-level text summarization from online news sources with sentiment analysis. In: 2017 International Conference on Networks & Advances in Computational Technologies. NetACT, pp. 19–24. http://dx.doi.org/10.1109/NETACT.2017.8076735.

Mohamed, M.A., 2016. Automatic text summarisation using linguistic knowledge-based semantics. https://api.semanticscholar.org/CorpusID:63350540.

Mohamed, M., Oussalah, M., Chang, V., 2023. SDbQfSum: Query-focused summarization framework based on diversity and text semantic analysis. Expert Syst. http://dx.doi.org/10.1111/exsy.13462.

MohammedBadry, R., Sharaf Eldin, A., Saad Elzanfally, D., 2013. Text summarization within the latent semantic analysis framework: Comparative study. Int. J. Comput. Appl. 81 (11), 40–45. http://dx.doi.org/10.5120/14060-2366.

Mohd, M., Jan, R., Shah, M., 2020. Text document summarization using word embedding. Expert Syst. Appl. 143, 112958. http://dx.doi.org/10.1016/j.eswa.2019.112958.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A., 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. System. Rev. 4 (1), 1. http://dx.doi.org/10.1186/2046-4053-4-1.

Moiyadi, H.S., Pawar, D., 2017. Semantic analysis based text summarization. https://api.semanticscholar.org/CorpusID:212571066.

Motilal Lodhi, P., Kharche, S., Kambri, D., Saleem Khan, S., 2022. Business meeting summarization system. In: 2022 2nd Asian Conference on Innovation in Technology. ASIANCON, pp. 1–6. http://dx.doi.org/10.1109/ASIANCON55314.2022.9908905.

Mridha, M.F., Lima, A.A., Nur, K., Das, S.C., Hasan, M., Kabir, M.M., 2021. A survey of automatic text summarization: Progress, process and challenges. IEEE Access 9, 156043–156070. http://dx.doi.org/10.1109/ACCESS.2021.3129786.

Mu, H., Zhang, S., Li, Q., Liu, T., Xu, H., 2023. Adapt-to-learn policy network for abstractive multi-document summarization. In: 2023 International Joint Conference on Neural Networks. IJCNN, pp. 1–8. http://dx.doi.org/10.1109/IJCNN54540.2023.10191102.

Muthiah, D.K., 2019. Automatic Coherent and Concise Text Summarization using Natural Language Processing MSc Research Project Kaarthic Muthiah School of Computing Supervisor : Divyaa Manimaran Elango.

Mutlu, B., Sezer, E.A., Akcayol, M.A., 2019. Multi-document extractive text summarization: A comparative assessment on features. Knowl.-Based Syst. 183, 104848. http://dx.doi.org/10.1016/j.knosys.2019.07.019.

Nazar, N., Hu, Y., Jiang, H., 2016. Summarizing software artifacts: A literature review. J. Comput. Sci. Tech. 31 (5), 883–909. http://dx.doi.org/10.1007/s11390-016-1671-1.

Nenkova, A., 2005. Automatic text summarization of newswire: Lessons learned from the document understanding conference. In: Proceedings of the National Conference on Artificial Intelligence. Vol. 3, pp. 1436–1441.

Nguyen, P.T., Dang, C.T.M., 2018. Statistical models to automatic text summarization. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): 11251 LNCS, pp. 486–498. http://dx.doi.org/10.1007/978-3-030-03192-3_37.

Niu, J., Sun, M., Rodrigues, J.J.P.C., Liu, X., 2019. A novel attention mechanism considering decoder input for abstractive text summarization. In: ICC 2019-2019 IEEE International Conference on Communications. ICC, pp. 1–7. http://dx.doi.org/10.1109/ICC.2019.8762040.

Okurowski, M.E., Wilson, H., Urbina, J., Taylor, T., Clark, R.C., Krapcho, F., 2000. Text summarizer in use. Inform. Sci. 4, 9–58. http://dx.doi.org/10.3115/1117575.1117581.

Ou, S., 2009. Text Summarization in Digital Libraries: Development and Evaluation of a Multi-Document Summarization Method for Research Abstracts. VDM Verlag.

PadmaLahari, E., Kumar, D.V.N.S., Prasad, S., 2014. Automatic text summarization with statistical and linguistic features using successive thresholds. In: 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies. pp. 1519–1524. http://dx.doi.org/10.1109/ICACCCT.2014.7019360.

Parmar, C., Chaubey, R., Bhatt, K., 2019. Abstractive text summarization using artificial intelligence. SSRN Electr. J. http://dx.doi.org/10.2139/ssrn.3370795.

Parums, D.V., 2021. Editorial: review articles, systematic reviews, meta-analysis, and the updated preferred reporting items for systematic reviews and meta-analyses (PRISMA) 2020 guidelines. Med. Sci. Monit. 27, http://dx.doi.org/10.12659/MSM.934475.

Patel, N., Mangaokar, N., 2020. Abstractive vs extractive text summarization (output based approach) - A comparative study. In: 2020 IEEE International Conference for Innovation in Technology. INOCON, pp. 1–6. http://dx.doi.org/10.1109/INOCON50539.2020.9298416.

Pei-ying, Z., 2009. Evaluation method of automatic summarization based on semantic similarity. Comput. Eng. Appl. https://api.semanticscholar.org/CorpusID:63671560.

Pokhrel, C., Adhikari, R., 2023. Automatic extractive text summarization for text in nepali language with bidirectional encoder representation transformers and K-Mean clustering. http://dx.doi.org/10.13140/RG.2.2.15823.87200, July.

Prakhar, Sethi, Saumitra, Khanwalker, 2020. Survey and analysis on automatic text summarization methods. https://api.semanticscholar.org/CorpusID:233218374.

Pramoda Devi, 2020. A comparative study on various approaches and complexities of text summarization. Inform. Sci. 10172, 10172–10183.

Pramudita, Y.D., Suzanti, I.O., Saifuddin Syarief, M., Solihin, F., 2022. Automatic text summarization of madura tourism articles using TF-IDF and K-medoid clustering. In: Proceeding - IEEE 8th Information Technology International Seminar. ITIS 2022, pp. 168–172. http://dx.doi.org/10.1109/ITIS57155.2022.10009983.

Purushotham Reddy, M., Sreekar, V., Srikanth, M., Siddartha, K., 2021. Text summarization of Telugu scripts. In: 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud). I-SMAC, pp. 1060–1064. http://dx.doi.org/10.1109/I-SMAC52330.2021.9640729.

Qi, S., Zhang, H., 2023. Text summarization quality detection based on GPT-3. Appl. Comput. Eng. 8 (1), 817–822. http://dx.doi.org/10.54254/2755-2721/8/20230251.

Quishpi, L., Carmona, J., Padró, L., 2020. Extracting Annotations from Textual Descriptions of Processes. pp. 184–201. http://dx.doi.org/10.1007/978-3-030-58666-9_11.

Rajalakshmi, R., Vidhya, S., Harina, D., Karna, R., Sowmya, A., 2023. Text summarization for news articles using latent semantic analysis technique. In: 2023 4th International Conference on Electronics and Sustainable Communication Systems, ICESC 2023 - Proceedings. pp. 1421–1425. http://dx.doi.org/10.1109/ICESC57686.2023.10193508.

Rajasekaran, A., R. Varalakshmi, D., 2018. Review on automatic text summarization. Int. J. Eng. Technol. 7 (3.3), 456. http://dx.doi.org/10.14419/ijet.v7i2.33.14210.

Rananavare, L.B., Reddy, P.V.S., 2017. An overview of text summarization. Int. J. Comput. Appl. 171 (10), 1–17.

Rane, N., Govilkar, S., 2019. Recent trends in deep learning based abstractive text summarization. Int. J. Recent Technol. Eng. 8 (3), 3108–3115. http://dx.doi.org/10.35940/ijrte.C4996.098319.

Raundale, P., Shekhar, H., 2021. Analytical study of text summarization techniques. In: 2021 Asian Conference on Innovation in Technology. ASIANCON, pp. 1–4. http://dx.doi.org/10.1109/ASIANCON51346.2021.9544804.

Rautray, R., Swain, L., Dash, R., Dash, R., 2018. A brief review on text summarization methods. Int. J. Eng. Technol. 7 (4.5), 728. http://dx.doi.org/10.14419/ijet.v7i4.5.25070.

Reddy, K.M., Guha, R., 2023. Automatic text summarization for conversational chatbot. In: 2023 IEEE 8th International Conference for Convergence in Technology. I2CT, pp. 1–7. http://dx.doi.org/10.1109/I2CT57861.2023.10126161.

Rudinac, S., Chua, T.-S., Diaz-Ferreyra, N., Friedland, G., Gornostaja, T., Huet, B., Kaptein, R., Lindén, K., Moens, M.-F., Peltonen, J., Redi, M., Schedl, M., Shamma, D.A., Smeaton, A., Xie, L., 2018. Rethinking Summarization and Storytelling for Modern Social Multimedia. pp. 632–644. http://dx.doi.org/10.1007/978-3-319-73603-7_51.

Sadiq, A., Hussain Ali, Y., Fadhil, M., 2013. Text summarization for social network conversation. Inform. Sci. 1, 3–18. http://dx.doi.org/10.1109/ACSAT.2013.11.

Saggion, H., Lapalme, G., 2000. Concept identification and presentation in the context of technical text summarization. pp. 1–10. http://dx.doi.org/10.3115/1567564.1567565.

Saggion, H., Poibeau, T., 2013. Automatic Text Summarization: Past, Present and Future. pp. 3–21. http://dx.doi.org/10.1007/978-3-642-28569-1_1.

Sakhare, D.Y., Kumar, R., 2014. Syntactic and sentence feature based hybrid approach for text summarization. Int. J. Inf. Technol. Comput. Sci. 6 (3), 38–46. http://dx.doi.org/10.5815/ijitcs.2014.03.05.

Salman, Z.A.-W., 2023. Text summarizing and clustering using data mining technique. Al-Mustansiriyah J. Sci. 34 (1), 58–64. http://dx.doi.org/10.23851/mjs.v34i1.1195.

Sharifi, B., Hutton, M.A., Kalita, J., 2010. Summarizing microblogs automatically. In: NAACL HLT 2010 - Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Proceedings of the Main Conference, June. pp. 685–688.

Sharma, A.D., Deep, S., 2014. Too Long-Didn't Read: A Practical Web Based Approach towards Text Summarization. pp. 198–208. http://dx.doi.org/10.1007/978-3-319-04126-1_17.

Sharma, G., Sharma, D., 2022. Improving extractive text summarization performance using enhanced feature based RBM method. Revue d'Intell. Artif. 36 (5), 777–784. http://dx.doi.org/10.18280/ria.360516.

Sharma, G., Sharma, D., 2023. Automatic text summarization methods: A comprehensive review. SN Comput. Sci. 4 (1), http://dx.doi.org/10.1007/s42979-022-01446-w.

Sheik, R., Nirmala, S.J., 2021. Deep learning techniques for legal text summarization. In: 2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering. UPCON 2021, pp. 1–5. http://dx.doi.org/10.1109/UPCON52273.2021.9667640.

Shinde, M., Mhatre, D., Marwal, G., 2021. Techniques and research in text summarization - A survey. In: 2021 International Conference on Advance Computing and Innovative Technologies in Engineering. ICACITE, pp. 260–263. http://dx.doi.org/10.1109/ICACITE51222.2021.9404670.

Shiva Prakash, B., Sanjeev, K.V., Prakash, R., Chandrasekaran, K., 2019. A Survey on Recurrent Neural Network Architectures for Sequential Learning. pp. 57–66. http://dx.doi.org/10.1007/978-981-13-1595-4_5.

Siddiqui, H., Siddiqui, S., Rawat, M., Maan, A., Dhiman, S., Asad, M., 2021. Text summarization using extractive techniques. In: 2021 3rd International Conference on Advances in Computing, Communication Control and Networking. ICAC3N, pp. 28–31. http://dx.doi.org/10.1109/ICAC3N53548.2021.9725501.

Singh, S., Deepak, G., 2021. Towards a Knowledge Centric Semantic Approach for Text Summarization. pp. 1–9. http://dx.doi.org/10.1007/978-981-16-4486-3_1.

Singh, J., Singh, G., Singh, R., Singh, P., 2018. Evaluation and Analysis of Word Embedding Vectors of English Text Using Deep Learning Technique. pp. 709–722. http://dx.doi.org/10.1007/978-981-10-8657-1_55.

Sinha, A., Yadav, A., Gahlot, A., 2018. Extractive text summarization using neural networks. arXiv, abs/1802.1. http://arxiv.org/abs/1802.10137.

Sornil, O., Gree-ut, K., 2006. An automatic text summarization approach using content-based and graph-based characteristics. In: 2006 IEEE Conference on Cybernetics and Intelligent Systems. pp. 1–6. http://dx.doi.org/10.1109/ICCIS.2006.252361.

Sri, S.H.B., Dutta, S.R., 2021. A survey on automatic text summarization techniques. J. Phys. Conf. Ser. 2040 (1), 1742–6596. http://dx.doi.org/10.1088/1742-6596/2040/1/012044.

Sung, Y.T., Chang, K.E., Liu, T.C., 2016. The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. Comput. Educ. 94, 252–275. http://dx.doi.org/10.1016/j.compedu.2015.11.008.

Taeho Jo, D., 2019. Validation of graph based K nearest neighbor for summarizing news articles. In: 2019 International Conference on Green and Human Information Technology. ICGHIT, pp. 66–69. http://dx.doi.org/10.1109/ICGHIT.2019.00022.

Tambe, R., Thaokar, D., Pachghare, E., Sawane, P., Mehendole, P., Kakde, P., 2023. Abstractive text summarization using deep learning. Int. J. Res. Appl. Sci. Eng. Technol. 11 (3), 68–72. http://dx.doi.org/10.22214/ijraset.2023.49329.

Taylor, G., Barabin, K., Sayre, K., 2015. An application of reinforcement learning to supervised autonomy. https://api.semanticscholar.org/CorpusID:1203120.

Teufel, S., 2018. Deeper summarisation: The second time around: An overview and some practical suggestions. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): 9624 LNCS, pp. 581–598. http://dx.doi.org/10.1007/978-3-319-75487-1_44.

Thakkar, K.S., Dharaskar, R.V., Chandak, M.B., 2010. Graph-based algorithms for text summarization. In: 2010 3rd International Conference on Emerging Trends in Engineering and Technology. pp. 516–519. http://dx.doi.org/10.1109/ICETET.2010.104.

Thirumoorthy, K., Britto, J.J.J., 2023. A hybrid approach for text summarization using social mimic optimization algorithm. Iranian J. Sci. Technol., Trans. Electr. Eng. 47 (2), 677–693. http://dx.doi.org/10.1007/s40998-022-00572-8.

Thomas, J., Sreeraj, A., Sreeraj, A., Varghese, M.M., Kuriakose, T., 2022. Automatic Text Summarization Using Deep Learning and Reinforcement Learning. pp. 769–778. http://dx.doi.org/10.1007/978-981-16-5157-1_60.

Thushara, M.G., Mownika, T., Mangamuru, R., 2019. A comparative study on different keyword extraction algorithms. In: 2019 3rd International Conference on Computing Methodologies and Communication. ICCMC, pp. 969–973. http://dx.doi.org/10.1109/ICCMC.2019.8819630.

Tsuchiya, S., Yoshimura, E., Watabe, H., 2009. An information arrangement technique for a text classification and summarization based on a summarization frame. In: 2009 International Conference on Natural Language Processing and Knowledge Engineering. pp. 1–5. http://dx.doi.org/10.1109/NLPKE.2009.5313816.

Tymoshenko, K., Moschitti, A., 2019. Shallow and deep syntactic/semantic structures for passage reranking in question-answering systems. ACM Trans. Inf. Syst. 37 (1), 1–38. http://dx.doi.org/10.1145/3233772.

Umadevi, K., Chopra, R., Singh, N., Aruru, L., Kannan, R.J., 2018. Text summarization of Spanish documents. In: 2018 International Conference on Advances in Computing, Communications and Informatics. ICACCI, pp. 1793–1797. http://dx.doi.org/10.1109/ICACCI.2018.8554839.

Upton, H., Guillot, R., 2023. The Metrics Maze : Navigating the Landscape of Evaluation Techniques for MT Systems.

Vaissnave, V., Deepalakshmi, P., 2023. Modeling of automated glowworm swarm optimization based deep learning model for legal text summarization. Multimedia Tools Appl. 82 (11), 17175–17194. http://dx.doi.org/10.1007/s11042-022-14171-6.

Vale, R., Lins, R.D., Ferreira, R., 2020. An assessment of sentence simplification methods in extractive text summarization. In: Proceedings of the ACM Symposium on Document Engineering. DocEng 2020, pp. 1–9. http://dx.doi.org/10.1145/3395027.3419588.

Van Lierde, H., Chow, T.W.S., 2019. Learning with fuzzy hypergraphs: A topical approach to query-oriented text summarization. Inform. Sci. 496, 212–224. http://dx.doi.org/10.1016/j.ins.2019.05.020.

Vassiliou, G., Troullinou, G., Papadakis, N., Stefanidis, K., Pitoura, E., Kondylakis, H., 2021. Coverage-Based Summaries for RDF KBs. pp. 98–102. http://dx.doi.org/10.1007/978-3-030-80418-3_18.

Verma, P., Tyagi, D.P., 2020. Supervised Machine Learning: Predispositions, Practices and Perspectives. https://api.semanticscholar.org/CorpusID:233228017.

Verma, P., Verma, A., Pal, S., 2022. A fusion of variants of sentence scoring methods and collaborative word rankings for document summarization. Expert Syst. 39 (6), http://dx.doi.org/10.1111/exsy.12960.

Verma, V.K., Yadav, A., Jain, T., 2019. Key Feature Extraction and Machine Learning-Based Automatic Text Summarization. pp. 871–877. http://dx.doi.org/10.1007/978-981-13-1501-5_76.

Vinzelberg, O., Jenkins, M.D., Morison, G., McMinn, D., Tieges, Z., 2023. Lay text summarisation using natural language processing: A narrative literature review. J. Jpn. Soc. Clin. Cytol. 43 (1), 202, http://arxiv.org/abs/2303.14222.

Vishwakarma, S.K., Sharma, A.K., Verma, S.S., Utmal, M., 2021. Text classification using FP-growth association rule and updating the term weight. In: Advances in Science, Technology and Innovation. pp. 401–405. http://dx.doi.org/10.1007/978-3-030-66218-9_47.

Wan, S., Dale, R., 2001. Merging Sentences using Shallow Semantic Analysis: A first Experiment. https://api.semanticscholar.org/CorpusID:3154263.

Wang, Xuping, Chen, Xiaohu, Yao, Chunjiang, Wu, Wenfeng, Ding, Yao, 2013. The application of automatic summarization technology in document management. In: 2013 IEEE 4th International Conference on Software Engineering and Service Science. pp. 919–921. http://dx.doi.org/10.1109/ICSESS.2013.6615455.

Wang, Z., Jiang, Z., Zhang, X., Soon, J., Zhang, J., Xiaoyao, W., Du, H., 2023. Beyond pure text: Summarizing financial reports based on both textual and tabular data. In: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. pp. 5233–5241. http://dx.doi.org/10.24963/ijcai.2023/581.

Wang, B., Kuo, C.-C.J., 2020. SBERT-WK: A sentence embedding method by dissecting BERT-based word models. IEEE/ACM Trans. Audio, Speech, Lang. Process. 28, 2146–2157. http://dx.doi.org/10.1109/TASLP.2020.3008390.

Wu, F., Zheng, T., Yao, L., Feng, H., 2021. A new unsupervised Algorithm for extracting relationship words between two entities. In: 2021 3rd International Conference on Advances in Computer Technology, Information Science and Communication. CTISC, pp. 161–165. http://dx.doi.org/10.1109/CTISC52352.2021.00037.

Xu, Y., Ma, Y., 2023. Evolutionary neural architecture search combining multi-branch ConvNet and improved transformer. Sci. Rep. 13 (1), 15791. http://dx.doi.org/10.1038/s41598-023-42931-3.

Yadav, D., Katna, R., Yadav, A.K., Morato, J., 2022. Feature based automatic text summarization methods: A comprehensive state-of-the-art survey. IEEE Access 10, 133981–134003. http://dx.doi.org/10.1109/ACCESS.2022.3231016.

Yadav, A.K., Maurya, A.K., Ranvijay, Yadav, R.S., 2021. Extractive text summarization using recent approaches: A survey. Ingénierie Des Syst. d Inf. 26, 109–121, https://api.semanticscholar.org/CorpusID:233187276.

Yang, X., Li, Y., Zhang, X., Chen, H., Cheng, W., 2023. Exploring the Limits of ChatGPT for Query or Aspect-based Text Summarization. http://arxiv.org/abs/2302.08081.

Yang, P., Li, W., Zhao, G., 2019. Language model-driven topic clustering and summarization for news articles. IEEE Access 7, 185506–185519. http://dx.doi.org/10.1109/ACCESS.2019.2960538.

Yao, C., Shen, J., Chen, G., 2015. Automatic document summarization via deep neural networks. In: 2015 8th International Symposium on Computational Intelligence and Design. ISCID, pp. 291–296. http://dx.doi.org/10.1109/ISCID.2015.83.

Yong-Kwang Kim, Y.-M.C., 2006. An experimental study on automatic summarization of multiple news articles. J. Korean Soc. Inf. Manag. 23 (1), 83–98. http://dx.doi.org/10.3743/kosim.2006.23.1.083.

Yuan, W., Neubig, G., Liu, P., 2021. BARTSCORE: Evaluating generated text as text generation. In: Advances in Neural Information Processing Systems, vol. 33, pp. 27263–27277, (NeurIPS).

Zala, A., Cho, J., Kottur, S., Chen, X., Oguz, B., Mehdad, Y., Bansal, M., 2023. Hierarchical video-moment retrieval and step-captioning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 23056–23065.

Zhang, Q., Cai, Y., Xu, X., 2013. Orthogonal discriminant improved local tangent space alignment based feature fusion for face recognition. J. Shanghai Jiaotong Univ. (Science) 18 (4), 425–433. http://dx.doi.org/10.1007/s12204-013-1417-y.

Zhang, T., Kishore, V., Wu, F., Weinberger, K.Q., Artzi, Y., 2020. Bertscore: Evaluating text generation with bert. In: 8th International Conference on Learning Representations. ICLR 2020, pp. 1–43.

Zhang, C., Zhang, L., Wang, C.-J., Xie, J.-Y., 2014. Text summarization based on sentence selection with semantic representation. In: 2014 IEEE 26th International Conference on Tools with Artificial Intelligence. pp. 584–590. http://dx.doi.org/10.1109/ICTAI.2014.93.

Zhao, W., Peyrard, M., Liu, F., Gao, Y., Meyer, C.M., Eger, S., 2019. Moverscore: Text generation evaluating with contextualized embeddings and earth mover distance. In: EMNLP-IJCNLP 2019-2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference. pp. 563–578, https://www.scopus.com/inward/record.uri?eid=2-s2.0-85084322921&partnerID=40&md5=12d34775ec2e266c426d412be7a200dd.

Zhou, Z., Yu, H., Fan, G., Huang, Z., Yang, K., 2023. Towards retrieval-based neural code summarization: A meta-learning approach. IEEE Trans. Softw. Eng. 49 (4), 3008–3031. http://dx.doi.org/10.1109/TSE.2023.3238161.

Zhou, F., Zhang, F., Yang, B., 2010. Graph-based text representation model and its realization. In: Proceedings of the 6th International Conference on Natural Language Processing and Knowledge Engineering. NLPKE-2010, pp. 1–8. http://dx.doi.org/10.1109/NLPKE.2010.5587861.

Zhu, G., Tao, H., Wu, H., Chen, L., Liu, Y., Liu, Q., Chen, E., 2022. Text classification via learning semantic dependency and association. In: 2022 International Joint Conference on Neural Networks. IJCNN, pp. 1–8. http://dx.doi.org/10.1109/IJCNN55064.2022.9892656.