Exploring Current Methods and Trends in Text Summarization: A Systematic Mapping Study

Muhammad Faris Faisal Ahmad Raddi [a,*], Rohayanti Hassan [a], Noor Hidayah Zakaria [a], Mohd Zanes Sahid [b], Nurul Aswa Omar [b], Ardian Firosha [c]

[a] Faculty of Computing, Universiti Teknologi Malaysia, Jalan Iman, Skudai, Malaysia
[b] Faculty of Computer Sciences and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM), Parit Raja, Malaysia
[c] Department of Information Technology, Politeknik Negeri Padang, Padang, Indonesia
Corresponding author: *[*]muhammadfarisfaisal@graduate.utm.my*

Abstract— **This paper presents a systematic mapping study of the current methods and trends in text summarization, a challenging task in natural language processing that aims to condense information from one or multiple documents into a concise and coherent summary. The paper focuses on applying text summarization for the Malay language, which has received less attention than other languages. The paper employs a three-phased quality assessment procedure to filter and analyze 27 peer-reviewed publications from seven prominent digital libraries, covering 2016 to 2024. The paper addresses two research questions: (1) What is the extent of research on text summarization, especially for the Malay language and the education domain? and (2) What are the current methods and approaches employed in text summarization, with a focus on addressing specific problems and language contexts? The paper synthesizes and discusses the findings from the literature review and provides insights and recommendations for future research directions in text summarization. The paper contributes to advancing knowledge and understanding of the state-of-the-art techniques and challenges in text summarization, particularly for the Malay language.**

Keywords— **Text; summarization; method; approach; Malay; language.**

I. INTRODUCTION

In the contemporary digital landscape, individuals grapple with an overwhelming influx of information, predominantly accessible online, resulting in a cognitive state known as inundation [1], [2]. However, the protracted nature of textual content introduces a challenge wherein pivotal information risks eluding readers' recollection [3], [4]. Consequently, providing a concise summary emerges as a strategic approach to minimize cognitive effort for readers while extracting valuable insights from the text [5].

Automatic Text Summarization (ATS) constitutes the process of condensing information from one or multiple documents to render it suitable for a designated task or user [6], [7], all the while upholding the precision of the content [8]. As posited by Alias, Mohammad, Hoon, and Ping [9], the fundamental role of a summarizer system is to selectively extract pertinent contextual elements from a given source and distill them into a succinct and manageable presentation.

Summarization manifests in either abstract or extractive forms [7]. In extractive summarization, pivotal sentences are directly derived from the source text and meticulously rephrased in the summary through its selection system [2]. Conversely, an abstractive summary involves the generation of an abstract comprising rephrased sentences that encapsulate identical concepts present in the original text(s) [7].

A. Existing Reviews on Text Summarization

Recognized as one of the challenging tasks within the realm of Natural Language Processing (NLP) [4], the evolution of ATS has spanned several decades, with researchers initiating ATS endeavors as early as the 1950s [3]. The application of ATS extends across diverse knowledge-based domains, each aiming to achieve specific summarization objectives. These domains include health [8], [10], natural disaster [11], sports [8], legal [12], technology [8], education [13], and, notably, certain studies incorporate inputs from social media [14], [15], [16].

Furthermore, research on ATS has expanded to its application as a system catering to multiple languages [16] or specific linguistic contexts, spanning languages such as Malay [11], [17], English [2], [3], Turkish [18] and more. However, despite the achievements in text summarization, there is no assured guarantee that current methodologies can consistently generate the requisite summaries, particularly concerning grammatical accuracy and preservation of information context [16], [11], [19].

B. Scope of Study

To the best of our understanding, previous research has identified the utilization of machine learning (ML) methodologies to enhance the accuracy of text summarization algorithm outputs. However, a comprehensive review synthesizing and illuminating the current methods and trends in text summarization is lacking. This study employs a systematic mapping methodology to explore the existing trend of methodologies addressing challenges related to text summarization accuracy. The findings from this review can assist in the application of various text summarization algorithms for the following purposes:

Determining the extent of research on text summarization, including identifying trends and applications. Investigate the current methods and approaches used in text summarization, focusing on addressing specific problems and language contexts. Examine the role of domain-specific knowledge and data sources in the Malay Language Text Summarization methods. To advance our understanding of how contemporary trends in text summarization approaches are implemented and to further our research objectives, we propose the following research questions (RQs):

RQ1: What is the extent of research on Text Summarization?
a. How many of these studies specifically focus on Text Summarization?
b. Among the studies that specifically focus on Text Summarization, how has the trend evolved over the years?
c. Among the studies, how many are applied in the education domain?
d. Within the education domain, how many of them are specifically applied in the context of the Malay language?

RQ2: What are the current methods and approaches employed in Text Summarization?
a. Among these studies, which these methods and approaches address specific problems in Text Summarization?
b. Within the studies, what language contexts are frequently emphasized in Text Summarization?
c. How has the trend in Text Summarization, particularly in the Malay language, evolved over the years among the selected studies?

RQ3: What domain-specific knowledge is essential in the methods employed for Malay Language Text Summarization?
a. Among the studies, what sources of data or information are utilized in the context of Text Summarization?

b. Within the studies, which problems, methods, and approaches have been explored in Malay Language Text Summarization?

C. Contributions

The study makes the following contributions:
a. This systematic study aims to offer scholarly insights into the impact of the current methodology trend on text summarization.
b. The study conducted 37 peer-reviewed literature pieces on text summarization. It analyzes how the findings could enhance future research in this domain and discusses key considerations when selecting methodologies to enhance text summarization accuracy.

This study is structured into five sections. Section II furnishes a comprehensive description of the systematic methodology employed for collecting, categorizing, and filtering out irrelevant data. In Section III, the results are scrutinized and elucidated in connection with each research question formulated in the study. Section III unfolds the discoveries, succeeded by a discourse and recommendations. Finally, this study is brought to a close with IV.

II. MATERIALS AND METHOD

A. Planning

The literature search done in this study encompassed publications from literature indexed in seven prominent databases, as outlined in Table I. These digital libraries were selected based on their robust scientific foundation, recognition, and relevance to the study.

TABLE I
ONLINE REPOSITORIES OF LITERATURE

No	Digital Library	URL
1	Scopus	https://www.scopus.com/
2	ACM Digital Library	https://dl.acm.org/
3	IEEE Xplore	https://ieeexplore.ieee.org/
4	ProQuest	https://www.proquest.com/
5	ScienceDirect	https://www.sciencedirect.com/
6	SpringerLink	https://link.springer.com/
7	Web of Science	https://www.webofscience.com/

Incorporating inclusion and exclusion criteria is imperative as it mitigates bias, delineates the search scope, pinpoints pertinent publications, and excludes studies incongruent with the study's objectives. The manual screening of peer-reviewed articles against these criteria enables reviewers to assess whether the implemented or proposed text summarization algorithm strategies align with research standards. Table II delineates the inclusion and exclusion criteria employed in this research. Commencing from the specified starting year 2016 and extending to 2024 or 2024 for certain libraries ensures the inclusion of only the latest literature, minimizing bias, sharpening the search focus, and capturing relevant publications while excluding studies misaligned with the research objectives.

TABLE II
INCLUSION AND EXCLUSION CRITERIA

No	Inclusion Criteria
1	Published later than 2015
2	Relevant to text summarization
3	Peer-reviewed and published literature
4	Proposed or used methodology

No	Exclusion Criteria
1	Published prior 2016
2	Not relevant to text summarization
3	Non-peer-reviewed literature
4	Retracted and duplicated literature
5	Literature without full text or abstract
6	Non-English, Indonesian, or Malay literature

The search string was crafted using expressions tied to the objectives of the research questions (RQs), intending to pinpoint pertinent phrases or synonyms employed in articles about sentiment analysis of popular perceptions of text summarization algorithms. Iterative pilot searches were performed, refining the search keywords as needed. The search string incorporates terms associated with text summarization, the Malay language, and various approaches, as Table III outlines.

TABLE III
SEARCH STRING

No	Search
S1	("text summarization" OR "text summarisation") AND
S2	("Malay" OR "Bahasa Malaysia" OR "Bahasa Melayu") AND
S3	("method" OR "approach" OR "technique")

The literature search involves successive searches starting with the first search (S1) with applied filters, which are detailed for each database as in Table IV. These filters encompass year, subject area, document type, language, and criteria associated with inclusion and exclusion, serving as an initial measure during the literature retrieval process.

TABLE IV
FILTERS ON DATABASES

Database	Filter Details
Scopus	Year: 2016-2024 Subject Area: Computer Science, Engineering, Decision Sciences, Social Sciences, Multidisciplinary Document Type: Conference Paper, Article, Conference Review, Book Chapter, Review, Book Language: English, Malay
ACM	Year: 2016-2024 Document Type: Proceedings, Research Article Media Type: PDF
IEEE	Year: 2016-2023 Document Type: Conferences, Journals, Early Access Articles, Books
ProQuest	Year: 2016-2024 Source Type: Books, Conference Papers & Proceedings, Dissertations & Theses, Scholarly Journal Language: English, Malay
Science Direct	Year: 2016-2024 Article Type: Review Articles, Research Articles, Book Chapters Subject Areas: Computer Science, Engineering, Social Sciences, Decision Sciences Language: English
Springer Link	Year: 2016-2024 Content Type: Chapter, Conference Paper, Book, Article, Conference Proceedings Language: English Disciplines: Computer Science, Engineering
Web of Science	Year: 2016-2024 Document Types: Proceeding Paper, Article, Early Access, Review Article, Book Chapters Language: English, Malay

B. Execution

This section outlines the process of processing and filtering the publications retrieved from digital library searches. Initially, a pool of 13,588 records was amassed from the first search and filters, with SpringerLink having the highest number of publications at 3,808 records. The lowest number is from ACM Digital Library, with 621 records, as shown in Table V.

TABLE V
LITERATURE FILTRATION

Database	S1	S2	S3	Inclusion	Exclusion	KIV
Scopus	3106	19	15	11	4	0
ACM	621	0	0	0	0	0
IEEE	903	7	5	4	1	0
ProQuest	2664	53	53	7	46	0
Science Direct	881	18	18	2	16	0
Springer Link	3808	111	111	2	109	0
Web of Science	1665	10	8	1	7	0
Total	13588	217	210	27	183	0

The search is further refined by extending the search strings S2 and S3, respectively, reducing the total literature outputs to 217 and 210. From the 210 results, metadata, including title, author, abstract, and published date, is extracted as a list for further filtering.
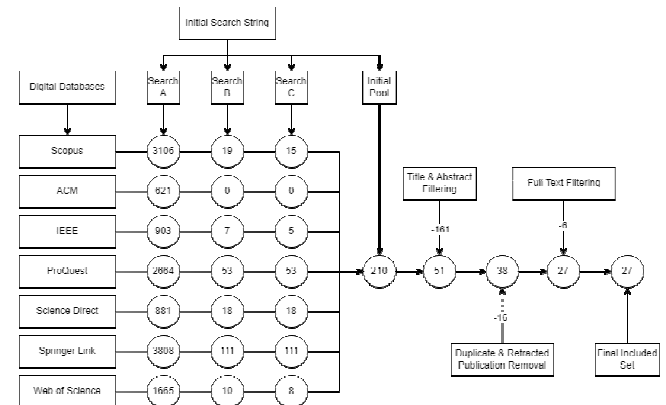

Fig. 1 Publication dataset filtering summary

A three-phased quality assessment procedure was enacted to exclude publications not meeting the inclusion criteria. The process involves three phases: title and abstract filtering, removal of duplicate and retracted publications, and full-text filtering. Table V elucidates the number of articles filtered at the end of step three.

In the first stage, the 210 publications underwent screening based on their title and abstract, removing 161 publications that did not utilize the discussion of text summarization and of the RQs. The titles, abstracts, and keywords were scrutinized to determine the articles' relevance for inclusion and exclusion based on the criteria listed in Table II. Articles were then cross-checked for redundancy, and publications were retracted in the second stage, removing 16 more articles. This results in the current exclusion publications 183 after full-text filtering, which is in stage 3. After applying the exclusion and inclusion criteria, 27 articles were retained for this study.

C. Synthesis

This section analyzes the information extracted to answer the offered research questions. We arranged the initial data collection based on the development of text summarization to answer RQ1. Metadata, including the paper's year of publication, is cataloged, offering an overview of the dataset. The dataset is further analyzed based on the approaches taken within the papers to address RQ2. This includes the dataset used, evaluation methods, and application language. Finally, the research's outcome was forwarded for peer review. Discussions are conducted within the reviewing process, and a final documented Excel spreadsheet is created manually.

TABLE VI
LIST OF PAPERS

No	Title	Year
1	A Malay Text Summarizer Using Pattern-Growth Method with Sentence Compression Rules	2016
2	An Extractive Malayalam Document Summarization Based on Graph Theoretic Approach	2016
3	A Malay Text Corpus Analysis for Sentence Compression Using Pattern-Growth Method	2016
4	Which Extractive Summarization Method for Malay Texts?	2017
5	Extract, Compress, and Summarize–An Experiment Using Malay News Article	2017
6	Rule-Based Approach on Extraction of Malay Compound Nouns in Standard Malay Document	2017
7	A Bootstrapped Approach to Multilingual Text Stream Parsing	2017
8	Understanding Human Sentence Compression Patterns for Malay Text Summarizer	2018
9	Exploring Graph Bushy Paths to Improve Statistical Multilingual Automatic Text Summarization	2018
10	MYTEXTSUM: A Malay Text Summarizer Model Using a Constrained Pattern-Growth Sentence Compression Technique	2018
11	Generation of News Headlines for the Malay Language Based on Term Features	2018
12	Determining Features of News Headlines in Malay News Document	2018
13	An Idea Based on Sequential Pattern Mining and Deep Learning for Text Summarization	2019
14	Summarizing Indonesian News Articles Using Graph Convolutional Network	2019
15	Classification of Short Possessive Clitic Pronoun Nya in Malay Text to Support Anaphor Candidate Determination	2020
16	Semantic Similarity and Text Summarization-Based Novelty Detection	2020
17	An Ontology-Based Information Extraction and Summarization of Multiple News Articles	2020
18	Bilingual Extractive Text Summarization Model Using Textual Pattern Constraints	2020
19	Technique On Malay Text Summarization: A Review	2020
20	Automatic Text Summarization for Malay News Documents Using Latent Dirichlet Allocation and Sentence Selection Algorithm	2021
21	Leading Sentence News TextRank	2021
22	A Syntactic-Based Sentence Validation Technique for Malay Text Summarizer	2021
23	Automatic Text Summarization Using Genetic Algorithm and Repetitive Patterns	2021
24	WikiDes: A Wikipedia-Based Dataset for Generating Short Descriptions from Paragraphs,	2022
25	Personalized News Recommendation Algorithm for Event Network	2022
26	A Visualized Hybrid Keyword-Cluster Approach for Extractive Text Summarizer Tool for STEM Education in Malaysia	2023
27	A Survey on Semantic Processing Techniques	2024

III. RESULTS AND DISCUSSION

Based on the method described in Section II, the findings for the applied RQs are presented as follows.

A. Research on Text Summarization

The trend of text summarization studies has increased steadily between 2016 and onwards in publications discussing the issue. Throughout the years, various languages, methodologies, results, and improvements have been studied and implemented to solve and innovate solutions regarding text summarization in different domains. In this mapping study, further filtration was done to analyze the specific topic of the research on text summarization methodology in the Malay language. As discussed in the previous section, the filtration results show that only a few studies are examining the area, with a total of less than 28 papers explicitly touching on the matter.
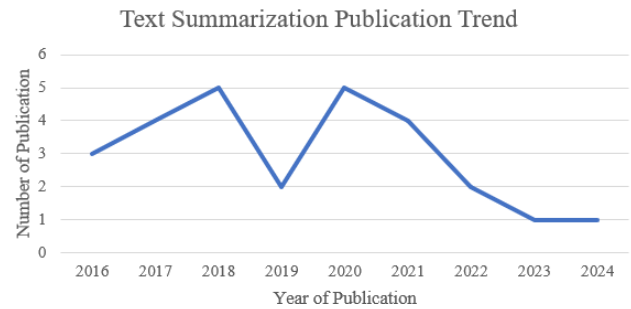

Fig. 2 Publication trend

The education domain as the focus of text summary study has not yet been a priority, as of this study, with only one paper discussing the issue, with the rest focusing more on summarizing articles and documents relating to news as its database and scope of study. As observed in [13], the study discusses the present extractive text summarization tool for the use of STEM education in Malaysia, proposes a new hybrid method for scoring the keywords, and displays summaries with highlighted keywords and word clouds. On the other hand, although research on Malay language text summarization has been conducted in large numbers in recent years, few researchers have studied the topic.
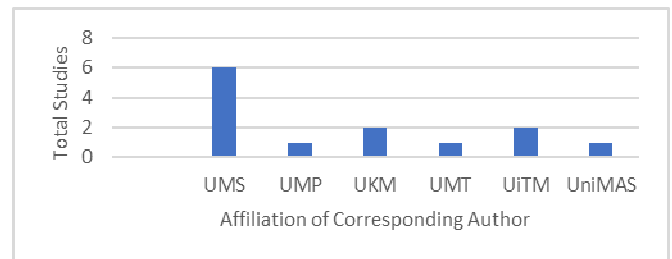

Fig. 3 Frequency of studies by universities

TABLE VII
LIST OF PAPERS BY UNIVERSITY

Affiliation of Corresponding Author	References
Universiti Malaysia Sabah	[1], [11], [20], [21], [5], [9]
Universiti Malaysia Pahang	[22]
Universiti Kebangsaan Malaysia	[17]
Universiti Teknologi Mara	[23], [8]
Universiti Malaysia Sarawak	[24]

B. Methods and Approaches

Various methods have been proposed and established within the studies analyzed within this mapping study. These include supervised, unsupervised, knowledge-based, and hybrid methodologies to solve the text summarization problem. In addition to the approaches used to develop text summarization with optimum accuracy for the Malay language, methods used from other languages, including Malayalam, English, and Indonesian, are included in this study's scope.

TABLE VIII
METHODS USED IN PAPERS

Method Used	Reference
Frequency Weightage	[1], [11], [17], [20], [21], [5], [9], [8], [25]
Genetic Algorithm	[2]
Graph-Based	[23], [26], [27]
Rule-Based	[22], [28], [29]
Ontology-Based	[30], [31]
Sentence Weightage	[17], [32]
Sentence Ranking	[33]
Hybrid	[13], [34]
No Proposed Method	[24], [34]

The approaches acquired from the results show that the advancement of Malay language summarization is in progress. By solving issues accordingly, the aim is to achieve optimum accuracy over time.

C. Discussion

The results of this study have shown that an abundant amount of research compromises the topic of text summarization. However, in the context of text summarization of the Malay language, there are many studies done to solve the specific development problem. The topic has been researched by solving issues accordingly towards achieving optimum accuracy in applying the text summarization using Malay language context, as can be seen by the continuous research done by Alias et. al [11], [13], [20], [21], [5], [25].

On the other hand, the methodologies used in this mapping study vary, with the majority using supervised, knowledge-based, and hybrid methods. Most authors have selected the approach of using annotated information and developing corpora to be proposed within their studies. Expanding research and development on text summarization, especially in the education domain, should be continued to ensure the advancement of this natural language processing application in enhancing knowledge. Besides its current focus for much research, which uses the materials from news articles as its database in its studies, the development of text summarization for the Malay language should be studied thoroughly. Its uniqueness compared to other languages requires a custom algorithm to achieve its optimum precision.

IV. CONCLUSION

This paper offers a systematic mapping study on the current methodologies and trends in text summarization, focusing on the Malay language and the education domain. The paper reviews 25 peer-reviewed publications from seven prominent digital databases. It analyzes them based on various criteria such as language-focused in-text summarization application, methodology, and domain-focused application to the system. The paper identifies the main challenges, applications, and gaps in text summarization and provides insights and recommendations for future research directions. The paper concludes that text summarization is a promising and evolving area of research with potential benefits.

REFERENCES

[1] S. Alias, S. K. Mohammad, G. K. Hoon, and T. T. Ping, "Extract, Compress and Summarize–An Experiment Using Malay News Article," Advanced Science Letters, vol. 23, no. 5, pp. 4336–4340, May 2017, doi: 10.1166/asl.2017.8317.

[2] E. Heidary et al., "Automatic Text Summarization Using Genetic Algorithm and Repetitive Patterns," Computers, Materials & Continua, vol. 67, no. 1, pp. 1085–1101, 2021, doi:10.32604/cmc.2021.013836.

[3] Z. Alami Merrouni, B. Frikh, and B. Ouhbi, "EXABSUM: a new text summarization approach for generating extractive and abstractive summaries," Journal of Big Data, vol. 10, no. 1, Oct. 2023, doi:10.1186/s40537-023-00836-y.

[4] N. Alami, M. Meknassi, N. En-nahnahi, Y. El Adlouni, and O. Ammor, "Unsupervised neural networks for automatic Arabic text summarization using document clustering and topic modeling," Expert Systems with Applications, vol. 172, p. 114652, Jun. 2021, doi: 10.1016/j.eswa.2021.114652.

[5] S. Alias, S. K. Mohammad, G. K. Hoon, and T. T. Ping, "A Malay text summarizer using pattern-growth method with sentence compression rules," 2016 Third International Conference on Information Retrieval and Knowledge Management (CAMP), pp. 7–12, Aug. 2016, doi: 10.1109/infrkm.2016.7806326.

[6] F. Carichon, F. Fettu, and G. Caporossi, "Unsupervised update summarization of news events," Pattern Recognition, vol. 144, p. 109839, Dec. 2023, doi: 10.1016/j.patcog.2023.109839.

[7] N. Ibrahim Altmami and M. El Bachir Menai, "Automatic summarization of scientific articles: A survey," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 4, pp. 1011–1028, Apr. 2022, doi: 10.1016/j.jksuci.2020.04.020.

[8] N. A. Rahman, S. N. A. Ramlam, N. A. Azhar, H. M. Hanum, N. I. Ramli, and N. Lateh, "Automatic Text Summarization for Malay News Documents Using Latent Dirichlet Allocation and Sentence Selection Algorithm," 2021 Fifth International Conference on Information Retrieval and Knowledge Management (CAMP), pp. 36–40, Jun. 2021, doi: 10.1109/camp51653.2021.9498029.

[9] S. Alias, S. K. Mohammad, G. Keng Hoon, and T. Tien Ping, "A Malay Text Corpus Analysis for Sentence Compression Using Pattern-Growth Method," Jurnal Teknologi, vol. 78, no. 8, Jul. 2016, doi: 10.11113/jt.v78.7413.

[10] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," IEEE Electron Device Letters, vol. 20, no. 11, pp. 569–571, Nov. 1999, doi:10.1109/55.798046.

[11] S. Alias, M. S. Sainin, and S. K. Mohammad, "A Syntactic-Based Sentence Validation Technique for Malay Text Summarizer," Journal of Information and Communication Technology, vol. 20, 2021, doi:10.32890/jict2021.20.3.3.

[12] Q. Wang et al., "Towards Legal Judgment Summarization: A Structure-Enhanced Approach," ECAI 2023, Sep. 2023, doi:10.3233/faia230553.

[13] S. Alias, M. Majalin, and N. Hayatin, "A Visualized Hybrid Keyword-Cluster Approach for Extractive Text Summarizer Tool for STEM Education in Malaysia," 2023 IEEE 8th International Conference On Software Engineering and Computer Systems (ICSECS), pp. 139–144, Aug. 2023, doi:10.1109/icsecs58457.2023.10256370.

[14] J. Fan, X. Tian, C. Lv, S. Zhang, Y. Wang, and J. Zhang, "Extractive Social Media Text Summarization Based on Mfmmr-Bertsum," 2023, doi: 10.2139/ssrn.4524718.

[15] F. A. Ghanem, M. C. Padma, and R. Alkhatib, "Automatic Short Text Summarization Techniques in Social Media Platforms," Future Internet, vol. 15, no. 9, p. 311, Sep. 2023, doi: 10.3390/fi15090311.

[16] N. Londhe, "A Bootstrapped Approach to Multilingual Text Stream Parsing.," *ProQuest LLC*, 2017.

[17] S. A. Mohd Noah, N. Mohamad Ali, and M. S. Hasan, "Penjanaan Ringkasan Isi Utama Berita Bahasa Melayu berdasarkan Ciri Kata (Generation of News Headline for Malay Language based on Term Features)," GEMA Online® Journal of Language Studies, vol. 18, no. 4, pp. 42–60, Nov. 2018, doi: 10.17576/gema-2018-1804-04.

[18] B. Baykara and T. Güngör, "Turkish abstractive text summarization using pretrained sequence-to-sequence models," Natural Language Engineering, vol. 29, no. 5, pp. 1275–1304, May 2022, doi:10.1017/s1351324922000195.

[19] S. Hou and R. Lu, "Knowledge-guided unsupervised rhetorical parsing for text summarization," Information Systems, vol. 94, p. 101615, Dec. 2020, doi: 10.1016/j.is.2020.101615.

[20] S. Alias, S. K. Mohammad, G. K. Hoon, and M. S. Sainin, "Understanding Human Sentence Compression Pattern for Malay Text Summarizer," 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP), pp. 1–6, Mar. 2018, doi: 10.1109/infrkm.2018.8464788.

[21] S. Alias, S. K. Mohammad, K. H. Gan, and T. T. Ping, "MYTextSum: A Malay Text Summarizer Model Using a Constrained Pattern-Growth Sentence Compression Technique," Computational Science and Technology, pp. 141–150, 2018, doi:10.1007/978-981-10-8276-4_14.

[22] N. H. Mohd Noor, S. A. Mohd Noah, and M. J. Ab Aziz, "Classification of Short Possessive Clitic Pronoun Nya in Malay Text to Support Anaphor Candidate Determination," Journal of Information and Communication Technology, vol. 19, 2020, doi:10.32890/jict2020.19.4.3.

[23] A. Aries, D. E. Zegour, and W. K. Hidouci, "Exploring Graph Bushy Paths to Improve Statistical Multilingual Automatic Text Summarization," Computational Intelligence and Its Applications, pp. 78–89, 2018, doi: 10.1007/978-3-319-89743-1_8.

[24] B. Ranaivo-Malançon and H. Iboi, "Which Extractive Summarization Method for Malay Texts?," in *Proceedings of the 6th International Conference on Computing and Informatics*, pp. 577–583, 2017.

[25] S. Alias, M. S. Sainin, and S. K. Mohammad, "Model Peringkasan Teks Ekstraktif Dwibahasa menggunakan Fitur Kekangan Corak Tekstual (Bilingual Extractive Text Summarization Model using Textual Pattern Constraints)," GEMA Online® Journal of Language Studies, vol. 20, no. 3, pp. 70–95, Aug. 2020, doi: 10.17576/gema-2020-2003-05.

[26] E. B. Ajmal and R. P. Haroon, "An Extractive Malayalam Document Summarization Based on Graph Theoretic Approach," 2015 Fifth International Conference on e-Learning (econf), pp. 237–240, Oct. 2015, doi: 10.1109/econf.2015.41.

[27] G. Garmastewira and M. L. Khodra, "Summarizing Indonesian News Articles Using Graph Convolutional Network," Journal of Information and Communication Technology, vol. 18, 2019, doi:10.32890/jict2019.18.3.4675.

[28] Z. A. Bakar, N. K. Ismail, and M. I. M. Rawi, "Rule-based Approach on Extraction of Malay Compound Nouns in Standard Malay Document," IOP Conference Series: Materials Science and Engineering, vol. 226, p. 012106, Aug. 2017, doi: 10.1088/1757-899x/226/1/012106.

[29] D. S. Maylawati, Y. J. Kumar, F. B. Kasmin, and M. A. Ramdhani, "An idea based on sequential pattern mining and deep learning for text summarization," Journal of Physics: Conference Series, vol. 1402, no. 7, p. 077013, Dec. 2019, doi: 10.1088/1742-6596/1402/7/077013.

[30] S. Kumar and K. K. Bhatia, "Semantic similarity and text summarization based novelty detection," SN Applied Sciences, vol. 2, no. 3, Feb. 2020, doi: 10.1007/s42452-020-2082-z.

[31] S. Venkatachalam, L. P. Subbiah, R. Rajendiran, and N. Venkatachalam, "An ontology-based information extraction and summarization of multiple news articles," International Journal of Information Technology, vol. 12, no. 2, pp. 547–557, Oct. 2019, doi:10.1007/s41870-019-00367-x.

[32] Y. Han, "Personalized News Recommendation Algorithm for Event Network," Mathematical Problems in Engineering, vol. 2022, pp. 1–10, Aug. 2022, doi: 10.1155/2022/7813457.

[33] H. T. Ta et al., "WikiDes: A Wikipedia-based dataset for generating short descriptions from paragraphs," Information Fusion, vol. 90, pp. 265–282, Feb. 2023, doi: 10.1016/j.inffus.2022.09.022.

[34] R. Mao et al., "A survey on semantic processing techniques," Information Fusion, vol. 101, p. 101988, Jan. 2024, doi:10.1016/j.inffus.2023.101988.