







### 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.

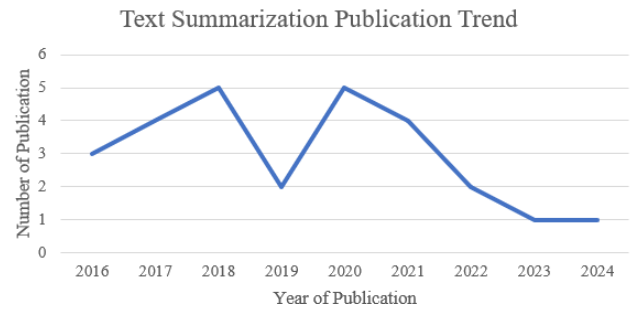


Fig. 2 Publication trend

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.

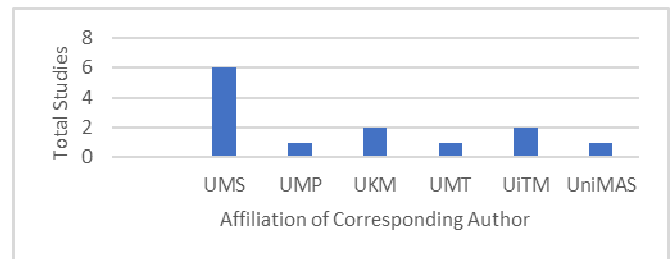


Fig. 3 Frequency of studies by universities

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.

## ACKNOWLEDGMENT

This work was supported by the Ministry of Higher Education, Malaysia (MOHE) under the Fundamental Research Grant Scheme (FRGS), FRGS/1/2023/ICT02/UTM/02/8.

## 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. Ramli, 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/infikm.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.