Indonesian Journal of Electrical Engineering and Computer Science
Vol. 40, No. 2, November 2025, pp. 907~915
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v40.i2.pp907-915

Query keyword extraction in discriminative marginalized probabilistic neural method for multi-document summarization

Bambang Subeno[1,2], Indra Budi[1], Evi Yulianti[1]
[1]Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia
[2]School of Computing, Telkom University, Bandung, Indonesia

Article Info

Article history:

Received Nov 20, 2024
Revised Aug 11, 2025
Accepted Oct 14, 2025

Keywords:

Background sentence query
DAMEN
Keyword extraction
Keyword query
Multi-document summarization

ABSTRACT

The large number of textual documents in the medical field makes it very difficult for readers to obtain comprehensive information. Users usually use a query approach to get the desired information. Using the correct query will produce relevant information. In the existing discriminative marginalized probabilistic neural method, referred to as DAMEN, used for multi-document summarization, a background sentence query is used to retrieve the top-K relevant documents and then generate a summary of these documents. However, the background sentence query used to retrieve the top-K documents did not provide accurate summary results. The author improved the DAMEN model by adding a keyword extraction process to the query background sentence. We call this model Q-DAMEN. Our model shows significant improvement over the original DAMEN method, with the best results achieved by the variation of using a keyword query entered into the discriminator component and a background sentence query entered into the generator component. The multipartieRank keyword extraction method shows the best results with a Rouge-1 value of 29.12, Rouge-2 of 0.79, and Rouge-L of 15.53. The results demonstrate that the more accurate the keywords extracted from the sentence background query, the more accurate the multi-document summaries generated.

Corresponding Author:

Indra Budi
Faculty of Computer Science, Universitas Indonesia
Depok, 16424, Indonesia
Email: indra@cs.ud.ac.id

1. INTRODUCTION

With the development of the Internet, digital text documents are increasingly used and grow into very large data [1]. With the many textual documents spread over digital media, it is not easy to get specific information desired by users. Special techniques are needed to get information spread over different digital media, such as machine learning (ML) [1], [2]. This has triggered the desire of many researchers to develop an effective approach that can automatically summarize text and produce a summary that contains important sentences and includes all relevant important information from the original document [2], [3].

ML is a branch of artificial intelligence (AI) that develops dynamic algorithms capable of making decisions based on data. Many tasks are developed to extract information contained in textual documents, such as clustering, classification, and summarization [2]. Clustering is a task to group data according to similarity of features and characteristics of data [4]. Classification is a task of identifying and grouping data based on predetermined categories. There are five steps to perform the classification process, namely text preprocessing, text representation, feature selection, classifier training, and effect evaluation [5]. Summarization is a task in the field of natural language processing (NLP) that involves condensing texts to

summarize the majority of the information from the source document [6], [7]. Summarization is necessary to support data analysis and to explore large data sets [8]. The text document summarization that has been carried out consists of single document summarization [9]–[12]. and multi-document summarization [13]–[15] with an extractive [16], [17] and abstractive [18], [19] summarization approach.

Single document summarization is a summarization that uses a single document as an input source [17]. Multi-document summarization is a summarization that automatically produces a short summary of a large collection of text documents, making it easier for readers to understand the main point of information discussed in the entire content of the document and increasing the accessibility of content from various existing topics [20]. Several approaches are basically carried out to produce a better summary that reflects the content of the main document [19], one of which is the query-based summarization approach. The query-based summarization approach can often take the form of a query for words or phrases that refer to a particular entity [21]. Several query-based summarization models, namely the TASA model [22], question-driven summarization (QDS) [21], and discriminative marginalized probabilistic method (DAMEN) [23], have not shown significant results. In the TASA model, sentences are scored based on words that have high weights to be used as summary candidates; this model is still limited to short sentences and less relevant to the topic [22]. The QDS model is a summarization based on questions; the answers to questions are called summaries. However, this model only deals with questions in each sentence and ignores mutual information between sentences and clusters [21]. The DAMEN model represents a method for abstractive multi-document summarization that involves the combination of background sentences into cluster documents. In this DAMEN model, the input documents must be in the form of clusters, and the background sentences are determined manually [23]. Although the summarization results are better than existing abstraction summarization, this model is still far from human-generated summaries.

The results of sentence query-based summarization show that there are still some limitations. This conclusion is in line with the results of the research conducted by [24], [25], which states that the sentence query-based multi-document summarization approach mostly fails to produce a better summary compared to the keyword query-based approach. In addition, users tend to prefer searching with keywords rather than with long sentences [25]. Therefore, in this study, the DAMEN model was modified by adding a keyword extraction process from background sentences before using them to retrieve documents. Background sentences are still used in the summarization process, so that summary candidates come from the results of document retrieval and the background sentences themselves. The summarization results are then compared with the word query approach and the background sentence query.

In the existing DAMEN query, the background sentence is used in the discriminator component to retrieve the top-K most relevant documents, then in the generator component, these documents are processed together with the background sentences to produce the final summary according to their clusters. In our proposed model, the query used to retrieve documents in the discriminator component does not use sentences, but uses keywords extracted from background sentences, with the aim of determining the effect of query use on summarization results. Before running the experimental process with a lot of data, the author conducted a small experiment to understand whether the use of keyword query can positively affect the multi-document summarization results using DAMEN method. The data used is a cluster with a process according to the stages in Figure 1. The existing DAMEN uses background sentences in the discrimination component, and the generator produces a Rouge-1 value of 23.52, Rouge-2 is 8.54, and Rouge-L is 13.44. Adding the keyword extraction process to DAMEN, which is performed in the discriminator component, produces a Rouge-1 value of 30.08, Rouge-2 is 7.21, and Rouge-L is 15.92. An example of a simple experimental procedure is shown in Figure 2. Based on the results obtained from the small experiment in Figure 2, a good keyword query from background sentence query can lead to better summarization results that is caused by a better top-K documents retrieved. Therefore, in this study, we propose the use of keyword extraction on DAMEN to add the keyword extraction query process to the background sentence by using larger data, with the hope that the results obtained show consistency.

2. RESEARCH METHOD

This section discusses the implementation of keyword extraction queries that convert background sentences into keywords. The process stage starts by taking MS2 data; clusters that have background are processed for the next stage. After the data containing the background sentence is clear, the document is processed by the indexer, and the background sentence of each cluster is extracted from the keyword words. Before extracting the keyword words, a preprocessing process is carried out to remove unimportant words. After the keyword extraction, the words are used as a query to retrieve documents in the discriminator component. The top-K retrieval results are combined with the original sentence query used for the generation summarization process. See Figure 1 for more details.

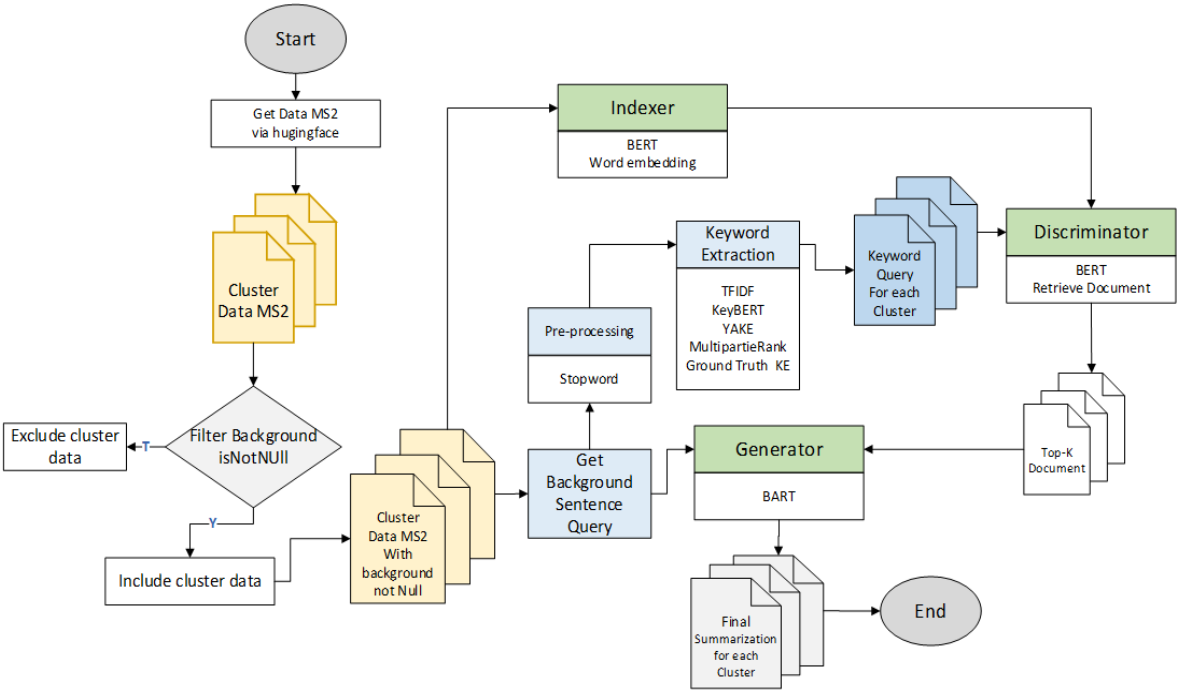

Figure 1. Flow of the keyword query extraction process for multi-document summarization

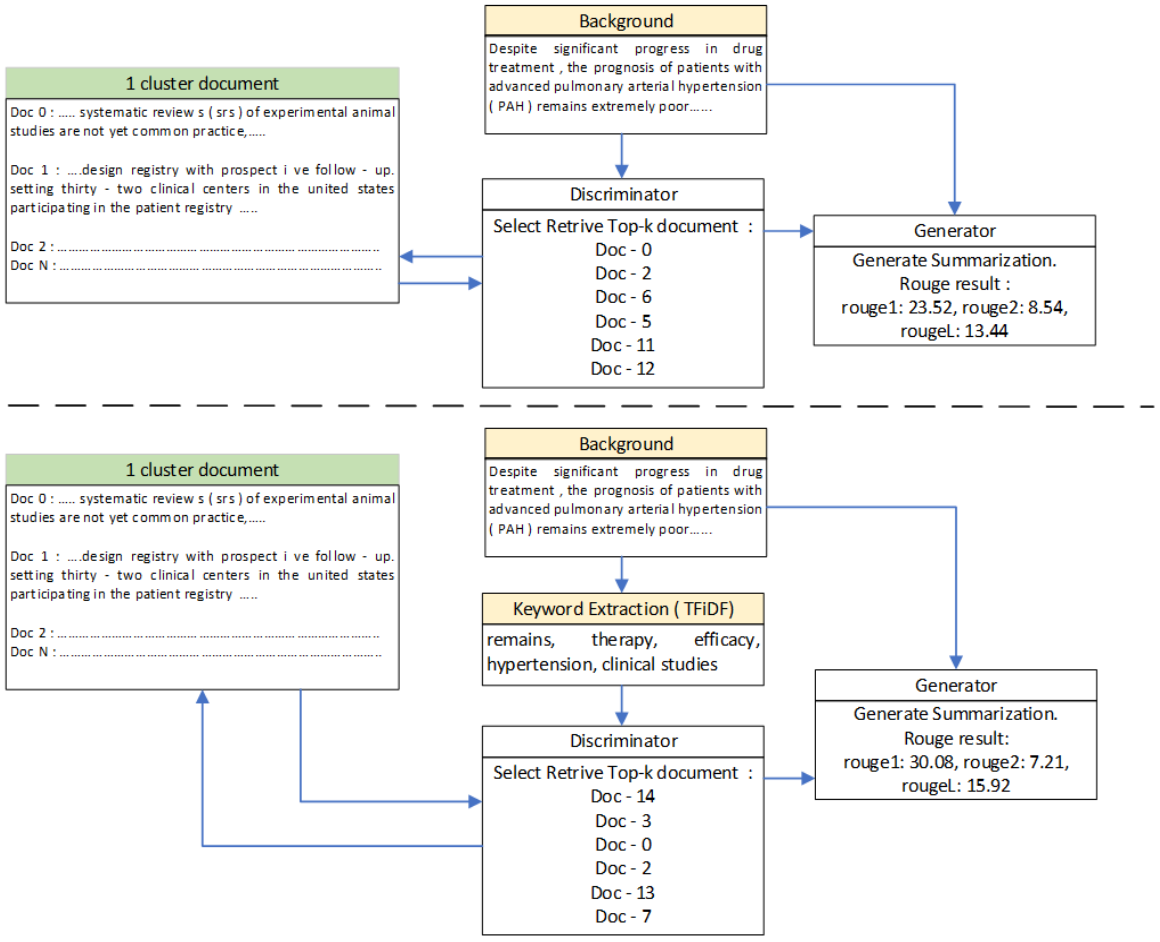

Figure 2. Illustration of a simple experimental process using background sentence and keyword extraction

2.1. Keyword extraction

Keyword extraction is a method that automatically extracts important words from a document containing five to ten words; these important words provide an instant summary of the document [26], [27]. Keywords are useful for readers to see the content of the document at a glance, for Internet users to find the most relevant part of a web page, or even for search engines to improve search results. The use of keyword extraction method can use a statistical approach such as term frequency–inverse document frequency (TFIDF), word frequency, Patricia tree [26], yet another keyword extractor (YAKE) [27], statistical graph-based such as MultipartieRank [28], or a ML approach such as keyword extraction using BERT (bidirectional encoder representations from transformers) (KeyBERT) [27]. In this study, keyword extraction was attempted in the experimental scenario using the TFIDF, KeyBERT, YAKE, and MultipartyRank methods. TFIDF is a classic keyword extraction method that is based on cultures based on word frequency [29]. KeyBERT is a simple keyword extraction method that uses the BERT model with a transformer library based on cosine similarity to find the most relevant words [30], [31]. YAKE is a method for extracting keywords based on word frequency, word position, word relatedness to context, and word difference in multilingual documents [26]. MultipartieRank is a method for extracting keyword phrases and topics in a multipartie graph structure based on the TaxtRank method [28].

2.2. DAMEN

DAMEN is a discriminative marginalized probabilistic neural method used to summarize the medical domain. The DAMEN model consists of three components, namely indexer, discriminator and generator [23]. At the indexer stage, each document in the cluster is indexed using the BERT model. At the discriminator stage, document retrieval is performed to distinguish important information in a cluster based on its background sentence query. Each document cluster is given a score of matching with the background using cosine similarity, and then the top K are selected to be used as candidates for the next process. The model used in this stage is BERT. At the generator stage, the summarization process is performed based on the background sentence query and the top K documents using the bidirectional and auto-regressive transformers (BART) model.

The model we developed consists of indexer, discriminator, and generator components based on the existing DAMEN model. However, there is a difference in the discriminator component; the background sentence is first processed by keyword extraction to obtain keywords before entering the discriminator component. The basic idea of adding keyword extraction is that the sentence query approach used in previous studies is less good than the word query used in extractive summaries [24], and based on the results of a small experiment according to Figure 2, which shows a correlation between the keyword query and the summarization results. The model we developed is called Q-DAMEN as shown in Figure 3.

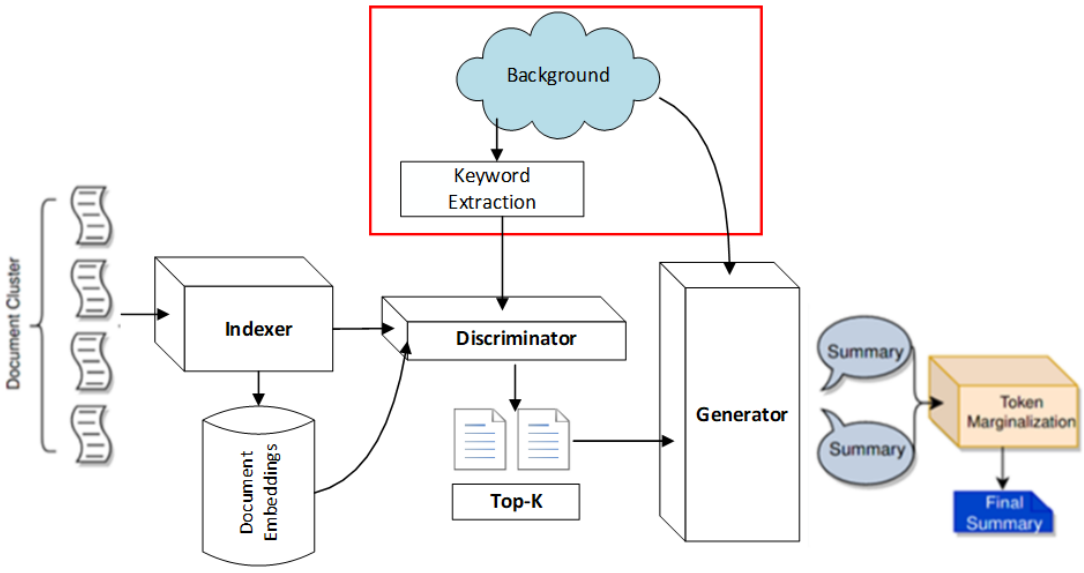

Figure 3. Query keyword extraction on DAMEN (Q-DAMEN)

3. RESULTS AND ANALYSIS

The third section consists of three subsections, namely dataset, experimental scenario, and experimental results and analysis. The dataset subsection discusses the dataset used and its characteristics. The experimental scenario subsection discusses several experimental scenarios conducted using 5 experimental scenarios. The experimental results and analysis subsection discusses each scenario result and provides its analysis.

3.1. Dataset

The dataset used in this study is the MS2 dataset, which is a collection of related data on medical data for multi-document summaries. The MS2 dataset is a collection of abstract documents that are already in the form of public clusters. Each cluster consists of a background, a collection of abstract documents, a target summary, and an example of data, as shown in Figure 4.

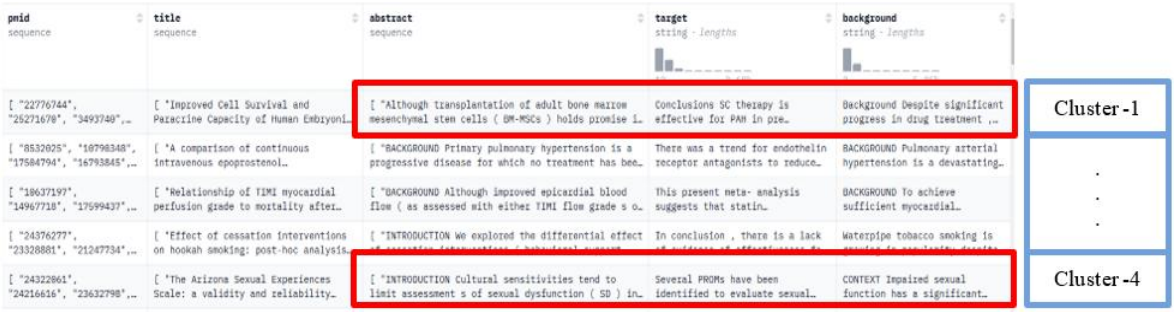

Figure 4. Example representation of the first 4 rows of the MS2 dataset

Background is a short sentence that is a query that describes a question or research topic according to its cluster. A collection of abstract documents is a collection of research abstracts grouped according to their background. The target summary is the result of the conclusion that is used as the ground truth of each cluster. The dataset used is shown in Table 1.

Table 1. MS2 data statistics used

Number of clusters	Average number of documents per cluster	Average number of words per cluster
1,000 cluster	22 documents	6,839 words

3.2. Experimental scenario

Based on the Q-DAMEN model in Figure 3 and the MS2 dataset, we conducted four experimental scenarios using the keyword extraction methods TFIDF, KeyBERT, YAKE, and MultipartyRank. Scenario 1 calculates the performance of keyword extraction. Scenario 2 calculates the use of the keyword query for all models with variations. (a) The background sentence query is processed by keyword extraction to obtain the keyword query, then enters the discriminator component, and the keyword query also enters the generator component. (b) The background sentence query is processed by keyword extraction to obtain the keyword query, then enters the discriminator component, and the background sentence query remains in the generator component. (c) The background sentence query enters the discriminator component, and the keyword query enters the generator component. Scenario 3 calculates the use of the keyword query generated by human annotators with three variations as in Scenario 2. Scenario 4 computes the T-test to determine the consistency of keyword extraction usage against summarization results. Scenario 5 calculates the correlation between keyword extraction performance and summarization results to determine the correlation between the performance of the keywords used and the resulting summarization.

3.3. Experiment results and analysis

The results of the 1st experimental scenario can be seen in Table 2. The results of the 2nd experimental scenario can be seen in Table 3. The results of the 3rd experimental scenario can be seen in the last two rows of the table. The results of the 4th experimental scenario can be seen in Table 4. The results of

the 5th experimental scenario can be seen in Table 5. Table 2 shows that the best keyword extraction performance uses the MultipartieRank model with a precision @5 value of 0.60.

Table 2. Keyword extraction performance results

Method	P @5	R @5	F1 @5
TFIDF	0.40	0.13	0.20
KeyBERT	0.26	0.08	0.13
YAKE	0.53	0.17	0.26
MultipartieRank	**0.60**	**0.19**	**0.30**

Based on the results in Table 3, of all the keyword extraction models used, the best variation is variation (b), where keyword query is inputted into the discriminator and background sentence query is inputted into the generator. Based on variation (b) in Table 3, the MultipartieRank keyword method shows the best results compared to other keyword extraction methods and baseline variations. This shows that using good keyword query words can result in good top-K document retrieval, which leads to better summarization results while still using the BART model generator. For variation (a), keyword query is inputted into both the discriminator and generator components, and for variation (c), background sentence query is inputted into the discriminator and keyword query is inputted into the generator, there is a tendency for less than good results. This shows that when the generator input in the form of keyword query words is combined directly with top-K documents, it produces less than good summarization. This is because the generator model used is the BART model, which is based on the generation of abstractive sentence context.

We also compared the results of manual keywords by comparing the results of variation (b) to find out whether good keyword query can also lead to good summarization results as in scenario 1. The results show that manual keywords provide better results than the baseline, with a difference of 7.71%. Compared to all keyword extraction models used, manual keywords still show the best performance. MultipartiteRank and YAKE are closest to the performance of manual keywords with a difference of 3.26% and 4.47%, respectively. This shows results that reinforce scenario 1, that the better the keyword query produced for document retrieval, the better the summarization results will be, even though the discriminator still uses BERT and the generator still uses BART.

In the 4th scenario, a Rouge T-test was performed based on the use of the keyword extraction method and the best variation of the baseline. This T-test was performed to determine the consistency of the best results with a statistical method approach. Table 4 shows that the keyword extraction method shows consistent results in all clusters with a T-test value below 0.05.

Table 3. The results of the evaluation of the use of keyword extraction on the summarization results

Summarization method	Keyword extraction method	Discriminator	Generator	Rouge-1	Rouge-2	Rouge-L
DAMEN	**Baseline*** (query sentence)	Sentence background	Sentence background	24.67	6.60	14.26
Q-DAMEN	TFIDF	Keyword background	Keyword background	18.90	2.57	13.35
		Keyword background	Sentence background	25.33	6.75	15.43
		Sentence background	Keyword background	14.941	1.06	10.62
Q-DAMEN	KeyBERT	Keyword background	Keyword background	18.03	2.24	12.10
		Keyword background	Sentence background	24.73	6.30	15.68
		Sentence background	Keyword background	18.34	2.11	12.46
Q-DAMEN	YAKE	Keyword background	Keyword background	6.25	0.0	6.25
		Keyword background	Sentence background	27.91	**11.90**	**16.53**
		Sentence background	Keyword background	17.20	4.39	10.75
Q-DAMEN	MultipartieRank	Keyword background	Keyword background	17.20	4.39	10.75
		Keyword background	Sentence background	**29.12**	7.92	15.53
		Sentence background	Keyword background	18.18	4.65	13.63

Table 4. Results of T-test performance keyword extraction

Method	T-test results
TFIDF	$1.22\ x\ 10^{-6}$
KeyBERT	$3.04\ x\ 10^{-2}$
YAKE	$5{,}54\ x\ 10^{-112}$
MultipartieRank	$\mathbf{2,80\ x\ 10^{-189}}$

In the 5th scenario, a correlation test was conducted between the performance keyword extraction results and the summary results variation (b) using Pearson's correlation coefficient. The purpose of this scenario is to ascertain whether there is a correlation between keyword extraction and summarization. Based on the results from Table 2, the most accurate keyword extraction method is the MultipartieRank method, and similar to the results shown in Table 5, the MultipartieRank method also shows the highest correlation with summarization, with a value of 0.2756. The least accurate keyword extraction method is the KeyBERT method, and similar to the results shown in Table 5, the KeyBERT method also shows the lowest correlation result with the summarization, with a value of 0.0812. This result reinforces that the better the keyword extraction, the better the summarization results.

Table 5. Results of pearson correlation performance keyword extraction with performance summarization

Method	Pearson correlation
TFIDF	0.1091
KeyBERT	0.0812
YAKE	0.2038
MultipartieRank	**0.2756**

4. CONCLUSION

The Q-DAMEN method which adds the process of extracting words from the background of sentences, shows better results than the baseline. The best results are shown in variation (b) with keyword query inputted into the discriminator component to retrieve top-K documents, and the generator still uses the background sentence query. The worst results are shown in variations (a), (c), which use the results of the keyword query inputted into the generator component. Among the keyword extraction methods used, the MultipartieRank method shows the best results with a Rouge-1 value of 29.12, Rouge-2 is 0.79, and Rouge-L is 15.53. The more accurate the keyword query, the more accurate the selection of top-K documents, which further results in the more accurate summarization results.

The use of keyword extraction in this study is an important factor in producing a better summary. In fact, traditional keyword extraction methods have been able to produce a summary that is superior to the baseline. Currently, many keyword extraction methods have been developed, ranging from traditional approaches to deep learning methods, which are able to overcome various limitations of traditional methods and are worth exploring in further research. Keyword extraction methods such as KeyGames and jointGL show the best performance for unsupervised data. Meanwhile, for supervised data, models such as SMART-KPE+Full and KIEMP provide the most optimal results. Therefore, the keyword extraction approach needs to be considered in future studies. In addition, considering the average document length of each cluster, the summary generation process needs to be tried with other models that can handle long documents, such as LED, Longformer, and BigBird-Pegasus. By making improvements adapted to variation (b), it is expected that the summarization results will be even better.

FUNDING INFORMATION

The authors state no funding is involved.

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY

Data availability does not apply to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: management, analysis and future prospects," *Journal of Big Data*, vol. 6, no. 1, Jun. 2019, doi: 10.1186/s40537-019-0217-0.

[2] M. Gambhir and V. Gupta, "Recent automatic text summarization techniques: a survey," *Artificial Intelligence Review*, vol. 47, no. 1, pp. 1–66, Mar. 2017, doi: 10.1007/s10462-016-9475-9.

[3] M. Allahyari *et al.*, "Text summarization techniques: a brief survey," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 10, 2017, doi: 10.14569/ijacsa.2017.081052.

[4] A. E. Ezugwu *et al.*, "A comprehensive survey of clustering algorithms: state-of-the-art machine learning applications, taxonomy, challenges, and future research prospects," *Engineering Applications of Artificial Intelligence*, vol. 110, p. 104743, Apr. 2022, doi: 10.1016/j.engappai.2022.104743.

[5] H. Zhu, X. Xia, J. Yao, H. Fan, Q. Wang, and Q. Gao, "Comparisons of different classification algorithms while using text mining to screen psychiatric inpatients with suicidal behaviors," *Journal of Psychiatric Research*, vol. 124, pp. 123–130, May 2020, doi: 10.1016/j.jpsychires.2020.02.019.
[6] A. Aries, D. E. Zegour, and K. W. Hidouci, "AllSummarizer system at MultiLing 2015: multilingual single and multi-document summarization," in *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 2015, pp. 237–244, doi: 10.18653/v1/W15-4634.
[7] N. Rane and S. Govilkar, "Recent trends in deep learning based abstractive text summarization," *International Journal of Recent Technology and Engineering*, vol. 8, no. 3, pp. 3108–3115, Sep. 2019, doi: 10.35940/ijrte.C4996.098319.
[8] M. Francia, M. Golfarelli, and S. Rizzi, "Summarization and visualization of multi-level and multi-dimensional itemsets," *Information Sciences*, vol. 520, pp. 63–85, May 2020, doi: 10.1016/j.ins.2020.02.006.
[9] D. R. Radev, S. Blair-Goldensohn, and Z. Zhang, "Experiments in single and multi-document summarization using MEAD," in *Proceedings of the first international conference on Human language technology research - HLT '01*, 2001, pp. 1–4.
[10] I.-C. Wu, C.-H. Tsai, and Y.-H. Lin, "Clustering and summarization topics of subject knowledge through analyzing internal links of Wikipedia," in *2013 IEEE 14th International Conference on Information Reuse & Integration (IRI)*, Aug. 2013, pp. 90–96, doi: 10.1109/IRI.2013.6642458.
[11] C. Wang, X. Yu, Y. Li, C. Zhai, and J. Han, "Content coverage maximization on word networks for hierarchical topic summarization," in *International Conference on Information and Knowledge Management, Proceedings*, Oct. 2013, pp. 249–258, doi: 10.1145/2505515.2505585.
[12] P. Yang, W. Li, and G. Zhao, "Language model-driven topic clustering and summarization for news articles," *IEEE Access*, vol. 7, pp. 185506–185519, 2019, doi: 10.1109/ACCESS.2019.2960538.
[13] T. Ma, H. Wang, Y. Zhao, Y. Tian, and N. Al-Nabhan, "Topic-based automatic summarization algorithm for Chinese short text," *Mathematical Biosciences and Engineering*, vol. 17, no. 4, pp. 3582–3600, 2020, doi: 10.3934/MBE.2020202.
[14] B. C. Wallace, S. Saha, F. Soboczenski, and I. J. Marshall, "Generating (factual?) narrative summaries of RCTs: Experiments with neural multi-document summarization," *AMIA Summits on Translational Science Proceedings*, vol. 2021, no. 5, pp. 605–614, 2021.
[15] X. Yan, Y. Wang, W. Song, X. Zhao, A. Run, and Y. Yanxing, "Unsupervised graph-based tibetan multi-document summarization," *Computers, Materials and Continua*, vol. 73, no. 1, pp. 1769–1781, 2022, doi: 10.32604/cmc.2022.027301.
[16] Á. Hernández-Castañeda, R. A. García-Hernández, Y. Ledeneva, and C. E. Millán-Hernández, "Language-independent extractive automatic text summarization based on automatic keyword extraction," *Computer Speech and Language*, vol. 71, p. 101267, Jan. 2021, doi: 10.1016/j.csl.2021.101267.
[17] N. Saini, S. Saha, D. Chakraborty, and P. Bhattacharyya, "Extractive single document summarization using binary differential evolution: optimization of different sentence quality measures," *PLoS ONE*, vol. 14, no. 11, p. e0223477, Nov. 2019, doi: 10.1371/journal.pone.0223477.
[18] P. Li, W. Cai, and H. Huang, "Weakly supervised natural language processing framework for abstractive multi-document summarization," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, Oct. 2015, vol. 19-23-Oct-, pp. 1401–1410, doi: 10.1145/2806416.2806494.
[19] X. Wang, J. Wang, B. Xu, H. Lin, B. Zhang, and Z. Yang, "Exploiting intersentence information for better question-driven abstractive summarization: algorithm development and validation," *JMIR Medical Informatics*, vol. 10, no. 8, p. e38052, Aug. 2022, doi: 10.2196/38052.
[20] L. Cagliero, P. Garza, and E. Baralis, "ELSA: a multilingual document summarization algorithm based on frequent itemsets and latent semantic analysis," *ACM Transactions on Information Systems*, vol. 37, no. 2, pp. 1–33, Jan. 2019, doi: 10.1145/3298987.
[21] X. Wang, J. Wang, B. Xu, H. Lin, B. Zhang, and Z. Yang, "Exploiting intersentence information for better question-driven abstractive summarization: algorithm development and validation," *JMIR Medical Informatics*, vol. 10, no. 8, p. e38052, Aug. 2022, doi: 10.2196/38052.
[22] Y. Cao *et al.*, "TASA: deceiving question answering models by twin answer sentences attack," in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 2022, pp. 11975–11992, doi: 10.18653/v1/2022.emnlp-main.821.
[23] G. Moro, L. Ragazzi, L. Valgimigli, and D. Freddi, "Discriminative marginalized probabilistic neural method for multi-document summarization of medical literature," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2022, vol. 1, pp. 180–189, doi: 10.18653/v1/2022.acl-long.15.
[24] L. Wang, H. Raghavan, V. Castelli, R. Florian, and C. Cardie, "A sentence compression based framework to query-focused multi-document summarization," *ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, vol. 1, pp. 1384–1394, 2013.
[25] Q. Wang, C. Nass, and J. Hu, "Natural Language query vs. keyword search: effects of task complexity on search performance, participant perceptions, and preferences," in *IFIP Conference on Human-Computer Interaction*, vol. 3585 LNCS, Springer Berlin Heidelberg, 2005, pp. 106–116.
[26] R. Campos, V. Mangaravite, A. Pasquali, A. Jorge, C. Nunes, and A. Jatowt, "YAKE! Keyword extraction from single documents using multiple local features," *Information Sciences*, vol. 509, pp. 257–289, Jan. 2019, doi: 10.1016/j.ins.2019.09.013.
[27] L. Kelebercová and M. Munk, "Search queries related to COVID-19 based on keyword extraction," *Procedia Computer Science*, vol. 207, pp. 2618–2627, 2022, doi: 10.1016/j.procs.2022.09.320.
[28] F. Boudin, "Unsupervised keyphrase extraction with multipartite graphs," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 2018, vol. 2, pp. 667–672, doi: 10.18653/v1/N18-2105.
[29] H. Shin, H. J. Lee, and S. Cho, "General-use unsupervised keyword extraction model for keyword analysis," *Expert Systems with Applications*, vol. 233, p. 120889, Dec. 2023, doi: 10.1016/j.eswa.2023.120889.
[30] R. Agrawal, H. Mishra, I. Kandasamy, S. R. Terni, and W. B. Vasantha, "Revolutionizing subjective assessments: a three-pronged comprehensive approach with NLP and deep learning," *Expert Systems with Applications*, vol. 239, p. 122470, Apr. 2024, doi: 10.1016/j.eswa.2023.122470.
[31] M. Song, Y. Feng, and L. Jing, "A survey on recent advances in keyphrase extraction from pre-trained language models," in *Findings of the Association for Computational Linguistics: EACL 2023*, 2023, pp. 2153–2164, doi: 10.18653/v1/2023.findings-eacl.161.

BIOGRAPHIES OF AUTHORS

Bambang Subeno is a doctoral student at the Faculty of Computer Science, Universitas Indonesia. He is a member of the IR-NLP Laboratory. He is also a lecturer at Telkom University, Bandung. His research interests include text mining, information extraction, information retrieval, and natural language processing. He can be contacted at email: bambang.subeno.if@gmail.com.

Prof. Dr. Indra Budi is a lecturer in computer science and information systems at the Faculty of Computer Science, Universitas Indonesia. He is also the head of the Information Retrieval and Natural Language Processing (IR-NLP) Laboratory. His research fields include information extraction, text mining, e-commerce, sentiment analysis, and social network analysis. He can be contacted at email: indra@cs.ui.ac.id.

Evi Yulianti is a lecturer and researcher at the Faculty of Computer Science, Universitas Indonesia. She received the B.Comp.Sc. degree from the Universitas Indonesia in 2010, the dual M.Comp.Sc. degree from Universitas Indonesia and Royal Melbourne Institute of Technology University in 2013, and the Ph.D. degree from Royal Melbourne Institute of Technology University in 2018. Her research interests include information retrieval and natural language processing. She can be contacted at email: evi.y@cs.ui.ac.id.