2021 Fifth International Conference on Information Retrieval and Knowledge Management (CAMP) | 978-1-6654-1237-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/CAMP51653.2021.9498029

Automatic Text Summarization for Malay News Documents Using Latent Dirichlet Allocation and Sentence Selection Algorithm

Nurazzah Abd Rahman
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
Shah Alam, Selangor, Malaysia
nurazzah@uitm.edu.my

Siti Nur Afiqah Ramlam
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
Shah Alam, Selangor, Malaysia
2016645066@isiswa.uitm.edu.my

Natasha Aleza Azhar
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
Shah Alam, Selangor.Malaysia
2020128021@isiswa.uitm.edu.my

Haslizatul Mohamed Hanum
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
Shah Alam, Selangor, Malaysia
haslizatul @uitm.edu.my

Noor Ida Ramli
Akademi Pengajian Bahasa
Universiti Teknologi MARA
Shah Alam, Selangor, Malaysia
noor032@uitm.edu.my

Najahudin Lateh
Akademi Pengajian Islam Kontemporari
Universiti Teknologi MARA
Shah Alam, Selangor, Malaysia
najahudin@uitm.edu.my

***Abstract*—The proliferation of internet newspapers making an Automatic Text Summarization is now a need to produce a summary that contains most of the important information from the original document. This study focused on the keyword extraction using Latent Dirichlet Allocation and Sentence Selection that used rule based concept approach to produce extractive summary. 100 Malay news documents covering general, sports, health and technology were collected from Utusan Online to evaluate the effectiveness of the system. This study used a single topic from LDA and top 10 words in the selected topic as the keywords. To evaluate, summary generated by the system was compared to summary generated by human expert using Precision Recall formula. The results showed the effectiveness of the summary generated by the system which is the best score 62.7 % that can help people read the Malay news documents in short time as the summary assist the readers to understand the important parts of the document without reading the whole document.**

Keywords—Information Retrieval, Text Summarization, Topic Modelling, LDA, Sentence Selection Algorithm, Malay Document Retrieval

I. Introduction

The purpose of automatic text summarization is to generate summary from a single document or multiple documents that should express the whole content in minimum number of words without losing its information content [1]. According to a study by Lee & Kim, extracting keywords from a massive amount of online news data is very useful because it can produce a short summary of news documents [2]. Therefore, an automated process that extracts keywords from news articles need to be established [2]. Automatic text summarization is one of the applications for text mining that can use the assistance of automatic keyword extraction to do the summary [3] [4]. It is a process of selecting words and phrases from the text document that can be the core sentiment of the document without any human intervention [5] [6].

From the keywords, it can summarize the content of articles and reflect the topic of articles [7]. Keyword extraction is important in order to identify the relative information in documents that contain some significant terms or words that best express the main point in the document [3] and keywords are a quick and efficient way of indicating the main topics of a text [8]. Automatic keyword extraction is the task to identify a small set of words, key phrases, keywords, or key segments from a document that can describe the meaning of the document [6].

Many text mining applications like automatic indexing, automatic summarization, automatic classification, and automatic clustering can take advantage from the automatic keyword extraction [3]. Therefore, the automatic text summarization can use the enormous assistance of the automatic keyword extraction.

This study mainly focuses on the role of Latent Dirichlet Allocation (LDA) as topic modelling for extracting the keywords from the documents [9-12]. The extracted keywords were used in Sentence Selection algorithm to summarize the Malay news documents [13-14].The objectives of this study are to identify the keywords from Malay news documents using LDA, to develop an automatic text summarization prototype for Malay news documents using LDA and Sentence Selection algorithm and to evaluate the effectiveness of automatic text summarization for summarizing Malay news documents.

The significance of this study is LDA technique that was applied as topic modelling to automatically extract keywords in the documents can identify the main points or important keywords [15]. The extracted keywords then, were used in Sentence Selection algorithm to create the extractive summary automatically without human guidance [16]. This study also help people read the Malay news documents in short time as the summary may assist the readers to understand the important parts of the texts without reading from the beginning to the end [17]. This study also contribute a lot in Natural Language Processing (NLP) [13][18] especially in summarization since there are only few works on Malay text summarization and no dedicated tool is immediately accessible to alleviate the reading of lengthy Malay texts [17]. [19-20] also use text summarization as hierarchical fuzzy logic ranking indicator on Malay text corpus.

II. Research Method

The whole process of text summarization is illustrated in Fig. 1. There were three processes involved which are preprocessing, keyword extraction using LDA algorithm and Sentence Selection to produce extractive summary from the input documents. For summarizer, the input was documents from the test collection. Each document consists of 15 – 60 sentences.

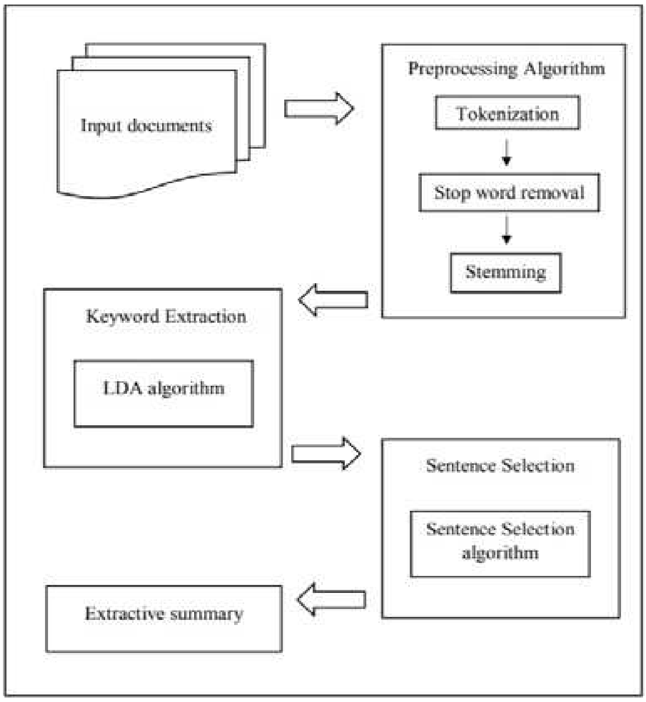

Fig. 1. Automatic text summarization architecture [17] [12]

This research focused on 100 Malay news documents covering general, sports, health and technology were collected from Utusan Online. The news from Utusan Online are extracted to txt file. For preprocessing, Tokenization breaks down paragraphs into sentences and each sentence is broken down into individual words or tokens [21] [22]. Stop words can be eliminated using a list of stop word such as "berkata", "seorang" and "katanya" and the advantage of using stop words is that it could reduce the number of terms that identify the document [22]. In stemming, all words are reduced to its root form [21]. For Malay language, words like "memakan", "dimakan" are reduced to their root form "makan" [23-24].

In keyword extraction process, LDA algorithm used to extract the keywords on the test [25]. Firstly, required parameters were set and trained [26] [27]. The results from the LDA were 10 topics and multiple words with weightage for each topic. Then topic composition was done to see which topic compose the documents [9] [10] [27]. The result file was imported into an output and sorted from highest to lowest probability for each document. The highest probability for the following topic indicated file which topic belong to that document [27]. For example, NGEN01.txt was assigned to topic 8 with 0.8935 weight probability and the selected top 10 words based on the LDA algorithm.

The selected top 10 words were used in the Sentence Selection algorithm. The keywords were mapped to the sentence in the original document to produce extractive summary. Rule based concept approach was used in this algorithm [28]. Rule based is "IF-THEN" condition. A few rules have been used in the algorithm. Threshold value was set to 2, if the sentence contains 2 mapped keywords, it will be selected as a part of the extractive summary. However, the generated summary must contain 30%, 40% and 50% from the original document. To comply these conditions, the algorithm used the rule based to check the highest number of mapped keywords in the sentence for the whole text and remove the lowest number of mapped keywords until the summary satisfied the conditions.

Relevant judgment also one of main element in this study [23]. This relevant judgment is produced by human expert to determine the relevant degree of summary resulted by the system with keyword extraction [29]. In this research, relevant judgment was determined by human expert from relevant Academy of Language Studies. Lastly, evaluation will take place to compare the efficiency of the system and the relevant judgment. This research using Precision-Recall Formula suggested by Zamin [30]. The result of the precision-recall formula shows the efficiency of the system. The following formula are applied:

Precision = correct / (correct + wrong)

Recall = correct / (correct + missed)

Score = 2 x (Precision x Recall) / (Precision + Recall)

Precision shows the accuracy of the extracted sentence, Recall reflects how many good sentences the system has missed, and Score is a weighted average of the Precision and Recall.

Where,

Correct is the number of sentences extracted by the system and the human. *Wrong* is the number of sentences extracted by the system but not by the human. *Missed* is the number of sentences extracted by the human but not by the system.

The generated summary is judged correct if it contains sentences that were tagged in the human's summary or partially correct if the summary provides sufficient context for the passage. The generated summary is judged wrong if needed context was totally misleading or if the summary did not contain the expected passage at all. Finally, the generated summary is judged wrong if there is insufficient context for the passage.

III. Result And Analysis

This section discussed on the result of LDA, sentence selection and evaluation. For LDA, there are three documents for each category used as examples. The results were based on the highest weight of topic probability for the documents and only top 10 terminologies were assigned in each topic as the keywords as shown in Table I.

Results shown in Table II are extractive summaries for NGEN01.txt document that automatically generated using Sentence Selection algorithm. The threshold value is set to 2. If the sentence only contains one keyword, the sentence will not be selected as extractive summary. The results also must follow the rule based that was applied in the process of selection.

TABLE I. LDA RESULTS BASED ON CATEGORY OF DOCUMENTS

Category of Documents	Documents Name	Topic	Weight of topic probability	Keywords (weight)
General News	NGEN01.txt	8	0.8935	tiaga (15.0) restoran (12.0) harga (8.0) makan (7.0) sawit (6.0) hidang (6.0) resit (5.0) malam (5.0) buah (5.0) caj (4.0)
	NGEN02.txt	9	0.7693	babit (10.0) operasi (10.0) siasat (9.0) letup (8.0) pagi (8.0) jalan (8.0) tiba (8.0) laku (8.0) jadi (8.0) kurang (8.0)
	NGEN03.txt	1	0.8346	anak (14.0) adik (10.0) kerja (9.0) hari (9.0) mohd (6.0) sama (6.0) zul (6.0) bapa (6.0) empat (6.0) pulang (5.0)
Sports News	NSPO01.txt	9	0.9697	dua (29.0) minit (27.0) jaring (19.0) gol (18.0) lepas (13.0) perak (12.0) terusi (12.0) terengganu (11.0) buat (9.0) penalti (9.0)
	NSPO02.txt	0	0.7854	atlet (10.0) badminton (5.0) pilih (5.0) latih (4.0) masa (4.0) negara (4.0) sihat (4.0) kerja (4.0) regu (3.0) pasang (3.0)
	NSPO03.txt	3	0.7669	negara (37.0) malaysia (29.0) sukan (28.0) lepas (25.0) tahun (24.0) jaya (20.0) hoki (17.0) dunia (15.0) aksi (15.0) peringkat (13.0)
Health News	NHEA01.txt	0	0.5430	produk (15.0) tiaga (9.0) operasi (7.0) lumpur (7.0) kuala (7.0) kosmetik (6.0) undang (6.0) biodegrasi (5.0) plastik (4.0) buah (4.0)
	NHEA02.txt	4	0.8174	frim (7.0) pokok (5.0) tanah (3.0) setiu (3.0) selidik (3.0) tanam (2.0) bris (2.0) kawasan (2.0) spf (2.0) stesen (2.0)
	NHEA03.txt	1	0.8335	kanak (10.0) anak (10.0) program (8.0) rawat (7.0) ijn (6.0) perlu (6.0) bantu (6.0) sakit (5.0) bedah (5.0) keluarga (5.0)
Technology News	NTEC01.txt	9	0.7531	angkasa (12.0) china (9.0) planet (8.0) bentuk (8.0) robot (8.0) kecil (7.0) selidik (7.0) misi (6.0) stesen (5.0) logam (5.0)
	NTEC02.txt	4	0.7218	siber (12.0) guna (8.0) kira (7.0) pintar (7.0) telefon (7.0) lamat (7.0) lalu (7.0) aplikasi (6.0) hack (5.0) maklumat (5.0)
	NTEC03.txt	4	0.73	siber (12.0) guna (8.0) kira (7.0) pintar (7.0) telefon (7.0) lamat (7.0) lalu (7.0) aplikasi (6.0) hack (5.0) maklumat (5.0)

The evaluation was made to calculate the effectiveness of automatic text summarization for Malay news documents that have been developed using Precision Recall formula as stated in Section II. Time constraint in construction of relevant judgment list by language expert only managed to summarize 20 documents manually. Therefore, the first three experiments conducted were made to compare the generated extractive summary with the relevant judgment list created previously. Three experiment were recorded for extractive summary of 30%, 40% and 50% length compared to original documents as illustrated in Table III, Table IV and Table V, respectively.

TABLE II. NGEN01.TXT EXTRACTIVE SUMMARIES

Length from original documents	Extractive Summary
30% (3 sentences)	Bermula dengan berniaga secara kecil-kecilan di bangunan Mara berhampiran stesen teksi Teluk Intan pada tahun 1990-an, Salwana Ali, 53, kini boleh berbangga kerana perniagaan makanan yang diusahakannya semakin maju. Saya memulakan perniagaan dengan menawarkan tomyam di restoran yang saya namakan Berkat Tomyam. Ketika itu gerai saya hanya ada empat meja sahaja dan dari situ saya mengembangkan perniagaan sehingga berjaya membuka restoran dengan 25 buah meja, katanya.
40% (5 sentences)	Bermula dengan berniaga secara kecil-kecilan di bangunan Mara berhampiran stesen teksi Teluk Intan pada tahun 1990-an, Salwana Ali, 53, kini boleh berbangga kerana perniagaan makanan yang diusahakannya semakin maju. Saya memulakan perniagaan dengan menawarkan tomyam di restoran yang saya namakan Berkat Tomyam. Ketika itu gerai saya hanya ada empat meja sahaja dan dari situ saya mengembangkan perniagaan sehingga berjaya membuka restoran dengan 25 buah meja, katanya. Memang banyak cabaran dan dugaan dalam perniagaan yang menjadi asam garam dan pahit manis dalam kehidupan saya sebagai usahawan, katanya. Di restoran miliknya, selain tomyam, Salwana menyajikan pelbagai menu lain antaranya hidangan istimewa nasi Wana, nasi goreng dan kambing bakar menggunakan arang.
50% (6 sentences)	Bermula dengan berniaga secara kecil-kecilan di bangunan Mara berhampiran stesen teksi Teluk Intan pada tahun 1990-an, Salwana Ali, 53, kini boleh berbangga kerana perniagaan makanan yang diusahakannya semakin maju. Saya memulakan perniagaan dengan menawarkan tomyam di restoran yang saya namakan Berkat Tomyam. Ketika itu gerai saya hanya ada empat meja sahaja dan dari situ saya mengembangkan perniagaan sehingga berjaya membuka restoran dengan 25 buah meja, katanya. Memang banyak cabaran dan dugaan dalam perniagaan yang menjadi asam garam dan pahit manis dalam kehidupan saya sebagai usahawan, katanya. Di restoran miliknya, selain tomyam, Salwana menyajikan pelbagai menu lain antaranya hidangan istimewa nasi Wana, nasi goreng dan kambing bakar menggunakan arang. Restoran ini merupakan perniagaan keluarga.

TABLE III. RESULT EXPERIMENT 1 FOR 30% LENGTH OF SUMMARY

Document Name	Precision	Recall	Score
NGEN01	0.667	0.400	0.500
NGEN02	0.750	0.273	0.400
NGEN03	0.750	0.231	0.353
NGEN04	1.000	0.286	0.444
NGEN05	0.667	0.400	0.500
NGEN06	1.000	0.667	0.800
NSPO02	0.667	0.400	0.500
NSPO03	1.000	0.400	0.571
NSPO04	1.000	0.286	0.444
NSPO05	1.000	0.500	0.667
NSPO06	1.000	0.333	0.500
NSPO22	1.000	0.250	0.400
NHEA01	0.500	0.333	0.400
NHEA02	1.000	0.286	0.444
NHEA03	0.333	0.125	0.182
NHEA04	0.667	0.222	0.333
NHEA08	1.000	0.400	0.571
NHEA16	1.000	0.500	0.667
NHEA17	1.000	1.000	1.000
NTEC01	0.500	0.143	0.222
NTEC16	0.000	0.000	0.000
NTEC17	0.333	0.250	0.286
NTEC18	0.667	0.333	0.444
NTEC19	0.333	0.167	0.222
NTEC20	1.000	0.333	0.500
Average	**0.753**	**0.341**	**0.454**

Fig. 2 shows the average score for the experiments done. Overall results show that Technology news shows the worst score. For example, NTEC19 only score 16.7% for all experiments. This might due to the word or term that have been used in the news are mostly number. However, for General and Sport news, the results were good. Experiment three obtained the best score which is 62.7% for the evaluation of effectiveness for the automatic text summarization for Malay news documents that have been developed. It can be concluded that automatic text summarization must produce extractive summary in 50% length from the original document to obtain a good summary result. In overall, the results shows all requirements are successfully tested.

TABLE IV. RESULT EXPERIMENT 2 FOR 40% LENGTH OF SUMMARY

Document Name	Precision	Recall	Score
NGEN01	0.600	0.600	0.600
NGEN02	0.800	0.364	0.500
NGEN03	0.833	0.385	0.526
NGEN04	1.000	0.286	0.444
NGEN05	0.500	0.400	0.444
NGEN06	1.000	0.833	0.909
NSPO02	0.800	0.800	0.800
NSPO03	0.667	0.400	0.500
NSPO04	1.000	0.429	0.600
NSPO05	1.000	0.667	0.800
NSPO06	1.000	0.333	0.500
NSPO22	1.000	0.250	0.400
NHEA01	0.500	0.333	0.400
NHEA02	1.000	0.429	0.600
NHEA03	0.600	0.375	0.462
NHEA04	0.750	0.333	0.462
NHEA08	1.000	0.600	0.750
NHEA16	1.000	1.000	1.000
NHEA17	0.750	1.000	0.857
NTEC01	0.667	0.286	0.400
NTEC16	0.500	0.500	0.500
NTEC17	0.500	0.500	0.500
NTEC18	0.750	0.500	0.600
NTEC19	0.200	0.167	0.182
NTEC20	1.000	0.667	0.800
Average	**0.777**	**0.497**	**0.581**

TABLE V. RESULT EXPERIMENT 3 FOR 50% LENGTH OF SUMMARY

Document Name	Precision	Recall	Score
NGEN01	0.667	0.800	0.727
NGEN02	0.667	0.364	0.471
NGEN03	0.875	0.538	0.667
NGEN04	1.000	0.429	0.600
NGEN05	0.600	0.600	0.600
NGEN06	1.000	0.833	0.909
NSPO02	0.667	0.800	0.727
NSPO03	0.750	0.600	0.667
NSPO04	1.000	0.571	0.727
NSPO05	1.000	0.833	0.909
NSPO06	1.000	0.333	0.500
NSPO22	1.000	0.500	0.667
NHEA01	0.333	0.333	0.333
NHEA02	1.000	0.571	0.727
NHEA03	0.500	0.375	0.429
NHEA04	0.800	0.444	0.571
NHEA08	1.000	0.600	0.750
NHEA16	1.000	1.000	1.000
NHEA17	0.600	1.000	0.750
NTEC01	0.750	0.429	0.545
NTEC16	0.500	0.500	0.500
NTEC17	0.400	0.500	0.444
NTEC18	0.800	0.667	0.727
NTEC19	0.167	0.167	0.167
NTEC20	1.000	0.667	0.800
Average	**0.763**	**0.578**	**0.637**

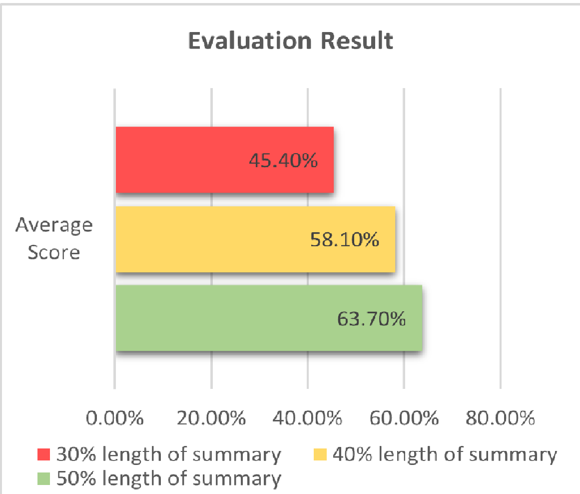

Fig. 2. Chart of average score for each experiment

IV. CONCLUSION

Objectives of this study has been successfully achieved. Relevant keywords have been identified using LDA algorithm. For comparative study, a prototype for the automatic text summarization for Malay news document using LDA and Sentence Selection algorithms has been developed. The prototype managed to get the best score 62.7% with the extractive summary length is 50% from the original document. In conclusion, it is clearly shown that human takes time to do the summarization. In this case, the human expert for this study cannot managed to summarize manually 100 documents within the given stipulated time. The human expert only managed to summarize 20 documents. We can conclude that manual summarization is difficult to do and time consuming, and the existence of automatic text summarization will assist in many ways for human being. However, LDA only generate a keyword from the selected document. In suggestion, the result for LDA keywords can be improve by combining LDA with word embedding to find the dependency of words. Based on the findings, we will adopt LDA and word embedding as part of our future work.

ACKNOWLEDGMENT

This research is based upon work supported by Universiti Teknologi MARA BESTARI Grant 600IRMI/DANA 5/3/BESTARI(112/2018) and Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia.

REFERENCES

[1] Y. K. Meena, and D. Gopalani, “Evolutionary algorithms for extractive automatic text summarization,” Procedia Computer Science, vol. 48(C), pp. 244–249, 2015.

[2] S. Lee, and H. J. Kim, “News keyword extraction for topic tracking,” in Proceedings of the 4th International Conference on Networked Computing and Advanced Information Management, NCM 2008, vol. 2, pp. 554–559, 2008.

[3] H. M. M. Hasan, F. Sanyal, and D. Chaki, “A Novel Approach to Extract Important Keywords from Documents Applying Latent Semantic Analysis,” in 2018 10th International Conference on Knowledge and Smart Technology: Cybernetics in the Next Decades, KST 2018, pp. 117–122, 2018.

[4] M. Y. Day, and C. Y. Chen, “Artificial intelligence for automatic text summarization,” in Proceedings - 2018 IEEE 19th International Conference on Information Reuse and Integration for Data Science, IRI 2018, pp. 478–484, 2018.

[5] S. K. Bharti, K. S. Babu, and A. Pradhan, “Automatic Keyword Extraction for Text Summarization in Multi-document e-Newspapers Articles,” in Proceedings of the International Conference on Informatics and Analytics - ICIA-16, vol. 4(6), pp. 1–8, 2016.

[6] C. Zhang, “Automatic Keyword Extraction from Documents Using Conditional Random Fields,” Information Systems Journal, vol. 3(3), pp. 1169–1180, 2008.

[7] K. Lin, C. Gao, X. Wang, and M. Qiu, “Research on Keyword Extraction Algorithm for Chinese Text Based on Document Topic Structure and Semantics,” In 2018 13th International Conference on Computer Science & Education (ICCSE), pp. 237–242, 2018.

[8] A. C. Junior, C. Magalhães, H. Caseli, and R. Zangirolami, “Topic Modeling for Keyword Extraction: using Natural Language Processing methods for keyword extraction,” Portal Min@s. Revista De Estudos Da Linguagem, vol. 23(3), pp. 695, 2015.

[9] P. M. Prihatini, I. Putra, I. A. D. Giriantari, and M. Sudarma, “Indonesian Text Feature Extraction using Gibbs Sampling and Mean Variational Inference Latent Dirichlet Allocation,” In 2017 15th International Conference on Quality in Research (QiR): International

Symposium on Electrical and Computer Engineering, pp. 40-44. IEEE, 2017.

[10] C. Hsu, “A Hybrid Latent Dirichlet Allocation Approach for Topic Classification,” in 2017 IEEE International Conference on Innovations in Intelligent SysTems and Applications (INISTA), pp. 312–315, 2017.

[11] X. Cai, and S. Cao, “A keyword extraction method based on learning to rank,” in Proceedings - 2017 13th International Conference on Semantics, Knowledge and Grids, SKG 2017, pp. 194–197, 2018..

[12] E. Laoh, I. Surjandari, and L. R. Febirautami, “Indonesian’s Song Lyrics Topic Modelling using Latent Dirichlet Allocation,” in 2018 5th International Conference on Information Science and Control Engineering (ICISCE), pp. 270– 274, 2018.

[13] K. Sintoris, and K. Vergidis, “Extracting business process models using natural language processing (NLP) techniques,” in Proceedings - 2017 IEEE 19th Conference on Business Informatics, CBI 2017, 1, 135–139, 2017.

[14] N. S. Syazhween Binti Zulkefli, N. B. Abdul Rahman, M. B. Puteh and Z. B. Abu Bakar, "Effectiveness of Latent Dirichlet Allocation Model for Semantic Information Retrieval on Malay Document," 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP), 2018, pp. 1-5.

[15] M. H. Haggag, “Keyword Extraction using Semantic Analysis,” International Journal of Computer Applications, vol. 61(1), pp. 975–8887, 2013.

[16] S. Siddiqi, & A. Sharan, (2015). “Keyword and Keyphrase Extraction Techniques: A Literature Review”. International Journal of Computer Applications, 109(2), 18–23.

[17] 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.

[18] W. Hong, X. Zheng, J. Qi, and W. Wang, "Project Rank: An Internet Topic Evaluation Model Based on Latent Dirichlet Allocation." In 2018 13th International Conference on Computer Science & Education (ICCSE), pp. 1-4, 2018.

[19] Rodzman, S.B.B., Ismail, N.K., Rahman, N.A., Nor, Z.M., & Noor, A.Y.M (2019). Domain Specific Concept Ontologies And Text Summarization As Hierarchical Fuzzy Logic Ranking Indicator On Malay Text Corpus. Indonesian Journal of Electrical Engineering and Computer Science, 2019, 15(3), pp. 1527–1534

[20] Bin Rodzman, S.B., Kamal Ismail, N., Rahman, N.A., ...Muhammad Naim Ku Khalif, K., Yunus Mohd Noor, A. (2019). Experiment With Text Summarization As A Positive Hierarchical Fuzzy Logic Ranking Indicator For Domain Specific Retrieval Of Malay Translated Hadith. ISCAIE 2019 - 2019 IEEE Symposium on Computer Applications and Industrial Electronics, 2019, pp. 299–304

[21] D. Hingu, D. Shah, and S. S. Udmale, “Automatic text summarization of Wikipedia articles,” in Proceedings - 2015 International Conference on Communication, Information and Computing Technology, ICCICT 2015, pp. 1–4, 2015.

[22] S. S. Naik, and M. N. Gaonkar, “Extractive text summarization by featurebased sentence extraction using rule-based concept,” in RTEICT 2017 - 2nd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, Proceedings, pp. 1364–1368, 2017.

[23] S. B. Rodzman, M. F. Izuan Abdul Ronie, N. K. Ismail, N. A. Rahman, F. Ahmad and Z. M. Nor, "Analyzing Malay Stemmer Performance Towards Fuzzy Logic Ranking Function on Malay Text Corpus," 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP), 2018, pp. 1-6.

[24] Z. A. Bakar and N. A. Rahman, "Evaluating the Effectiveness of Thesaurus and Stemming Methods in Retrieving Malay Translated Al-Quran Documents", International Conference on Asian Digital Libraries Lecture Notes in Computer Science 2911, pp. 653-662, 2003.

[25] D. Blei, and J. Lafferty, “Dynamic topic models,” in International Conference on Machine Learning, ACM, New York, NY, USA, pp. 113–120, 2016.

[26] D. M. Blei, and J. D. Lafferty, “Visualizing Topics with Multi-Word Expressions,” arXiv:0907.1013v1 [stat.ML], pp. 1–12, 2009.

[27] A. K. McCallum, " Mallet: A machine learning for language toolkit." http://mallet. cs. umass. Edu, 2002.

[28] L. Dut, “A Fast and Accurate Rule-Base Generation Method for Mamdani Fuzzy Systems,” IEEE Transactions on Fuzzy Systems, vol. 26(2), pp. 715–733, 2018.

[29] N. Moratanch, and S. Chitrakala, “A survey on extractive text summarization,” in International Conference on Computer, Communication, and Signal Processing: Special Focus on IoT, 2017.

[30] N. Zamin, and A. Ghani, “A Hybrid Approach for Malay Text Summarizer,” in Proceedings of the International Multi-Conference on Engineering and Technological Innovation, 2010.