ISSN 2278-3091
Volume 8, No.6, November – December 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse78862019.pdf
https://doi.org/10.30534/ijatcse/2019/78862019

Sequential Pattern Mining and Deep Learning to Enhance Readability of Indonesian Text Summarization

Dian Sa'adillah Maylawati[1,2,*], Yogan Jaya Kumar[3], Fauziah Binti Kasmin[4], Basit Raza[5]

[1]Centre for Advanced Computing Technology, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Malaysia, p031810030@student.utem.edu.my

[2]Department of Informatics, UIN Sunan Gunung Djati Bandung, Indonesia, diansm@uinsgd.ac.id

[3]Centre for Advanced Computing Technology, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Malaysia, yogan@utem.edu.my

[4]Centre for Advanced Computing Technology, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Malaysia, fauziah@utem.edu.my

[5] Department of Computer Science, COMSATS University Islamabad (CUI), Islamabad, 45550, Pakistan: basit.raza@comsats.edu.pk

ABSTRACT

Readability is a big challenge that must be solved in automatic text summarization research. The aim of this study is to comprehensively and systematically investigate the literatures that are related with text summarization needs, in particular, to prepare efficient algorithm for Indonesian text using Deep Learning and Sequential Pattern Mining. Evidence from previous literatures shows that there are several studies which use Sequential Pattern Mining to extract representation of text for document clustering and classification of Indonesian text. However, not much attention was given to text representation in document summarization, specifically for Indonesian language. As readability is a major concern in the text summarization community, determining a better text representation to maintain the meaning of the generated text summary is deemed necessary. This paper gives an opportunity to take a deeper look into how to design an efficient and effective text representation which can further enhance text summarization readability. Besides the general systematic literature review, we discuss an idea to combine Deep Learning and Sequential Pattern Mining to improve readability of summary result to develop in the future.

Key words: readability, sequential pattern, systematic review, text mining, text representation, text summarization

1. INTRODUCTION

One of the text analytic and natural language processing tasks that extract the main idea or key essance from documents using machine learning technique is called automatic text summarization [1], [2]. In its development, many techniques to summarize single or multiple text documents have been proposed. The big challange for text summarization is not only to focus in the summary information, but readability is also an important aspect [3], [4]. Therefore, it is important to prepare text representation to minimize the gap between summary result and reader understanding. Text representation is important so that the meaning of text is not lost. Moreover, text representation greatly determines the relationship between words, sentences, paragraphs and even between documents. We believe that the impact of well-prepared text representation in text summarization can improve its readability.

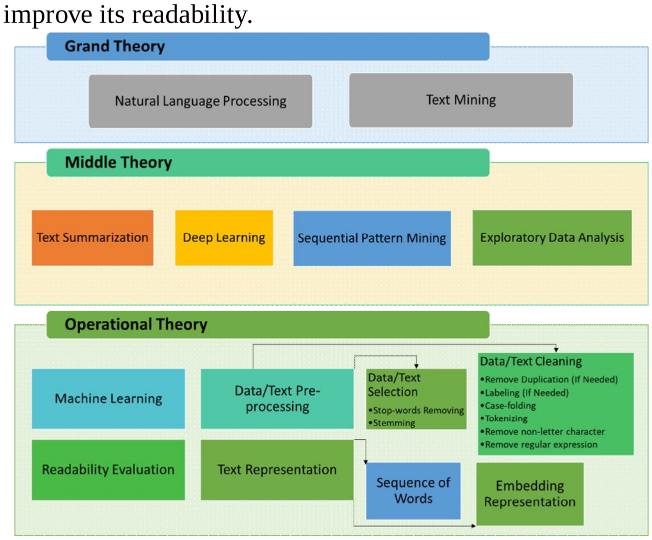

Figure 1: Classification of Research Literature/Theory

This paper brings the knowledge and patterns of the needs of text summarization research using current technologies. Figure 1 presents the classification of theories that is related to readability enhancement of Indonesian text summarization. The theories are divided into 3 class, among others Grand Theory, Middle Theory, and Operational Theory [5], [6]. Hierarchically, grand theory is a general theory as the main foundation of more specific theories in it. Middle theory is an elaboration of grand theory, while operational theory is where technical basic theories are immediately applied. The grouping of theories is very dependent on the problem to be solved. In this case, grand theory contains the main and basic theories of text summarization.

The grand theory of automated text summarization research is Natural Language Processing (NLP) and Text Mining which are part of Artificial Intelligent (AI) and Data Science studies. AI is a computational technology that has intelligent behavior to support daily social life with minimal human intervention [7]. While, data science concerns to prepare, process, analyze, manage, visualize, preservation, and collect insight information from large data collection (even big data) using mathematical or computational skill [8], [9]. Natural Language Processing is a specific technique in AI for language processing, either to analyze or discover the insight knowledge of data, such as text data and speech data [10]–[12]. In common, many NLP researches process the language using Text Mining technique (which is part of Data Mining).

In the middle theory, there are text summarization, Sequential Pattern Mining, Deep Learning and Exploratory Data Analysis (EDA). Sequential Pattern Mining (SPM) is a mining technique to find sequence of item in database transaction [13]–[16]. But, in Text Mining, Sequential Pattern Mining is used to prepare text representation that is commonly called Sequence of Words (SoW) [17]–[20]. Deep Learning is a popular method that is developed from Artificial Neural Networks (ANN) with multilayer between input layer and output layer [21]–[23] as illustrated in Figure 2. Although Deep Learning has been used for Natural Language Processing as ilustrated in Figure 3, there is still wide opportunity for it to be studied [10], [11]. Another middle theory is Exploratory Data Analysis to analyze and prepare data [24], [25], where different treatment is given between structured and unstructured data.

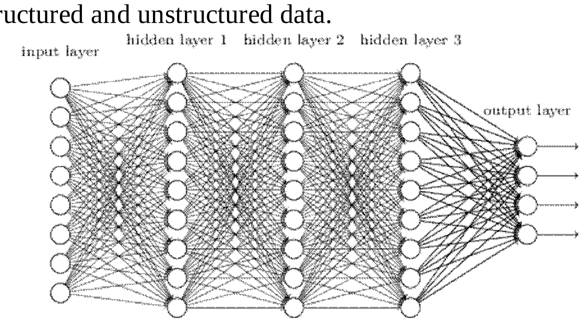

Figure 2: Deep Neural Network Model (http://neuralnetworksanddeeplearning.com/chap6.html)

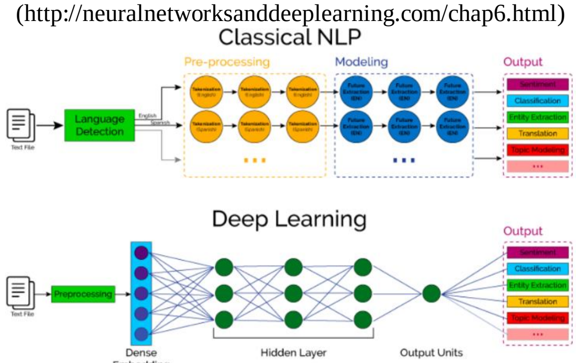

Figure 3: Deep Learning for NLP (https://www.upwork.com/hiring/for-clients/artificial-intelligence-and-natural-language-processing-in-big-data/)

In operational theory, there are Machine Learning theory, text pre-processing, text representation and readability evaluation. Machine Learning is widely used to imitate human behavior in solving problems or doing automation [26]–[28]. Machine Learning can process the various data types, either structured data, semi-structured data, or unstructured data. Text summarization today widely use Vector Space Model (VSM) and common progress in text summarization research use effective and efficient process, such as neuro-fuzzy [29], Recurrent Neural Network (RNN) [4], [30] and others. Text pre-processing is an important phase to prepare text data before moving to the next phase [31]–[34], including preparing structured text representation. This study will discuss types of text representation, including Sequence of Words and word embedding representation.

Sequence of Words (SoW) is a part of Multiple of Words (MoW) representation. Many researches in text processing use Multiple of Words as text representation compared to using Bag of Word (BoW) [35]–[39]. Multiple of Words pay attention to relations between words, even between sentences and paragraph. Sequential Pattern Mining technique produce Sequence of Words such as Frequent Word Sequence (FWS) [40], [41], Frequent Word Itemset (FWI) [42], and Maximal Frequent Word Sequence (MFS) which had been able to study text representation [19], [43]. Moreover, FWS and FWI were proven to maintain the meaning of text well, either for good grammar or unstructured grammar such as slang, especially for Indonesian language [35], [40]. Every language is unique and need different treatment occasionally, including Indonesian language that has its own characteristic, structure and grammar.

Therefore, based on the facts, it show there is still opportunity to use and improve Sequential Pattern Mining to extract better text representation, especially for Indonesian text. This study will elaborate and investigate some literatures to find better text representation for Indonesian text in keeping the meaning of text and achieve readability of summary. Then we will further investigate Deep Learning literatures as the current trend technique, either in Natural Language Processing or Text Mining. Then, we elaborate the literatures that support the opportunities to combine Deep Learning and Sequential Pattern Mining (DeepSPM) for future works.In the next section, this paper will present the methodology of the research, then the taxonomy of text summarization. Followed by structured representation for text, deep learning for text summarization, readability evaluation for NLP, and discussion about the related works. Lastly, the paper ends with conclusion and future works recommendation.

2. METHODOLOGY

A systematic approach will be used in this study to review some literatures on Text Summarization, structured representation for text, including SPM to produce text representation, SPM for Indonesian text, SPM for text summarization, DL for text summarization and readibility metrics to measure text summary result. Systematic approach is one of the research method that is used to collect the evidences by identifiying, assessing and interpreting those evidences to provide answers for specific research questions [44], [45].

In accordance with Figure 1, the SLR result of this research presents the literatures and related works that are classified into grand theory, middle theory and operational theory. Next, all of literatures are classified into seven categories based on the needs of research, among others: literature about SPM for text representation, SPM for Indonesian text, SPM for text summarization, Indonesian text summarization, text summarization using machine learning, text summarization using DL, and readability measurement and metrics for summary result. We will elaborate and investigate the findings from literatures and conclude the gap that can be solved.

Generally, this paper has inclusion and exclusion criteria of references as the scope of the relevant sources to be discussed. The inclusion criteria of this paper is to present the references of grand theory definition, then to discuss more details about trend (current research), types, models, techniques and implementations of middle and operational theories. For the exclusion criteria, this paper discussing in detail the related references to the operational theory used. In addition, DL and SPM theories and researches are provided for text data. Also, text representation that is generated from sequential pattern techniques is not exemplified using languages other than Indonesian language (Bahasa).

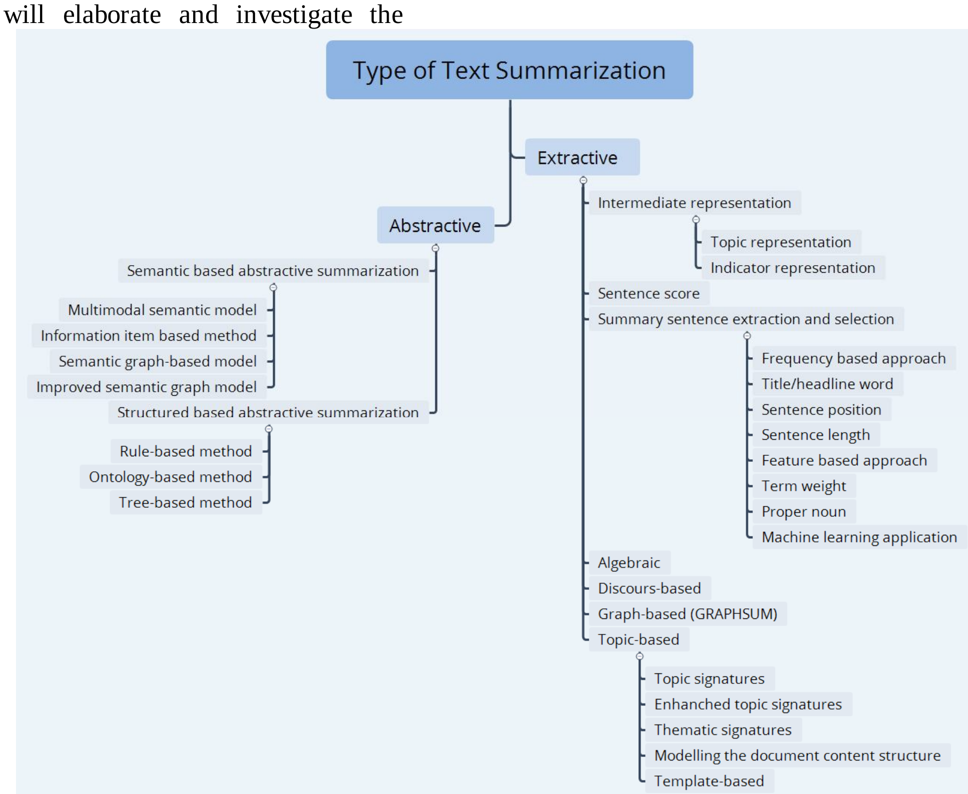

Figure 4: Taxonomy of text summarization types

3. TAXONOMY OF TEXT SUMMARIZATION

Text summarization is a technique to summarize a document automatically based on analytic result from the main idea or key essence of documents [1], [2]. Summarizing produces text document in a shorter form using sentence reduction by eliminating infrequent sentence or unimportant sentence for extractive summary; and restructuring or paraphrasing sentence using synonyms of some words that are used in the summary for abstractive summary [1], [46]–[49]. Besides sentence reduction and paraphrasing, there are several ways to summarize the text document, such as sentence split and join, generalization concept and specification concept [50].
Basically, automated text summarization has two types which are extractive and abstractive that have been described in Figure 4. Text summarization research is developed rapidly enough along with the development of text analytic research. The big challenge in text summarization research is how to produce readable summary [3], [4], [51], it means that the gap between summary result and reader understanding is not high. The elaboration of current text summarization can be made based on summarization approach (the way to summarize) [49], [50]; type of summary result [3], [52]–[54]; requirement of user [48], [55]; output style [48]; sentence weight [48], [56]; summarization impact [48]; available dataset [53], [56], [57] document type, summary unit [57], sources [57]; target and task-based [48]; language-based [48], [56] approach, objective of text summarization [48]; indicator representation [48], [52]; and document types [48], [57] as provided in Figure 5.

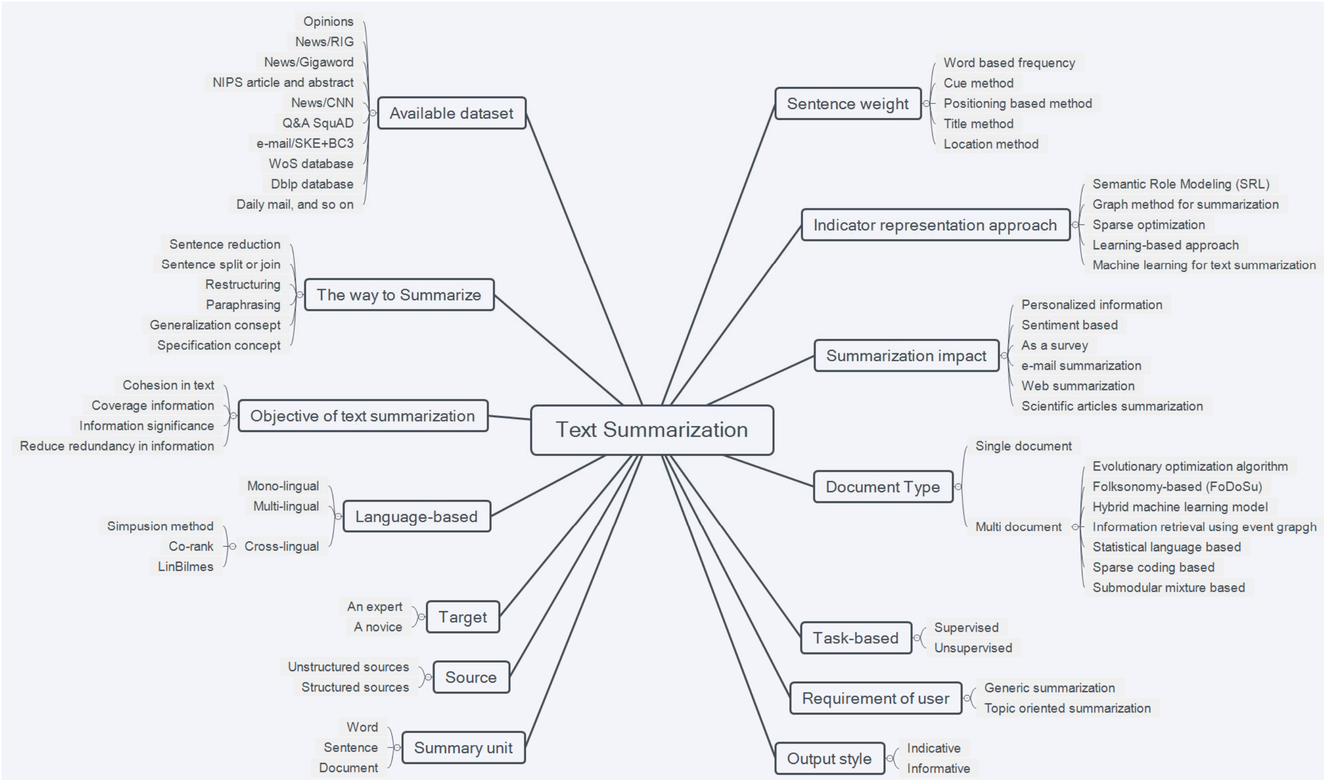

Figure 5: Taxonomy of text summarization based on document type, language, objective and indicator representation approach

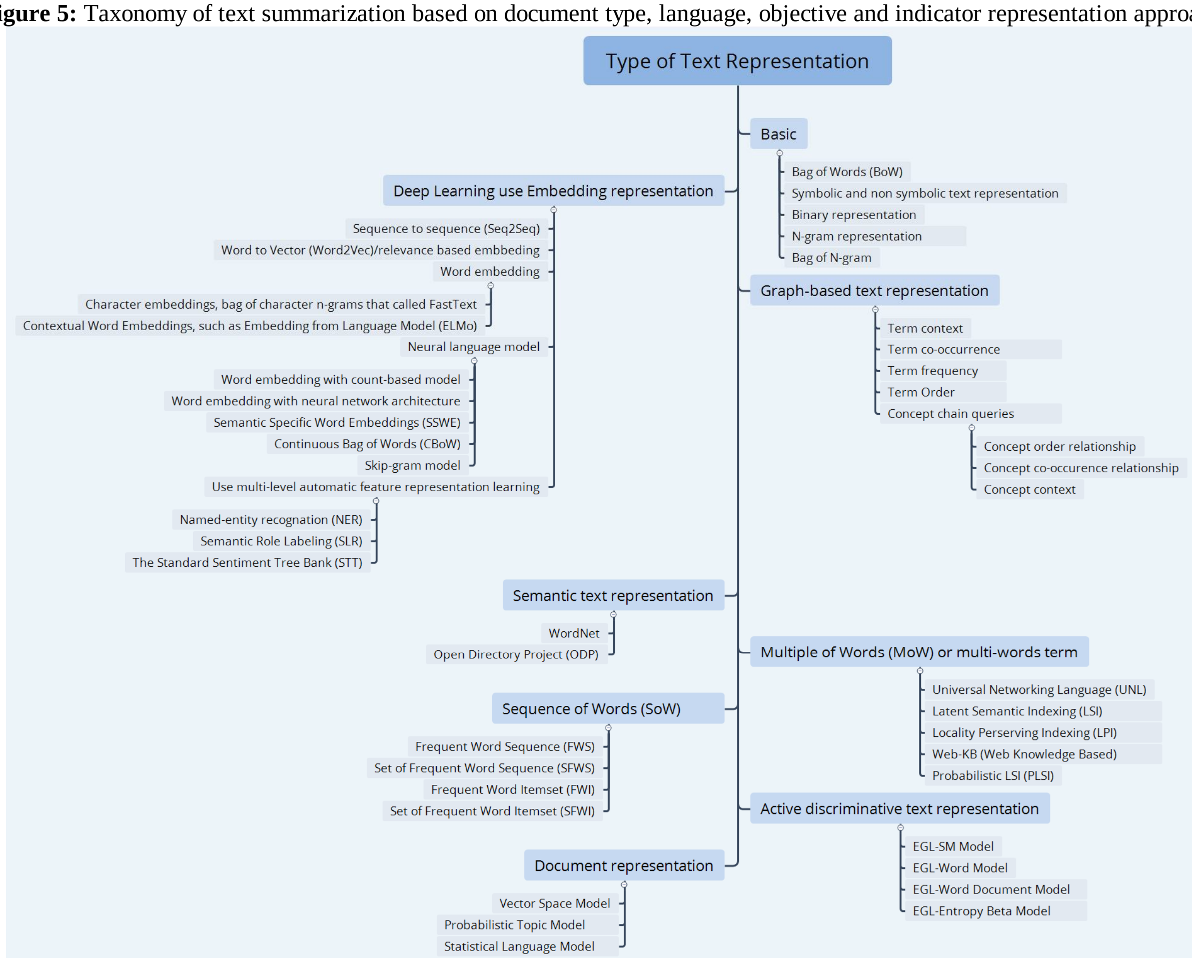

Figure 6: Structured Representation of Text Data

4. STRUCTURED REPRESENTATION FOR TEXT

Text is an unstructured data that must be structured prior to conducting the next process such as analyzing or mining. Text representation means the structured form of text data which is prepared from text pre-processing so that it is ready to be mined [35], [58]. Text representation has important role because it contains the meaning of text. Therefore, the impact of well-prepared text representation will affect the text analytics result, even in text summarization. Figure 6: Structured Representation of Text Datadescribes current research that explains and use several structured representation for text, such as: Bag of Words (BoW) [58], [59]; symbolic and non-symbolic text representation [60]; binary representation [58]; n-gram representation [58]–[61]; bag of n-gram [60]; semantic text representation [62]; sequence of words [17], [40], [63]–[66]; text representation for deep learning called embedding representation [11], [67]–[74]; multiple of words or multi-words term [58]; document representation [75]; active discriminative text representation [74]; and graph-based text representation [76]. Text representation was developed beginning from the BoW as the basic of text representation. However, BoW is considered weak in maintaining the meaning of the text. Whereas, the most frequently used are text representations that pay attention to the order in which words appear or their semantics include MoW, SoW, and semantic text representation.

SoW is one of multiple of word that represent text data by observing the order in which the words appear in the document. Moreover, SoW is developed not only for appearance of word, but also appearance of phrase and sentence in the document. SoW had been proven can increase the text mining result (either in classification or clustering) [17], [40], information retrieval [37], [39], [77], and even in text summarization [78], [79]. It is an evidance that SoW can maintain the meaning of text well. There are several previous research (provided in Figure 6). that use SoW as data text representation, or SPM as methodology to extract the SoW. However, not many research with Indonesian text that use SoW as text representation. Based on previous research, not all SoW type have been proven to maintain the meaning of Indonesian text data. FWS has been proven good in keeping the meaning of text with good grammar structure for English [65]. Moreover, FWS is developed to be Set of FWS (SFWS) that has been proven better than FWS in keeping the meaning of English text. However, FWS has been used for natural Indonesian text which contain many slang from social media [80]. For FWI, there is a research that uses FWI and improve it to be Set of FWI (SFWI) for natural Indonesian text with slang from social media [42]. This research proves that FWI can maintain the meaning of Indonesian text with slang well enough.

In text summarization, most of research use SoW for English. As far as this SLR is conducted, no research has been found that specifically uses SoW or SPM algorithms for Indonesian text summarization. There is an Indonesian text summarization study that uses "Frequent term" and used Tf-Idf to find frequent terms [81].

Basically, SPM algorithm produces sequential pattern that has several pattern, among others [13], [15], [20], [82]: itemset which is a set of items that are not empty, for example itemset denoted by i, where $i = (i_j, i_{j+1}, i_{j+2}, ..., i_n)$ and i_j are item, then the sequence is an ordered list of several itemsets, if the sequence is denoted by s, then $s = \{s_j, s_{j+1}, s_{j+2}, ..., s_n\}$ where s_j is an itemset. A sequence $A = \{a_1, a_2, a_3, ..., a_n\}$ is said to be subsequence of sequence $B = \{b_1, b_2, b_3, ..., b_n\}$ and B is supersequence of A, if integers $i_1 < i_2 < i_3 < ... < i_n$ and $a_1 \subseteq b_{i1}$, $a_2 \subseteq b_{i2}$, $a_3 \subseteq b_{i3}$, ..., $a_n \subseteq b_{in}$. For example, a sequence {(6) (7, 8) (10)} is subsequence of sequence {(6) (7, 8) (10) (7, 8, 9) (10)}, because (7) ⊆ (7, 8), (7, 8) ⊆ (7, 8, 9), and (10) ⊆ (10), but sequence {(6) (7)} is not subsequence of sequence {(6, 7)} and vice versa. This is because in sequence {(6) (7)} items appear in sequence one itemets with other itemsets, while in sequence {(6,7)} items appear together in one itemset.

In NLP research, SoW is used as text representation and also the algorithms which produce SoW are implemented to mine text data corpus. Generally, SoW can be implemented in any language, but specifically every language has different grammar structure that influences the meaning of text. For example, in English and Dutch language the structure is Explained-Explain, but in Indoensia and Germany language the structure is Explain-Explained, such as "beautiful girl" in English, but "gadis (girl) cantik (beautiful)" in Indonesia. Furthermore, there are many different grammar structure between every language. SoW can maintain the meaning of text, because "good research" and "research good" are different.

It is important to get the main idea of a document. Each paragraph will have its main idea, where it contains the main sentence and explanatory sentence [83], [84]. In Indonesian language, the main idea is not always in the first sentence, main idea can be in the beginning, in the end, or in both. The main idea that contains the meaning of text will have an impact on text summarization result. However, in automatic text summarization, the result is not only from one paragraph, but also from other related paragraphs.

Text Pre-processing is an important phase in natural language processing [85], [86]. Generally, text pre-processing prepares Indonesian document corpus with several process, among others tokenizing each paragraph and each sentence, lowering case, removing character non-letter and regular expression, removing Indonesian stopwords (remove unuseful words), and stemming using Porter algorithm for Indonesian language (where stemming process is the change of words with affixes into its basic words). Porter algorithm is widely used because it is simple and this algorithm has been adapted to the needs and rules of use of affixes in Indonesian language [87]. The example of Indonesian text pre-processing to get the SoW is illustrated below:

Document Example:
<u>Document 1:</u>
Untuk menyelesaikan Ph.D, pastikan lakukan yang Anda sukai dan yang benar-benar dapat Anda jangkau. Terus membaca dan menulis apapun setiap hari, meski hanya satu kalimat, tulis apa yang Anda inginkan, nikmati perjalanan studi Anda. Dan jangan biarkan diri Anda merasa sendirian!!!
(To complete a Ph.D, make sure to do what you interest with and that you really can catch up. Keep reading and writing anything every day, even if it is only one sentence, write what you want, enjoy your study trip. And don't let yourself feel alone !!!)
<u>Document 2:</u>
Mencari topik membuat tesis Ph.D harus sesuai kesukaan anda. Ditambah lagi rutin membaca dan menulis setiap hari, meski hanya satu kalimat, tulis apa yang Anda inginkan, nikmati saja perjalanan studi Anda.
(Looking for topics to make a Ph.D thesis should be your liking. Plus routine reading and writing every day, even if only one sentence, write what you want, just enjoy your study trip.)

Tokenizing, removing character non letter, regular expression, and lowering case:
<u>Document 1:</u>
<untuk menyelesaikan phd pastikan lakukan yang anda sukai dan yang benar benar dapat anda jangkau>
<terus membaca dan menulis apapun setiap hari meski hanya satu kalimat tulis apa yang anda inginkan nikmati perjalanan studi anda>
<dan jangan biarkan diri anda merasa sendirian>
<u>Document 2:</u>
<mencari topik membuat tesis phd harus sesuai kesukaan anda>
<ditambah lagi rutin membaca dan menulis setiap hari meski hanya satu kalimat tulis apa yang anda inginka nikmati saja perjalanan studi anda>

Removing stopwords:
<u>Document 1:</u>
<menyelesaikan phd pastikan lakukan sukai benar benar dapat jangkau>
<terus membaca menulis setiap hari satu kalimat tulis inginkan nikmati perjalanan studi>
<jangan biarkan diri merasa sendirian>
<u>Document 2:</u>
<mencari topik membuat tesis phd harus sesuai kesukaan>
<ditambah rutin membaca menulis setiap hari satu kalimat tulis inginkan nikmati perjalanan studi>

Stemming:
<u>Document 1:</u>
<selesai phd pasti lakukan suka benar benar dapat jangkau>
<terus baca tulis setiap hari satu kalimat tulis ingin nikmat jalan studi>
<jangan biar diri rasa sendiri>
<u>Document 2:</u>
<cari topik buat tesis phd harus sesuai suka>
<tambah rutin baca tulis setiap hari satu kalimat tulis ingin nikmat jalan studi>

The example of Sequence of Words from Document 1 and 2, if minimum support value is 75%:
{(phd)}: 2
{(phd, suka)}: 2
{(baca, tulis, nikmat, studi)}: 2
{(phd, suka), ((baca, tulis, nikmat, studi)}: 2

5. DEEP LEARNING FOR TEXT SUMMARIZATION

Many researches in text summarization use neural network, such as neuro-fuzzy, RNN and others. Figure 7 shows type of DL that have been used for NLP and text summarization that contain unsupervised deep learning [22], [88]; deep reinforced model with reinforced algorithm [89], [90]; deep generative model [91]–[93]; and many other deep learning methods [10], [55], [72], [94]–[96].

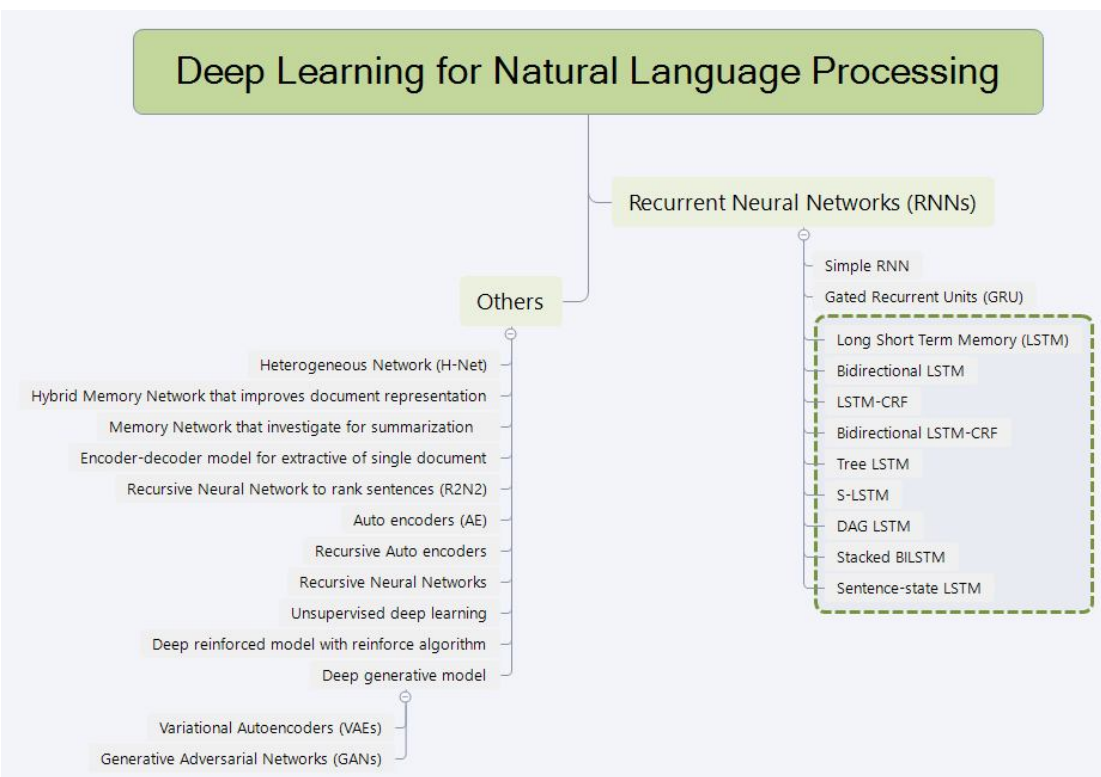

Figure 7: Scheme of Deep Learning for Natural Language Processing

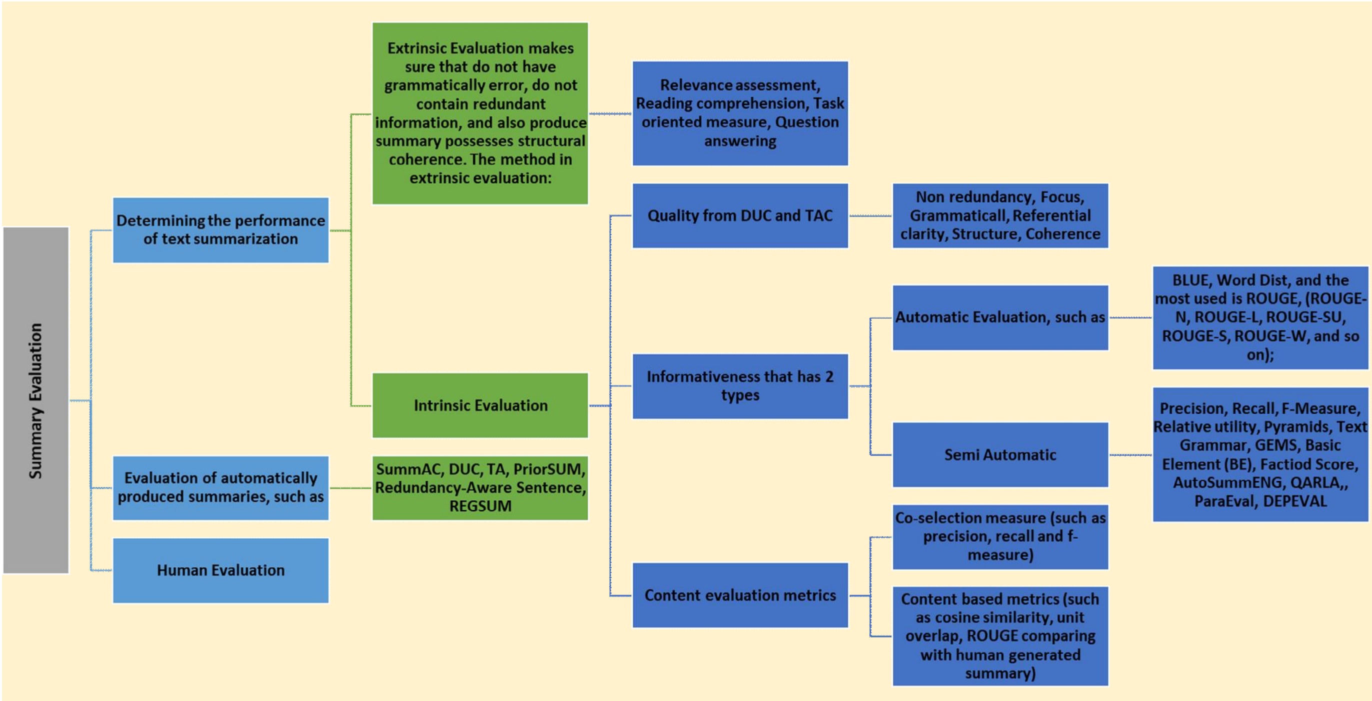

Figure 8: Types of summary evaluation

6. READABILITY EVALUATION FOR NATURAL LANGUAGE PROCESSING

Summary result are mostly measured with Recall-Oriented Understudy for Gisting Evaluation (ROUGE) metric [48], [97]–[99]. To measure readability of a summary, it is important to check the similarity of the summary with expert summary that is the gold standard. Direct survey or interview for the readers and expert can also be conducted to evaluate our summary result whether the generated summary is readable and easy to understand. Figure 8 describes types of summary evaluation that is divided into 3 general types, among others evaluation to determine the performance of text summary result, automated evaluation of summary result, and human evaluation [43], [45], [52], [54], [97].

7. TABULATION OF PREVIOUS RELATED WORKS

In this section, we summarize our related literatures findings from Table 1.

Table 1: SoW in Text Mining and NLP Research

Categories	Methods	Languages
Sequence of Words in Text Mining and Natural Language Processing Research	Frequent Adjacent Sequential Pattern (FASP) [36], Keyphrase Candidate Search using sequential Pattern (KCSP) [39], Sequential Pattern Mining Wildcard (SPMW) [77], Keyphrase Candidate Search using sequential Pattern (KCSP) [37], DIMASP-C and DIMASP-D [19], Prefix Span [17], Prefix span and KNN [40]	FASP is used in English and Malay, else in English
Sequence of Words in Sequence of Words in Text Mining and Natural Language Processing Research with Indonesian Text	Frequent Pattern Growth (FP-Growth) [100], BIDE – PrefixSpan – TruleGrowth [80], Frequent Pattern Growth (FP-Growth) and Compact Pattern Tree (CP-Tree) [101]	Indonesian
Text Summarization using Sequence of Words	Sequential Pattern Mining in English and Malay [79], Frequent Pattern and TF-IDF [102], FASP (Frequent Adjacent Sequential Pattern) and FASPe (Frequent Eliminated Pattern) for Malay [78], MWI-Sum (Multilingual Weighted Itemsetbased Summarizer) [103], Frequent itemset sequence generation algorithm [104], Term-based and ontology-based methods [105] Closed Frequent Patterm and Genetic Algorithm [106], Term Weighting and Sentence Selection [107].	Malay, English
Text Summarization using Machine Learning and Deep Learning	Naive Bayes Classifier [108], K-means [109], Ensable Noisy Auto-Encoder (ENAE) [96], Recurrent Neural Network (RNN) [30], [67], [110], Neuro-Fuzzy [29], ANFIS (Adaptive Neuro-Fuzzy Inference System) [111], Deep Learning [112], Deep Q-Network (DQN) [89], Timestap approach in Naive Bayes Classifier [113], Sentence similarity [114], Convolutional Neural Network (CNN) [67].	English

Categories	Methods	Languages
Indonesian Text Summarization using Machine Learning and Deep Learning	Vector Space Model (VSM) [115], Deep Learning [116], Latent Dirichlet Allocation (LDA) and Genetic Algorithm (GA) [117], Adaptive K-Nearest Neighbor (AKNN) [118], TF-IDF and Phrase Reinforcement algorithm [119], TF-IDF and Machine Learning [120], Naive Bayes Classifier [121], Frequent term based and Sentence weighting [81].	Indonesian

In text summarization research, SoW representation is also widely used, among others: representing sentence based on word sequence that can be used for any languages (tested with Spanish and English) [108]; multi document summarization using dual pattern-based as representation for sentence that implemented closed frequent pattern and sentence score ranking [106]; and MFS is used as model for English text summarization from single document based on sentence clustering using K-Means algorithm [109].

Frequent sequence is used as maximal frequent sentences for extracting text summary [107], and sequential sentence for multi document summarization to find the similarities and correlation between text resources using sequential pattern to keep the meaning of semantic knowledge of sentence in the multiple document (with Malay language) [79]. For another Malay language, sequential is used as text representation for text summarizer uses FASP (Frequent Adjacent Sequential Pattern) and FASPe (Frequent Eliminated Pattern) to represent Malay text and it focused on readability of text summary [78], and multilingual text summarization research uses itemset-based model to summarize the collection of document begin from the same topic.

The sentence selection process is analyzed based on frequent weighted itemsets and easily applicable for document collections in different languages, other than English [103]. There is a research that use sentence features (such as length, position, centroid, noun, and noun-verb pair) and frequent itemsets are combined to increase the quality of text summary result [104]. Moreover, frequent itemset representation has been used for text summarization with ItemSum as the itemset-based model to evaluate and select the most relevant sentences to include in the summary, where sentence evaluation and selection are scored using TF-IDF [102].

Specifically for Indonesian language, SoW had been used for document clustering [40]; Indonesian slang text representation [42]; semantic similarity from Indonesian text, not only semantic textual similarity but also explain about semantic similarity of pair sentence and pair of chunk [122]; and 5H1W (why, who, where, when, what, and how) event extraction from news based on sequence labeling from paragraph [116]. Several research on Indonesian text summarization, among others: generating abstract automatically for scientific article based on text summarization approach using Vector Space Model [123]; extracting important sentence for text summarization with counts noun and verb terms frequency [81]; Indonesian text summarization using hybrid Latent Dirichlet Allocation and Genetic Algorithm for weight of sentence feature determination [117].

Indonesian text summarization from social media (Twitter) using TF-IDF and Phrase Reinforcement algorithm that utilize Twitter hashtag had been conducted [119]. Naive Bayes algorithm is also used for Indonesian text summarization that determine sentence weight based on text features and Latent Semantic Analysis used to investigate the effect of semantic feature [121].

Term Frequency-Inverse Document Frequency (TF-IDF) is widely used in automatic text summarization [120]; sentence scoring and clustering for text summarization evaluation which proves that clustering is not significantly affect summarization result because the most important sentences are at the beginning of the paragraph (with English document) [111]; analyzing the opinion of consumer from online hotel review with text summarization approach [114]; generating scientific journal article using text summarization approach [118]; deep learning method for text summarization such as deep Auto Encoder and Ensamble Noisy Auto-Encoder [96], Recurrent Neural Networks [110], and neuro-fuzzy approach [29].

Several study about text summarization using deep learning method, among others: Text summarization using unsupervised deep learning that propose query-oriented extraction for document summarization with deep auto encoder (AE), they propose Ensable Noisy Auto-Encoder (ENAE) because AE adds noise to the input text and selects the top sentece from an enseble of noisy runs [96]; another neural network model for learning phrase representation called RNN Encoder-Decoder [110]. There is a study for removing redundant sentence for efficient text summarization using Restricted Boltzmann Machine (RBM) which is stochastic neural network [112]; Recurrent Neural Networks (RNNs) is also used in text summarization research to get meaningful summaries [30], [67]. Another research present text summarization using Deep Q-Network (DQN) for extractive summarization tasks [89].

Readability of text can be measured with various metrics, among others: ROUGE with its variants, such as ROUGE-1, ROUGE-2, and ROUGE-SU4 that evaluate f-measure [67], [98], [124], manual evaluation utilizing crowdsourcing tool to evaluate summary result in http://www.crowdflower.com/ [125]. Many research do evaluate summary result by human evaluation [81], [89], [113], [114], [125], [126].

The results of the literature review show that there are possibilities for further improvement with regard to summary quality. Specifically, text summarization, different language has its own characteristics, so the treatment will be different. Research on text summarization can also be done based on various points of view; each stage of the process, representation of the text, evaluating the summary results. This paper shows a variety of text representations that can be further investigated, such as the SPM technique that can be used to produce text representations sequentially by preserving the meaning of the text. Likewise, with the techniques used, Deep Learning is currently a popular technique for text summarization. Combining SPM and Deep Learning as Deep Sequential Pattern Mining (DeepSPM) could take the advantages of both approach to produce a better summarization model and worth to be investigated.

8. CONCLUSION

This paper presents the analyze and investigate result of the literatures related to text representation, Sequential Pattern Mining and Deep Learning for text summarization. The study found the opportunities in: (1) using Sequence of Words as text representation or Sequential Pattern Mining as a technique for Indonesian text summarization to achieve summary result readability; (2) this study offers a preliminary step to improve Deep Learning and Sequential Pattern Mining as Deep Sequential Pattern Mining (DeepSPM) that is expected to be able to produce readable text summary efficiently and effectively. Overall, this study can be used as foundation to design, implement, develop DeepSPM to enhance the efficiency and effectiveness of the automatic text summarization process for Indonesian languange with good quality of summary result. We regard this as our future work.

LIST OF ABRREVIATIONS

AI	=	Artificial Intelligent
AKNN	=	Adaptive K-Nearest Neighbor
ANFIS	=	Adaptive Neuro-Fuzzy Inference System
ANN	=	Artificial Neural Network
BIDE	=	Bi-Directional Extention
BoW	=	Bag of Words
CNN	=	Convolutional Neural Network
CP-Tree	=	Compact Pattern Tree
DL	=	Deep Learning
DQN	=	Deep Q-Networks
EDA	=	Exploratory Data Analysis
ENAE	=	Ensable Noisy Auto-Encoder
FASP	=	Frequent Adjacent Sequential Pattern
FASPe	=	Frequent Eliminated Pattern
FP-Growth	=	Frequent Pattern Growth
FPM	=	Frequent Pattern Mining
FWI	=	Frequent Word Itemsets
FWS	=	Frequent Word Sequence
GA	=	Genetic Algorithm
KCSP	=	Keyphrase Candidate Search using sequential Pattern
LDA	=	Latent Dirichlet Allocation
MFS	=	Maximal Frequent Sequence
ML	=	Machine Learning
MoW	=	Multiple of Words
MWI-Sum	=	Multilingual Weighted Itemsetbased Summarizer
NLP	=	Natural Language Processing
RBM	=	Restricted Boltzmann Machine
RNN	=	Recurrent Neural Networks
ROUGE	=	Recall-Oriented Understudy for Gisting Evaluation
SFWI	=	Set of Frequent Word Itemset
SFWS	=	Set of Frequent Word Sequence
SLR	=	Systematic Literature Review
SoW	=	Sequence of Words
SPM	=	Sequential Pattern Mining
SPMW	=	Sequential Pattern Mining Wildcard
TF-IDF	=	Term Frequency – Inverse Document Frequency
TM	=	Text Mining
VSM	=	Vector Space Model

ACKNOWLEDGMENT

This research work supported by Universiti Teknikal Malaysia Melaka (UTeM). We would like to thank to Muhammad Ali Ramdhani as a Professor of Information Technology in UIN Sunan Gunung Djati Bandung who is willing to become an expert in evaluating and giving many suggestions for this SLR study.

REFERENCES

[1] V. Thiagarajan, D. B. Shah, and N. Lalithamani, "Text summarization," *J. Adv. Res. Dyn. Control Syst.*, 2018.

[2] D. Sarkar, *Text Analytics with Python - A Practical Real-World Approach to Gaining Actionable Insights from Your Data*. 2016.

[3] Y. J. Kumar, O. S. Goh, H. Basiron, N. H. Choon, and P. C. Suppiah, "A review on automatic text summarization approaches," *Journal of Computer Science*. 2016. https://doi.org/10.3844/jcssp.2016.178.190

[4] J. ge Yao, X. Wan, and J. Xiao, "Recent advances in document summarization," *Knowl. Inf. Syst.*, 2017. https://doi.org/10.1007/s10115-017-1042-4

[5] M. A. Ramdhani, "Perancangan sistem pendukung keputusan kriteria majemuk pada pengambilan keputusan kelompok," Bandung, 2001.

[6] A. Ramdhani, M. Ramdhani, and A. Amin, "Writing a Literature Review Research Paper: A step-by-step approach," *Int. J. Basic Appl. Sci.*, 2014.

[7] H. Lu, Y. Li, M. Chen, H. Kim, and S. Serikawa, "Brain Intelligence: Go beyond Artificial Intelligence," *Mob. Networks Appl.*, 2018. https://doi.org/10.1007/s11036-017-0932-8

[8] S. Gowrishankar, A. Veena, S. Gowrishankar, and A. Veena, "Introduction to Data Science," in *Introduction to Python Programming*, 2019. https://doi.org/10.1201/9781351013239

[9] B. Manoj, K. V. K. Sasikanth, M. V. Subbarao, and V. Jyothi Prakash, "Analysis of data science with the use of big data," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 7, no. 6, pp. 87–90, 2018.

https://doi.org/10.30534/ijatcse/2018/02762018

[10] A. Kulkarni and A. Shivananda, "Deep Learning for NLP," in *Natural Language Processing Recipes*, 2019. https://doi.org/10.1007/978-1-4842-4267-4

[11] F. Chaubard, G. Genthial, R. Socher, M. Fang, R. Mundra, and K. Clark, "Natural Language Processing with Deep Learning," *CS224n*, 2017. .

[12] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: An introduction," *J. Am. Med. Inform. Assoc.*, 2011. https://doi.org/10.1136/amiajnl-2011-000464

[13] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," *J. Comput. Sci. Technol.*, vol. 15, no. 6, pp. 487–499, 1994.

[14] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and a I. Verkamo, "Fast discovery of association rules," *Advances in knowledge discovery and data mining*, vol. 12. pp. 307–328, 1996.

[15] R. Srikant and E. Agrawal, "Mining Sequential Patterns: Generalization and Performance Improvements," *5th Int. Conf. Extending Database Technol. (EDBT '96)*, pp. 3–17, 1996. https://doi.org/10.1007/BFb0014140

[16] J. Pei *et al.*, "Mining sequential patterns by pattern-growth: The prefixspan approach," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 11, pp. 1424–1440, 2004. https://doi.org/10.1109/TKDE.2004.77

[17] A. Hoonlor, "Sequential Patterns and Temporal Patterns for Text Mining," New York, 2011.

[18] H. Mahgoub, D. Rösner, N. Ismail, and F. Torkey, "A Text Mining Technique Using Association Rules Extraction," *Int. J. Comput. Intell.*, vol. 4, no. 1, pp. 21–28, 2008.

[19] R. A. García-Hernández, J. F. Martínez-Trinidad, and J. A. Carrasco-Ochoa, "Finding Maximal Sequential Patterns in text document collections and single documents," *Inform.*, 2010.

[20] D. S. Maylawati, H. Aulawi, and M. A. Ramdhani, "The Concept of Sequential Pattern Mining for Text," in *AASEC 2018*, 2018.

[21] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*. 2015.

[22] J. Ahmad, H. Farman, and Z. Jan, "Deep Learning Methods and Applications," in *SpringerBriefs in Computer Science*, 2019.

[23] J. R. B. Del Rosario, "Development of a face recognition system using deep convolutional neural network in a multi-view vision environment," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 3, pp. 369–374, 2019. https://doi.org/10.30534/ijatcse/2019/06832019

[24] M. Abzalov, "Exploratory data analysis," in *Modern Approaches in Solid Earth Sciences*, 2016.

[25] B. M. Bolker, "Exploratory Data Analysis and Graphics," in *Ecological Models and Data in R*, 2019.

[26] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, Fourth Edi. United States of America: Morgan Kaufman, 2012.

[27] H. Jiawei, M. Kamber, J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2006.

[28] J. Goyal and B. Kishan, "Progress on Machine Learning Techniques for Software Fault Prediction," *Prog. Mach. Learn. Tech. Softw. Fault Predict.*, vol. 8, no. 2, pp. 305–311, 2019. https://doi.org/10.30534/ijatcse/2019/33822019

[29] M. Azhari and Y. Jaya Kumar, "Improving text summarization using neuro-fuzzy approach," *J. Inf. Telecommun.*, vol. 1, no. 4, pp. 367–379, 2017.

[30] G. Rossiello, "Neural abstractive text summarization," in *CEUR Workshop Proceedings*, 2016.

[31] D. S. Maylawati, "Pembangunan Library Pre-Processing untuk Text Mining dengan Representasi Himpunan Frequent Word Itemset (HFWI) Studi Kasus: Bahasa Gaul Indonesia," Bandung, 2015.

[32] D. S. Maylawati, H. Aulawi, and M. A. Ramdhani, "Flexibility of Indonesian text pre-processing library," *Indones. J. Electr. Eng. Comput. Sci.*, 2019. https://doi.org/10.11591/ijeecs.v13.i1.pp420-426

[33] S. Vijayarani, J. Ilamathi, and M. Nithya, "Preprocessing Techniques for Text Mining - An Overview," *Int. J. Comput. Sci. Commun. Networks*, vol. 5, no. 1, pp. 7–16, 2015.

[34] S. Kannan *et al.*, "Preprocessing Techniques for Text Mining," *Int. J. Comput. Sci. Commun. Networks*, 2015.

[35] D. Sa'Adillah Maylawati and G. A. Putri Saptawati, "Set of Frequent Word Item sets as Feature Representation for Text with Indonesian Slang," in *Journal of Physics: Conference Series*, 2017.

[36] S. Alias, S. K. Mohammad, G. K. Hoon, and T. T. Ping, "A text representation model using Sequential Pattern-Growth method," *Pattern Anal. Appl.*, vol. 21, no. 1, pp. 233–247, 2018.

[37] Q. Wang, V. S. Sheng, and C. Hu, "Keyphrase Extraction Using Sequential Pattern Mining and Entropy," in *Proceedings - 2017 IEEE International Conference on Big Knowledge, ICBK 2017*, 2017. https://doi.org/10.1109/ICBK.2017.20

[38] F. Xie, X. Wu, and X. Zhu, "Efficient sequential pattern mining with wildcards for keyphrase extraction," *Knowledge-Based Syst.*, 2017.

[39] Q. Wang, V. S. Sheng, and X. Wu, "Keyphrase Extraction with Sequential Pattern Mining," in *AAAI*, 2017. https://doi.org/10.1109/ICBK.2017.20

[40] D. Rahmawati, G. A. P. Saptawati, and Y. Widyani, "Document Clustering using Sequential Pattern(SP)," *2015 Int. Conf. Data Softw. Eng. (ICoDSE 2015)*, pp. 98–102, 2015.

[41] H. Ahonen-Myka, "Discovery of Frequent Word Sequences in Text," *Proc. ESF Explor. Work. Pattern Detect. Discov.*, vol. {LNCS} (24, no. Teollisuuskatu 23, pp. 180–189, 2002.

[42] D. Sa'Adillah Maylawati and G. A. Putri Saptawati, "Set of Frequent Word Item sets as Feature Representation for Text with Indonesian Slang," in *Journal of Physics: Conference Series*, 2017, vol. 801, no. 1.

[43] H. Ahonen-Myka, "Finding All Maximal Frequent Sequences in Text," *Proc. ICML Work. Mach. Learn. Text Data Anal.*, pp. 11–17, 1999.

[44] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering - A systematic literature review," *Information and Software Technology*. 2009. https://doi.org/10.1016/j.infsof.2008.09.009

[45] B. A. Kitchenham, "Systematic review in software engineering," 2012.

[46] V. Gupta, N. Bansal, and A. Sharma, "Text Summarization for Big Data: A Comprehensive Survey," in *International Conference on Innovative Computing and Communications*, 2019. https://doi.org/10.1007/978-981-13-2354-6_51

[47] M. Allahyari *et al.*, "Text Summarization Techniques : A Brief Survey," 2017.

[48] M. Gambhir and V. Gupta, "Recent automatic text summarization techniques: a survey," *Artif. Intell. Rev.*, 2017.

[49] M. Roig, "Avoiding plagiarism, self-plagiarism, and other questionable writing practices: A guide to ethical writing," in *Getting to Good: Research Integrity in the Biomedical Sciences*, 2018.

[50] S. M. Alzahrani, N. Salim, and A. Abraham, "Understanding plagiarism linguistic patterns, textual features, and detection methods," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*. 2012.

[51] K. Nandhini and S. R. Balasundaram, "Improving readability through extractive summarization for learners with reading difficulties," *Egypt. Informatics J.*, vol. 14, no. 3, pp. 195–204, 2013.

[52] M. Allahyari *et al.*, "Text Summarization Techniques: A Brief Survey," *Int. J. Adv. Comput. Sci. Appl.*, 2017. https://doi.org/10.14569/IJACSA.2017.081052

[53] M. Patel, A. Chokshi, S. Vyas, and K. Maurya, "Machine Learning Approach for Automatic Text Summarization Using Neural Networks," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 7, no. 1, pp. 194–202, 2018.

[54] A. Khan *et al.*, "Abstractive Text Summarization based on Improved Semantic Graph Approach," *Int. J. Parallel Program.*, 2018.

[55] M. Day and C. Chen, "Artificial Intelligence for Automatic Text Summarization," *2018 IEEE Int. Conf. Inf. Reuse Integr.*, pp. 478–484, 2018.

[56] M. Gupta, "Automatic Text Summarization Manish Gupta," 2018.

[57] M. E. Johnson, "Automatic Summarization of Natural Language Literature Review and Synthesis," 2018.

[58] B. S. Harish, D. S. Guru, and S. Manjunath, "Representation and classification of text documents: A brief review," *IJCA, Spec. Issue Recent Trends Image Process. Pattern Recognit.*, no. 2, pp. 110–119, 2010.

[59] H. M. Wallach, "Topic Modeling: Beyond Bag-of-Words," *ICML*, no. 1, pp. 977–984, 2006.

[60] H. Schuetze, H. Adel, and E. Asgari, "Nonsymbolic Text Representation," *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguist. Comput. Linguist.*, vol. 1, pp. 785–796, 2017. https://doi.org/10.18653/v1/E17-1074

[61] A. Sethy and B. Ramabhadran, "Bag-of-word normalized n-gram models," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2008, pp. 1594–1597.

[62] N. H. Ali and N. S. Ibrahim, "Porter Stemming Algorithm for Semantic Checking," in *ICCIT*, 2012.

[63] A. Doucet, "Advanced document description, a sequential approach," University of Helsinki, Finland, 2006.

[64] A. Doucet and H. Ahonen-Myka, "An efficient any language approach for the integration of phrases in document retrieval," *Lang. Resour. Eval.*, vol. 44, no. 1–2, pp. 159–180, 2010.

[65] R. As. Nalistia, "Feature-Based Clustering untuk Pengelompokan Dokumen dengan Representasi Text Terstruktur," Bandung, 2008.

[66] D. S. Maylawati, "Pembangunan Library Pre-processing untuk Text Mining dengan Representasi Himpunan Frequent Word Itemset. Studi Kasus Bahasa Gaul Indonesia," Bandung Institute of Technology, 2015.

[67] T. Shi, Y. Keneshloo, N. Ramakrishnan, and C. K. Reddy, "Neural Abstractive Text Summarization with Sequence-to-Sequence Models," pp. 1–28, 2018.

[68] M. Kamkarhaghighi and M. Makrehchi, "Content Tree Word Embedding for document representation," *Expert Syst. Appl.*, 2017.

[69] M. Naili, A. H. Chaibi, and H. H. Ben Ghezala, "Comparative study of word embedding methods in topic segmentation," in *Procedia Computer Science*, 2017. https://doi.org/10.1016/j.procs.2017.08.009

[70] H. Zamani and W. B. Croft, "Relevance-based Word Embedding," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '17*, 2017.

[71] S. Banerjee, "Word2Vec — a baby step in Deep Learning but a giant leap towards Natural Language Processing," 2018. [Online]. Available: https://medium.com/explore-artificial-intelligence/word2vec-a-baby-step-in-deep-learning-but-a-giant-leap-towards-natural-language-processing-40fe4e8602ba.

[72] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing [Review Article]," *IEEE Comput. Intell.*

Mag., vol. 13, no. 3, pp. 55–75, 2018.

[73] M. Kageback, O. Mogren, N. Tahmasebi, and D. Dubhashi, “Extractive Summarization using Continuous Vector Space Models,” *Proc. 2nd Work. Contin. Vector Sp. Model. their Compos.*, 2014.

[74] Y. Zhang, M. Lease, and B. C. Wallace, “Active Discriminative Text Representation Learning,” *Proc. Thirty-First AAAI Conf. Artif. Intell.*, pp. 3386–3392, 2016.

[75] K. N. Singh and H. M. Devi, “Document representation techniques and their effect on the document Clustering and Classification : A Review,” *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 5, pp. 1780–1784, 2017.

[76] W. Jin and R. K. Srihari, “Graph-based text representation and knowledge discovery,” *SAC*, p. 807, 2007. https://doi.org/10.1145/1244002.1244182

[77] F. Xie, X. Wu, and X. Zhu, “Efficient sequential pattern mining with wildcards for keyphrase extraction,” *Knowledge-Based Syst.*, vol. 115, pp. 27–39, 2017.

[78] S. Alias, S. K. Mohammad, G. K. Hoon, and T. T. Ping, “A Malay text summarizer using pattern-growth method with sentence compression rules,” in *2016 3rd International Conference on Information Retrieval and Knowledge Management, CAMP 2016 - Conference Proceedings*, 2017.

[79] S. Alias and S. K. Muhammad, “Sequential pattern based multi document summarization - An exploratory approach,” in *International Conference on Research and Innovation in Information Systems, ICRIIS*, 2013.

[80] D. Sa’Adillah Maylawati, M. Irfan, and W. Budiawan Zulfikar, “Comparison between BIDE, PrefixSpan, and TRuleGrowth for Mining of Indonesian Text,” *J. Phys. Conf. Ser.*, vol. 801, no. 1, 2017.

[81] M. Fachrurrozi, N. Yusliani, and R. U. Yoanita, “Frequent Term based Text Summarization for Bahasa Indonesia,” *Int. Conf. Innov. Eng. Technol.*, pp. 30–32, 2013.

[82] P. Fournier-Viger, J. Chun, -Wei Lin, R. U. Kiran, Y. S. Koh, and R. Thomas, “A Survey of Sequential Pattern Mining,” *Ubiquitous Int.*, vol. 1, no. 1, pp. 54–77, 2017.

[83] Anonymous, “Ide Pokok, Pengetian, Cara Menentukan Ide Pokok, Ide Pokok dalam Paragraf Dedektif dan Paragraf Induktif, Main Idea,” *berbagaireviews.com*, 2018. [Online]. Available: https://www.berbagaireviews.com/2018/07/ide-pokok-pengetian-cara-menentukan-ide.html. [Accessed: 26-Jan-2019].

[84] Kresnoadi, “5 Langkah Menemukan Ide Pokok dalam Paragraf,” *Ruang Guru*, 2018. [Online]. Available: https://blog.ruangguru.com/5-langkah-menemukan-ide-pokok-dalam-paragraf. [Accessed: 26-Jan-2019].

[85] A. F. Hidayatullah and M. R. Ma’arif, “Pre-processing Tasks in Indonesian Twitter Message,” in *International Conference on Computing and Applied Informatics (ICCAI)*, 2016.

[86] D. S. Maylawati, H. Aulawi, and M. A. Ramdhani, “Flexibility of Indonesian text pre-processing library,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 13, no. 1, pp. 420–426, 2019.

[87] D. S. Maylawati, W. B. Zulfikar, C. Slamet, and M. A. Ramdhani, “An Improved of Stemming Algorithm for Mining Indonesian Text with Slang on Social Media,” in *The 6th International Conference on Cyber and IT Service Management (CITSM 2018)*, 2018. https://doi.org/10.1109/CITSM.2018.8674054

[88] L. Deng, “Deep Learning: Methods and Applications,” *Found. Trends® Signal Process.*, 2014.

[89] K. Yao, L. Zhang, T. Luo, and Y. Wu, “Deep reinforcement learning for extractive document summarization,” *Neurocomputing*, 2018.

[90] J. Schmidhuber, “Deep Learning in neural networks: An overview,” *Neural Networks*. 2015.

[91] L. Mescheder, S. Nowozin, and A. Geiger, “Adversarial variational bayes: Unifying variational autoencoders and generative adversarial networks,” in *34th International Conference on Machine Learning, ICML 2017*, 2017.

[92] S. Latif, R. Rana, J. Qadir, and J. Epps, “Variational autoencoders to learn latent representations of speech emotion,” in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2018.

[93] J. Luo and J. Huang, “Generative adversarial network: An overview,” *Yi Qi Yi Biao Xue Bao/Chinese Journal of Scientific Instrument*. 2019.

[94] Y. Zhang, Q. Liu, and L. Song, “Sentence-State LSTM for Text Representation,” *ACL2018*, 2018.

[95] R. Socher, “Recursive Deep Learning for Natural Language Processing and Computer Vision,” *PhD thesis*, 2014.

[96] M. Yousefi-Azar and L. Hamey, “Text summarization using unsupervised deep learning,” *Expert Syst. Appl.*, 2017. https://doi.org/10.1016/j.eswa.2016.10.017

[97] V. Gupta and G. S. Lehal, “A Survey of Text Summarization Extractive techniques,” in *Journal of Emerging Technologies in Web Intelligence*, 2010, vol. 2, no. 3, pp. 258–268.

[98] N. Nazari, M. A. Mah, and Davi, “A survey on Automatic Text Summarization,” *J. AI ang Data Min.*, vol. 7, no. 1, pp. 121–135, 2019.

[99] F. Kiyani and O. Tas, “A survey automatic text summarization,” *Pressacademia*, vol. 5, no. 1, pp. 205–213, 2017.

[100] D. S. Maylawati and G. A. P. Saptawati, “Set of Frequent Word Item sets as Feature Representation for Text with Indonesian Slang,” in *International Conference on Computing and Applied Informatics*, 2016, pp. 1–6.

[101] D. S. A. Maylawati, M. A. Ramdhani, A. Rahman, and W. Darmalaksana, “Incremental technique with set of frequent word item sets for mining large Indonesian text data,” in *2017 5th International Conference on Cyber*

and IT Service Management, CITSM 2017, 2017.

[102] E. Baralis, L. Cagliero, S. Jabeen, and A. Fiori, "Multi-document summarization exploiting frequent itemsets," in *Proceedings of the 27th Annual ACM Symposium on Applied Computing - SAC '12*, 2012.

[103] E. Baralis, L. Cagliero, A. Fiori, and P. Garza, "MWI-Sum: A Multilingual Summarizer Based on Frequent Weighted Itemsets," *ACM Trans. Inf. Syst.*, 2015.
https://doi.org/10.1145/2809786

[104] J. Jayabharathy, S. Kanmani, and Buvana, "Multi-document summarization based on sentence features and frequent itemsets," in *Advances in Intelligent and Soft Computing*, 2012.

[105] J. P. Qiang, P. Chen, W. Ding, F. Xie, and X. Wu, "Multi-document summarization using closed patterns," *Knowledge-Based Syst.*, vol. 99, pp. 28–38, 2016.

[106] Y. Wu, Y. Li, and Y. Xu, "Dual pattern-enhanced representations model for query-focused multi-document summarisation," *Knowledge-Based Syst.*, vol. 163, pp. 736–748, 2019.

[107] Y. Ledeneva, A. Gelbukh, and R. A. García-Hernández, "Terms derived from frequent sequences for extractive text summarization," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4919 LNCS, pp. 593–604, 2008.

[108] E. Villatoro-tello, L. V. Pineda, and M. Montes-y-gómez, "Using Word Sequences for Text Summarization," *TSD, Text, Speech Dialogue*, pp. 293–300, 2006.

[109] R. A. García-Hernández and Y. Ledeneva, "Word sequence models for single text summarization," *Proc. 2nd Int. Conf. Adv. Comput. Interact. ACHI 2009*, pp. 44–48, 2009.

[110] K. Cho *et al.*, "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1724–1734.
https://doi.org/10.3115/v1/D14-1179

[111] M. AZHARI, Y. KUMAR, O. GOH, and B. RAZA, "Text Summarization Evaluation Based on Sentence Scoring and Clustering," *Jestec.Taylors.Edu.My*, pp. 144–151, 2018.

[112] G. PadmaPriya and K. Duraiswamy, "An approach for text summarization using deep learning algorithm," *J. Comput. Sci.*, 2014.

[113] N. Ramanujam and M. Kaliappan, "An automatic multidocument text summarization approach based on naïve Bayesian classifier using timestamp strategy," *Sci. World J.*, 2016.

[114] Y. H. Hu, Y. L. Chen, and H. L. Chou, "Opinion mining from online hotel reviews – A text summarization approach," *Inf. Process. Manag.*, vol. 53, no. 2, pp. 436–449, 2017.

[115] C. Slamet, A. R. Atmadja, D. S. Maylawati, R. S. Lestari, W. Darmalaksana, and M. A. Ramdhani, "Automated Text Summarization for Indonesian Article Using Vector Space Model," in *IOP Conference Series: Materials Science and Engineering*, 2018.

[116] M. L. Khodra and Y. Wibisono, "Ekstraksi Event berbasis Paragraf dari Artikel Berita Bahasa Indonesia," *Semin. Tah. Lingusitik*, 2018.
https://doi.org/10.31227/osf.io/9wgs2

[117] Silvia, P. Rukmana, V. R. Aprilia, D. Suhartono, R. Wongso, and Meiliana, "Summarizing text for indonesian language by using latent dirichlet allocation and genetic algorithm," *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, vol. 1, no. August, pp. 148–153, 2014.

[118] J. W. G. Putra and M. L. Khodra, "Automatic Title Generation in Scientific Articles for Authorship Assistance: A Summarization Approach," *J. ICT Res. Appl.*, vol. 11, no. 3, p. 253, 2017.

[119] W. Hariardi, N. Latief, D. Febryanto, and D. Suhartono, "Automatic Summarization from Indonesian Hashtag on Twitter Using TF-IDF and Phrase Reinforcement Algorithm," *Int. Work. Comput. Sci. Eng. (WCSE 2016)*, no. Wcse, pp. 17–19, 2016.

[120] H. Christian, M. P. Agus, and D. Suhartono, "Single Document Automatic Text Summarization Using Term Frequency-Inverse Document Frequency (Tf-Idf)," *ComTech Comput. Math. Eng. Appl.*, vol. 7, no. 4, pp. 285–294, 2016.

[121] A. Najibullah, "Indonesian Text Summarization based on Naïve Bayes Method," *Proceeding Int. Semin. Conf. 2015 Golden Triangle*, pp. 67–78, 2015.

[122] R. C. Rajagukguk and M. L. Khodra, "Interpretable Semantic Textual Similarity for Indonesian Sentence," *2018 5th Int. Conf. Adv. Informatics Concept Theory Appl.*, pp. 147–152, 2018.

[123] C. Slamet, A. R. Atmadja, D. S. Maylawati, R. S. Lestari, W. Darmalaksana, and M. A. Ramdhani, "Automated Text Summarization for Indonesian Article Using Vector Space Model," in *IOP Conference Series: Materials Science and Engineering*, 2018.

[124] C. Y. Lin, "Rouge: A package for automatic evaluation of summaries," in *Proceedings of the workshop on text summarization branches out (WAS 2004)*, 2004.

[125] T. Oya, Y. Mehdad, G. Carenini, and R. Ng, "A Template-based Abstractive Meeting Summarization: Leveraging Summary and Source Text Relationships," 2015.
https://doi.org/10.3115/v1/W14-4407

[126] X. Wang, M. Nishino, T. Hirao, K. Sudoh, and M. Nagata, "Exploring Text Links for Coherent Multi-Document Summarization," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016.