Turkish Online Journal of Qualitative Inquiry (TOJQI)

Volume 12, Issue 3, July 2021:3123- 3142

Research Article

Research on Automatic text summarization techniques: a survey

G.Prathap[1],

Research Scholar, Department of Computer Applications, Madurai Kamaraj University, Madurai.

Dr.Rathina Sabapathy[2],

Associate Professor, Department of Computer Applications, Madurai Kamaraj University, Madurai.

Abstract

Automatic text summarization is a technique where it produces automatically a text summary with only the important points. Text plays a very important role in everyday life. Everywhere there is a communication through text, it may vary from small text to large text. When the sender sends the short text it is easy for the receiver to read it, incase if the sender side text is more it is difficult to read the summary. And there is no guarantee that everyone will read the content in that text. But how to transfer the important information to the receiver side is a technique that has evolved as text summarization. The large text is summarized into small with only important points. So many researches are happening in this novel endeavor because of the popularity of the communication through text. Here in our review paper we have shown some of the concepts of the text summarization, techniques and usage of this, and the impacts of this in today's world.

Keywords: Term Frequency-Inverse Document Frequency (TF IDF), Text Summarization, Machine Learning (ML), Natural Language Processing (NLP)

1. Introduction:

Text summarization is a method for compressing a lengthy chunk of text into a concise summary. It means that an only important aspect from the text has been out bounded to the reader. Several problems exist in Natural Language Processing (NLP) and Machine Learning (ML) on the basis of autonomous text summarization. [1,2,3] We all know that in today's world most of the communication are not happening face to face. So if they want to communicate most of the people are using text as communication medium. So now in the domain of machine learning there are so many researchers are involved and they started to work on summarizing text into important point which will give more benefit to the readers. So far it was done manually but now it has been implemented automatically and it is called automatic text summarization. Because of this new technique the time taken to read the content will be less. Because the main problem in the world is time, most people won't give that much time to read the content so it is better to summarize the text.

In text summarization there are three main steps are involved in order to summarize the text into a concise concept. It comprises of topic identification, interpretation and at last text summarization.[4,5,6,7] These are the main steps to be followed in order to summarize the text into concise gist.

2. Benefits of Text Summarization

The aim of text summarization is to extract the most significant information from a prescribed text in order to communicate with end-users [32] and the important contents which will be easy for the reader to read the content. The following content will describe further features of text summarization. According to statistics, significant terms are those that appear frequently. The word that occurs repeatedly will increase the phrase's score. Term Frequency-Inverse Document Frequency (TF IDF) [9,10]] is the most extensively used technique for calculating word frequency. Sentence frequency is the technique where it will calculate the number of sentences containing a particular word. The idea of this TF-IDF is , if there are more specific words appears in the document of the given sentence, it is assumed that those sentence are relatively more important. Location [19] is one of the key aspects to be considered in text summarization, it is based on the assumption that certain texts are located in specific position or location. Mostly the first and last line of the text will be added in the summary because of the assumption that most of the important information will be at the starting and ending of the text. The hypothesis behind the cue-based method is that the relevance of a phrase is determined in the summary either by the presence or absence of word lists found in a dictionary called a cue dictionary. [11] The weight of the sentence will indicate the importance of the effects of positivity or negativity of the word in the summary. Mostly the words present in the topic field or the headlines are most important where the sentences are related positively to the summary. The word present in the title indicates the main topic or core subject of the document. When we evaluate the summary we should give importance to the sentence length. Because [12] while drafting the summary very short sentence as well as very long sentence will not be suitable for formation of the summary from the document. Linguistic knowledge is utilized to determine the similarities between documents, sentences, document title, and sentences themselves. Some of the words such as proper noun shall be identified as important words for text summarization from the given text summary which will contain the core aspect of the text. For example person name, organization name, and place where the organization resides are important words which will be associated with crux of the textual information provided. When it pertains to forming relationships between entities, the distance between text units is a deciding element [54].

3. Different phases in Text Summarization

The text extraction procedure is separated into 2 phases: pre-processing and post-processing. The representation of the original text's structure is the preprocessing phase [16].

3.1Pre-Processing Phase:

- Sentence Boundary Identification:

Basically in text the user wants to know where it ends. It can be identified with the full stop in the text. So this boundary can be identified by finding where the dot is present in the text [8].

3.1.1Elimination of Stop word:

When texts are phrased there might be appearance of most common words which doesn’t have any special meaning or add any additional information. The most prevalent words in any natural language are stop words. These stop words might not provide significant value to the semantics of the document when examining textual information and constructing NLP algorithms [33]. These can be neglected externally without compromising the sentence's meaning. During the process of sentiment analysis or text classification, it is required to eliminate the stop words because they do not contribute any information, thereby removing undesired terms from our corpus. When it comes to stop words, there is no stimulating and active rule. Some of the recognized operations are Auto-Tag Generation, Caption Generation, Language Classification, Sentiment Analysis, Spam Filtering, or maybe something about Text Summarization, Machine Translation, Question-Answering problems, Text Classification, and

Language Modeling. For these kinds of operations stop words must be completely eradicated because they are an important part of these applications.

Stemming is the reduction of derivative terms to their root. The stem does not produce a similar word to the morphological root. Typically, acceptable words map to the same stem, even if that stem isn't a valid root in and of itself.

3.2 Post - Processing Phase:

With the help of the weight learning method [18] weights for features are assigned and using the weights, the relevance of the sentence is obtained in Post-processing phase. By using the future weight_the final score of each sentence is determined. Sentence which has a top score is selected for the final summary. Because in text summarization final summary evaluation is very important. This summarization can be processed using intrinsic and extrinsic actions [9].

4. Automatic Text summarization Methods

Text summarization is achieved through two different methods popularly known as extractive method and abstractive method. Each and every method has its own advantages and disadvantages. The diagram given below describes the tree structure of the methods which have been used in the text summarization to extract important points from the entire summary. Each and every approach has own concepts and methods. [23]

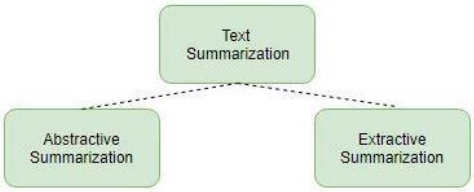

Fig: 1Automatic Text Summarization

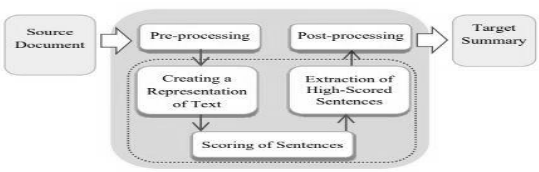

Fig: 2 Extractive Text Summarization

4.1 Extractive Text Summarization:

The framework of the extractive text summarization process is illustrated in the diagram above. Initially, the framework deals with the source document, it proceeds to provide the methodologies to arrive at the targeted summary, which is the final concise summary after all the processes. In this diagram, between the source and target we have two phases which are preprocessing phase and post processing phase. The preprocessing phase takes the input from the source document. Once the source file is input to the system it creates a representation of text and it is forwarded to sentence scoring. Sentence scoring is the link process which connects the preprocessing phase and the post processing phase. Once the original text representation passes to the scoring process from there the highest scored sentences are extracted with the post processing phase [3]. Once the scores are computed, the sentences with highest score from the summary are extracted, which are the targeted text. There are many research works have been proposed using many different techniques to extract the summary. As an example some of the summaries are generated fixing the threshold on the compression rate or length of the summary in order to limit the summary and some of the generated sentences of the input text are preserved.

4.1.1 Neural network based approach The relevant aspects of phrases that can be chosen in the article description are investigated using a neural network [45]. Thereafter, the neural network is customized to aggregate and generalize the key features revealed in summary sentences. Subsequently, the improved neural network is employed as a filter to generate news article descriptions.

4.1.2 Concept obtained approach Based on HowNet [46][47], this method is used to retrieve word conceptions. Instead of using a word, it emphasizes the concept as a feature. This technique generates a basic summary using a theoretical vector space model and then evaluates the level of semantic comparisons between words to reduce duplication. During the evaluation, Hownet is utilized to acquire a text concept and create a conceptual vector space model, after which it measures the concept's relevance using the conceptual vector space model. Eventually, it produces the conclusion by estimating phrase quality and removing summarization repetition.

4.2 Abstractive Text summarization

The figure given below depicts the architecture of the abstractive text summarization. It shows how the process of the summarization text will take place. When we compare this with the extractive summarization there is a minute difference.

In both extractive and abstractive summarization methods, there is preprocessing, intermediate representation of text and post processing phases. But there is a change in other phases. In extractive method, the summary is extracted from the existing text but in abstractive method it is not just extracting the contents. It has to be generated from the existing representation of the text.

This technique will not extract the content from the summary, but it will create the new sentences.[4] The first process in this is which will create the intermediate representation from the summary and with the help of the natural language processing the summary is generated.

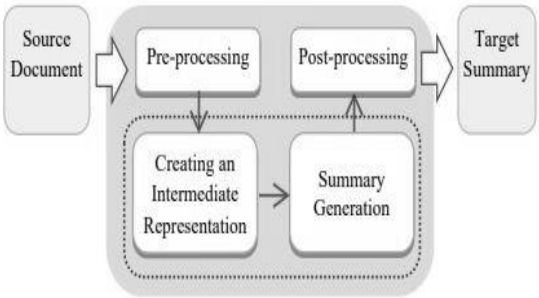

Fig: 3 Abstractive Text Summarization

Identified and combining the redundant information a concise abstractive summary is generated,

4.3 Hybrid Text Summarization:

Hybrid the name itself indicates that it is the combination of the other methods. . In the hybrid method the extractive approach and abstractive approach are combined to generate the summary. The below figure depicts the architecture of the hybrid text summarization [14] which contains the functionality of the extractive and abstractive as its name indicates.

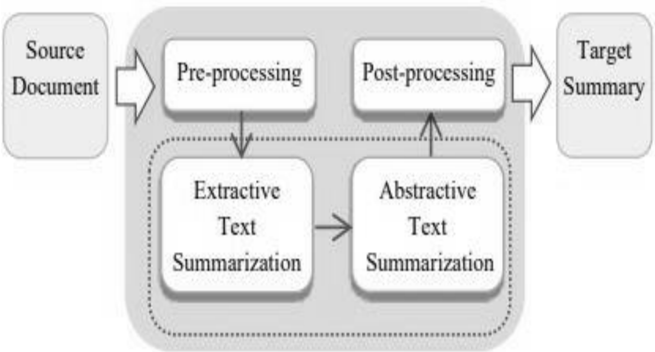

Fig:4 Hybrid Text Summarization

In abstractive summarization various techniques have been used. In order to achieve abstractive summarization in the tree-based approach [55], a theory-based discourse tree is synthesized into a dependency-based tree. In template based [56], first template is generated from either human authored summaries or extracting contents containing keywords[57] in the given document then the required information is updated into a template to generate the summary. The graph structure is utilized to describe the content in the graph-based technique [58]. Each node in the graph describes the word that exists in the text as well as its position and POS tag.

5. Literature Survey of Text Summarization

The concept of automatic text summarization is to condense the source text into the small versions by retaining its overall meaning and content information.

There are different techniques of text summarization used by different researchers. All these can be broadly classified into the above discussed techniques namely extraction based summarization technique, abstraction based summarization and hybrid summarization.

Pankaj Gupta et al. [48] have reviewed different techniques of Sentiment analysis and different techniques of text summarization. Sentiment Analysis (SA) is a method of machine learning where a system analyzes and interprets the sentiments and emotions in a text. Machine learning techniques such as the Naive Bayes Classifier and Support Machine Vectors (SVM) are employed. The emotions and sentiments in the text are determined using these methods. In this article, a survey on text summarization and sentiment analysis was undertaken to discover new research areas while keeping in view the benefits and drawbacks of present approaches and strategies.

Centroid-based Summarization [50]: Every cluster is made up of two to ten items from various sources that are arranged in chronological order are called topic detection. Employing the TF-IDF vector, an agglomerative clustering technique appends texts to clusters and recomputes the centroids. Because TFIDF scores are used to build clusters, centroids are referred to as pseudo-documents. After that, centroids are used to identify sentences from each cluster that characterize the issue.

Harsha Dave et al. [49] The author of this investigation proposes a method that uses the WordNet ontology to construct an abstractive summary from an extractive summary. The multiple documents with varied composition such as pdf files, text format files and word files had been used. The author has discussed various text summarization techniques and elaborates on step by step procedures of the multiple document text summarization approaches. The experiment's outcome is evaluated for various online extraction tools as well as with human generated summaries and shows the proposed system gives good results.

DharmendraHinhu et al. [52] provided an automatic summarization system using the extractive text summarization. In this research, the author considers Wikipedia Articles as input for summarising and acquiring text scoring. At the initial stage, the sentences are tokenized using regular expressions to match patterns. Later, the data were obtained in the form of a set of words by eradicating the stop words. The stemming of the words follows. The sentences are then rated using standard procedures. Scoring assists in the classification of sentences whether or not they should be included in the summary. It has been discovered that ranking sentences on the basis of citation yields superior results.

P.Krishnaveni, Dr.S. R. Balasundaram [36], the authors investigate that the automatic text summarization is a concept that compresses a large source information into smaller versions while retaining the overall sense and material information. It usually highlights the input document utilizing local scoring and ranking systems to generate a heading-by-heading summary. In this approach, similar attributes were applied to all document sentences and later intended to rank the sentences heading by heading to select the top n sentences from each heading. The compression ratio is used to get the value of n. The proposed heading-by-heading summary method keep the overall information intact without giving rise to any coherence gap which may be the disadvantage in other text summarization methods. Also they proved that their study approach maintains overall content in a concise form and the better understanding of the summary text. A survey of text summarization extractive techniques is described in this study.

J.N.Madhuri and Ganesh Kumar. R [37], the authors examined the text summarization based on a single document. The sentence separation approach facilitates the organization of the input in a more compact format so as the sentences are ordered by assigning weights. Highly ranked sentences are considered as the key sentences and they are extracted from the input document. Further, this process assits in obtaining the high-quality of the input document which in turn is saved in the audio format.

YashasviSwapnesh Kumar Parikh, NarsinganiAmishaDarshakbhai, HetalGaudani have employed Extractive Text Summarization Using Text Summary Generation Techniques [35]. The main objective of extractive summarization is to abstract text documents. The initial step in the text summary generating technique was to divide the text into sentences and sentences into words. All the content was formatted and converted into lowercase. All the punctuation marks, words having fewer than 3 characters, all the stop words, were removed. Afterwards convert the words third person to first person and past, future tense to present tense. This step was followed by the reduction of words to their root form. All the stop words were removed from the content. The subsequent action is to measure the frequency of words in the content. The final summary is generated by including the sentences that have the most frequent words in them.

Fuzzy Logic, Yan du and huahuo[38], in their study, explained the significance of automatic text summarization. A new automatic summarization model for news content is introduced based on multi-features, fuzzy logic principles, and the Genetic algorithm (GA). They assigned a score to each word and identified terms that scored higher than the threshold as keywords. News text would exhibit more information when it occurs with specialized features like time of the happening of the news, the persons involved, place where it has occurred. Such words are also identified and given prominence for finding feature words. Researchers added weights to sentences using a weighted genetic algorithm, and the linear composition of these attributes demonstrates the relevance of each sentence. Finally using fuzzy logic, the scores for sentences are found. The scores are used to extract the important sentences as text summary.

6. Statistical based Text Summarization Approach

6.1 Term Frequency:

Term frequency is a statistical tool which provide the information that how important words are distributed in the document.[9, 10] This strategy works by increasing the values of the TF-IDF based on the frequency of instances of the word in the document. Sentence frequency is the technique where it will calculate the number of words occurring a particular sentence. The idea of this TF-IDF is , if there are more specific words appears in the document of the given sentence, it is assumed that those sentence are relatively more important.

6.2 Topic Representation Approach

This topic representation approach may differ extremely. In this content we are going to represent the most common topic approach method. The method which is called as topic word [34] is really one of the recent work of the summarization of the text. A document may contain several topics. These topics will occur more often in the document. The measure of significance for words and sentences are found using frequency distribution. The sentences scoring highest significance are extracted to produce the abstract. The identification of topic by using the distribution of words is done using these statistical measures. Luhn [21] idea is most helpful in identifying the words which are descriptive of the given input. These types of descriptive words are traditionally called as the "topic signatures" in the literature summarization. Because of the topic signature the selection of important content of the multi document will be easier. These types of words which occur very often in the input but on the other text it will be rare. Actually it provides a way to set a threshold value in order to split the words which are in the input to see if they are descriptive. This type of prospect which is in the background corpus is categorized into the two way of the assumption like H1and H2. H1 is the input that probably the same in the background of B and H2 is the input which has the higher probability than the background B.

H1: p (w|I) =p (w|B) =p (where the w is not descriptive)

H2: p (w|I) = P_I and p (w|B) =p_B and p_I>p_B

H1 indicates that w is not a topic word. In this case, w appears in I and B with equal frequency. H2 indicates that w is a topic word. The word appears with higher probability in I than in B.

6.3 Frequency Driven Approaches [20]

When we consider the topic word approach there are two potentials which immediately come into the mind. As in topic word approaches the weight which represents the topic words are not required to decide whether words are better connected to the topic using binary weights (0 or 1) or real-value (continuous) weights.

In this approach the contents are presented in TF.IDF and the probability of the word. Comparatively the binary words will give much constant importance of the sentence than the probability of the word and the TF.IDF. Because of the conceptual simplicity the approaches are overviewed. In this approach, the lexical chains are utilized to identify the significance of the sentence. In difference of the other approaches the database of the lexical and the information which is derived from the wordnet are used. The information, which is derived from the wordnet and based on the frequency the prominence, can be tracked of the various topics. The term's probability is calculated by dividing the number of times it occurs by the total number of words in the document.

$p(w)= (c(w))/N$, $p(w)$ is the probability of the word, $c(w)$ is the frequency of occurrence of the word in the document.

There is one more representation using the regularity of the words happening in the document. This representation of including a probability of the word, expresses the importance of the word in the document, it does not represent the importance of the sentence. Suppose there is a sentence having more frequent words, then that sentence should be given more importance. This is achieved using, one more representation which is called as sum basic [22].The use of this system is to find the importance of the sentence.

$$\text{Weight}(s_j) = \frac{(\sum w_i \in s_j\, p(w_i))}{(|\{w_i \mid w_i \in s_j\}|)}$$

Here, s_j stands for sentences that have been altered. Each sentence should have its own paragraph. Therefore, s_j assigns a weight to the content words in the sentence based on the average probability p (wi) determined with the input for summarization.

The Greedy Method assists in choosing the best optimal at each stage, which identifies the term with the greatest likelihood of appearing. It is considered that the sentence is significant based on the selection criteria. Following the selection of the best sentence, the likelihood of every word in the sentence is changed. The looping is performed until the preferred result is obtained.

6.4 TF.IDF Method

Term frequency-inverse document frequency (TF-IDF) is a numeric measure to determine the relevant words in the document. The frequency of a sentence is defined as a couple of times a term appears in the source document's sentences. If the words appear frequently that means it is vital and should be given a high score than the word appearing less frequently [39].

For calculating TF – IDF formula is:

tf (w) = [Total appearance of a term w in document (D) / Total terms in D] ....(1)

idf (w) = loge [Total number of documents / Number of documents with term w in it.]

…. (2)

As a result, for a word w in the document, TF-IDF is calculated by using the following:

TF- IDF (w) = (1) * (2).

7. Machine Learning Approach in Text Summarization - Latent Semantic Analysis

This LSA is an unsupervised strong technique for the derivation of semantic of the text which is based on the co-occurrence of the words that are observed. Gong etol states that to have the identification of the most important content of the document in single-level and multi-level, the Latent semantic analysis can be utilized. The weight will be zero if the sentence has no words, or it will be equal to the TF*IDF word weight if the phrase contains words. Such a representation in which a row corresponds to words and sentences correspond to columns will form a matrix of order nxm. The matrix obtained is known as term sentence matrix. In order to represent the matrix as a product of the three matrices the singular value decomposition from the linear algebra is applied which is denoted as A=U∑VT, where ∑ is the diagonal matrix of the singular values. Several standard libraries include configured implementations of matrix decomposition. The main aim of the proposed method is to select most important sentences. From the representation the the value of matrx D can be obtained as $D = \sum V^T$. Then the weight of the sentences are found using the following formula,

$$\text{Weight } (s_i) = \sqrt{\sum_{j=1}^{m} d_{ij}^2}$$

8.RelatedWorkSummary

Reference	Techniques	Evaluation	Advantages	Disadvantages
Bayesian Learning in text summarization _ Proceedings of human language technology conference	Methodology used in this summarization is the Bayes rule of machine learning approach	The problems are classified into models in order to summarize the task of the particular text.	Selecting sentence form the large set of data for the summary are improved with this technique.	In order to the automatic summarization, human interpretation are mostly required which is little bit time consuming.
Automatic text summarization with neural network - IEEE	In this methodology they have used the artificial neural network	The idea behind this method is to train a neural network where the phrases can be classed as summaries or non-summaries, allowing the neural network to filter and classify the sentences.	According to the human reader style the network has been trained. the users need and requirements can be achieved easily	In order to train the data it requires human interruption. In training phase and application phase the neural network is slow.
CRF based future extraction applied for supervised automatic text summarization - Elsevier	The conditional random field (CRF) has been used in this methods.	CRF can be used as a sequence problem by the approach of the statistical model.	Better representation of the sentence can be provided by identifying the correct features and sentences are grouped appropriately.	For training this method needs an external domain specific and limitation of the language is not there.

Graph based framework for automatic text summarization - Elsevier	Here they have used the graph-based technique for summarization	The relationship of the sentence can be framed or identify by constructing the graph.	Coherency is improved and the repeated information in the sentences is captured easily.	This technique will not focus much o the problems.
An approach to concept obtained text summarization – IEEE	In this the methodology used is concept oriented approach.	In a sentence the important concepts are captured by using the knowledge from the Wikipedia etc.	There is no cooperation of the similarity in order to reduce the repeated information from the sentence.	This technique also will not identify some of the problems.
Automatic text summarization using customizable fuzzy features and attention on the context and vocabulary – IEEE	The approach used in this methodology is the fuzzy logic based approach.	By using the various sets of fuzzy rules the summarization is happened.	The quality of service of the text summarization has been improved by maintaining the coherency.	Some of the membership function of the fuzzy logic is not supported.

9. Conclusion:

Text summarization is one of the rapid growing technologies. Text summarization is to summarize the text into a precise summary which is very much useful for the reader to fetch only the important points from the total amount of the content. Even though this automatic text summarization is a very old challenge but the current scenario of research directs this trend to various sectors like education, blogs, biomedicine and product review. Because in these sectors information are overloaded, automated summarization using natural language processing is very much an imperative aspect in these areas. There are still so many researches are happening in this field for the text summarization.

This paper introduced the research over text summarization, elaborated the various techniques and methods, various features to achieve the text summarization and benefits of text summarization in today's technological advancement.

And it also elaborated the three methods to extract important points from original text to arrive at targeted one with advantages and disadvantages of each and every method.[15] All the approaches are also explained in a precise way which will help the researchers to study and understand more about the concept and make the disadvantages in the approach into an advantageous one. In both commercial as well as research community the text summarization has its importance. At the end of the analysis it is clearly shown that comparatively more learning and reasoning is needed for the abstractive approach, and more meaningful information can be provided by the abstractive approach. Lot of research has to be undertaken in the methods to improve the summarization process and make it more effective.

References:

[1] Saranyamol C S, Sindhu L, "A Survey on Automatic Text Summarization", International Journal of Computer Science and Information Technologies, 2014,Vol. 5 Issue 6.

[2] Reeve Lawrence H., Han Hyoil, NagoriSaya V., Yang Jonathan C., Schwimmer Tamara A., Brooks Ari D., "Concept Frequency Distribution in Biomedical Text Summarization", ACM 15th Conference on Information and Knowledge Management (CIKM), Arlington, VA, USA,2006.

[3] Khan Atif, SalimNaomie, "A review on abstractive summarization Methods", Journal of Theoretical and Applied Information Technology, 2014, Vol. 59 No. 1.

[4]SuneethaManne,ZaheerParvezShaikMohd. , Dr. S. Sameen Fatima, "Extraction Based Automatic Text Summarization System with HMM Tagger", Proceedings of the International Conference on Information Systems Design and Intelligent Applications, 2012, Vol. 132, P.P 421-428.

[5] Gupta Vishal, "A Survey of Text Summarizers for Indian Languages and Comparison of their Performance", Journal of Emerging Technologies In Web Intelligence, 2013, Vol. 5, No. 4.

[6] Vishal Gupta,"A Survey of Recent Keywords and Topic Extraction Systems for Indian Languages", International Journal of Engineering Trends and Technology (IJETT) , 2013, Vol. 6 No. 6

[7] Gupta V. And Lehal G. S., "A Survey of Text Summarization Extractive Techniques", International Journal of Emerging Technologies in Web Intelligence, 2010, Vol. 2. , pp. 258-268.

[8] ViswanathMeghana, "Thesis: Ontology-Based Automatic Text Summarization", M. Sc Thesis, Vishweshwaraiah Institute of Technology, India, 2009.

[9]Reeve Lawrence H., Han Hyoil, NagoriSaya V., Yang Jonathan C., Schwimmer Tamara A., Brooks Ari D., "Concept Frequency Distribution in Biomedical Text Summarization", ACM 15th Conference on Information and Knowledge Management (CIKM), Arlington, VA, USA,2006.

[10].Khan Atif, SalimNaomie,"A review on abstractive summarization Methods", Journal of Theoretical and Applied Information Technology,2014, Vol. 59 No. 1.

[11] Ibrahim Imam, NihalNounou, AlaaHamouda, Hebat Allah Abdul Khalek, "Query Based Arabic Text Summarization", IJCST, 2013, Vol. 4, Issue Spl - 2.

[12] KhosrowKaikhan, "Text Summarization Using Neural Networks", Proceedings. 2004 Second IEEE International Conference on Intelligent Systems, 2004, Vol. 1.

[13] Dixit Rucha S., Apte S. S., "Improvement Of Text Summarization Using Fuzzy Logic Based Method", IOSR Journal Of Computer Engineering (IOSRJCE) ISSN: 2278-0661, ISBN: 2278-8727, 2012, Vol. 5, Issue 6, PP 05-10.

[14] Fachrurrozi M., Yusliani Novi, and YoanitaRizkyUtami, "Frequent Term based Text Summarization for Bahasa Indonesia", International Conference on Innovations in Engineering and Technology Bangkok (Thailand), 2013.

[15] Babar S.A. and Thorat S.A., "Improving Text Summarization using Fuzzy Logic & Latent Semantic Analysis", International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2014, Vol. 1 Issue 4.

[16] Marcus V. C. Guelpeli Ana Cristina B. Garcia ,AntónioHortaBranco,The process of summarization in the pre-processing stage in order to improve measurement oftextswhenclustering,

https://www.researchgate.net/publication/236672418_The_process_of_summarization_in_the_pre-processing_stage_in_order_to_improve_measurement_of_texts_when_clustering

[17] Jabbar, A., Iqbal, S., Tamimy, M.I. *et al.* Empirical evaluation and study of text stemming algorithms. *ArtifIntell Rev* **53,** 5559–5588 (2020).https://doi.org/10.1007/s10462-020-09828-3

[18] Meng, Z., Zhang, Z., Li, G. *et al.* An active weight learning method for efficient reliability assessment with small failure probability. *Struct Multidisc Optim* **61,** 1157–1170 (2020).https://doi.org/10.1007/s00158-019-02419-z

[19] Deepali K. Gaikwad1 and C. NamrataMahender2, A Review Paper on Text Summarization, International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 3, March 2016

[20] Matthew Mayo, KDnuggets**,** Approaches to Text Summarization: An Overview,https://www.kdnuggets.com/2019/01/approaches-text-summarization-overview.html

[21 Hans Peter Luhn. 1958. The automatic creation of literature abstracts. IBM Journal of research and development 2, 2 (1958), 159–165.

[22] https://www.statmt.org/book/slides/07-language-models.pdf
http://ivan-titov.org/teaching/nlmi-15/lecture-3.pdf

[23] https://www.queppelin.com/how-nlp-is-helping-in-automatic-text-summarization-2/

[24] H. Sak, A. Senior, F. Beaufays, Long short-term memory recurrent neural network architectures for large scale acoustic modeling, in: Fifteenth annual conference of the international speech communication association.

[25]C. Olah, Understanding lstm networks, GITHUB blog, posted on August 27 (2015) 2015
https://colah.github.io/posts/2015-08-Understanding-LSTMs/

[26] SeppHochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. Neural computation 9, 8 (1997), 1735–1780.

[27] DzmitryBahdanau, Kyunghyun Cho, and YoshuaBengio. 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014).

[28] IlyaSutskever, OriolVinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In Advances in neural information processing systems. 3104–3112

[29] YoshuaBengio, Patrice Simard, and Paolo Frasconi. 1994. Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks 5, 2 (1994), 157–166.

[30] RazvanPascanu, Tomas Mikolov, and YoshuaBengio. 2013. On the difficulty of training recurrent neural networks. In International Conference on Machine Learning. 1310–1318

[31] Smagulova, Kamilya ,James, Alex - 2019/05/23,A survey on LSTMmemristive neural network architectures and applications,The European Physical Journal Special Topics

[32]D. Hingu, D. Shah and S. S. Udmale, "Automatic text summarization of Wikipedia articles," *2015 International Conference on Communication, Information & Computing Technology (ICCICT)*, Mumbai, India, 2015, pp. 1-4, doi: 10.1109/ICCICT.2015.7045732.

[33]SHUBHAM SINGH, AUGUST 21, 2019, NLP Essentials: Removing Stopwords and Performing Text Normalization using NLTK and spaCy in Python,https://www.analyticsvidhya.com/blog/2019/08/how-to-remove-stopwords-text-normalization-nltk-spacy-gensim-python/

[34]Leonhard Hennig DAI Labor, TU Berlin Berlin, Germany, Topic-based Multi-Document Summarization with Probabilistic Latent Semantic Analysis.International Conference RANLP 2009 - Borovets, Bulgaria, pages 144–149.

[35] Text Summary Generation Techniques YashasviSwapnesh Kumar Parikh, NarsinganiAmishaDarshakbhai, HetalGaudani, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-7S, May 2020

[36] Automatic Text Summarization by Local Scoring and Ranking for Improving Coherence, P.Krishnaveni, Dr.S. R. Balasundaram.

[37] Extractive Text Summarization Using Sentence Ranking, J.N.Madhuri, Ganesh Kumar.R,

[38] News Text Summarization Based on Multi-Feature and Fuzzy Logic, Yan du and huahuo

[39] Jaiswal, A.& Bhatia, N., (2016, January). Automatic text summarization and it's methods-a review. In 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence) (pp. 65-72). IEEE.

[40]Saranyamol, C. S., &Sindhu, L. (2014).A survey on automatic text summarization.International Journal of Computer Science and Information Technologies, 5(6), 7889-7893.

[42]Mahender, C. N., &Gaikwad, D. K., (2016).A review paper on text summarization. International Journal of Advanced Research in Computer and Communication Engineering, 5(3), 154-160.

[43]Takamura, H., Okumura, M., Nagata, M. ,Hirao, T., & Kikuchi, Y., (2014, June). Single document summarization based on nested tree structure. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (pp. 315-320).

[44]Khan, A.,&Salim, N. (2014). A review on abstractive summarization methods. Journal of Theoretical and Applied Information Technology, 59(1), 64-72.

[45]Kaikhah, K. (2004). Text summarization using neural networks.

[46] Wang, M., Wang, X., &Xu, C. (2005, October). An approach to concept-obtained text summarization.In IEEE International Symposium on Communications and Information Technology, 2005.ISCIT 2005.(Vol. 2, pp. 1337-1340).IEEE.

[47] Zamanifar, A., Minaei-Bidgoli, B., &Sharifi, M. (2008, August). A new hybrid farsi text summarization technique based on term co-occurrence and conceptual property of the text. In Software Engineering,

Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2008. SNPD'08. Ninth ACIS International Conference on (pp. 635-639). IEEE.

[48] Pankaj Gupta, RituTiwari and Nirmal Robert, “Sentiment Analysis and Text Summarization of Online Reviews: A Survey.” International Conference on Communication and Signal Processing, 2016.

[49]Harsha Dave, Shree Jaswal, “Multiple Text Document Summarization System using Hybrid Summarization Technique.” 1st International Conference on Next Generation Computing Technology (NGCT), 2015

[50] Radev, D. R., Jing, H., Stys, M., and Tam, D. (2004). Centroid-based summarization of multiple documents.Information Processing and Management 40 (2004), 40:919-938.

[51]Akshi Kumar, Aditi Sharma, Sidhant Sharma, ShashwatKashyap, “Performance Analysis of Keyword Extraction Algorithms Assessing Extractive Text Summarization.” International Conference on Computer, Communication, and Electronics (Comptelix), 2017.

[52]DharmendraHingu, Deep Shah, SandeepS.Udmale, “Automatic Text Summarization of Wikipedia Articles.” International Conference on Communication, Information & Computing Technology (ICCICT), 2015.

[53]. Taeho Jo, “K Nearest Neighbor for Text Summarization using Feature Similarity.” International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE), 2017.

[54] Reeve Lawrence H., Han Hyoil, NagoriSaya V., Yang Jonathan C., Schwimmer Tamara A., Brooks Ari D., “Concept Frequency Distribution in Biomedical Text Summarization”, ACM 15th Conference on Information and Knowledge Management (CIKM), Arlington, VA, USA,2006.

[55] IEEE/ACM Transactions on audio, speech, language processing, vol. Xx, no. X, December 2014 , Summarizing a Document by Trimming the Discourse Tree Tsutomu Hirao*, Masaaki Nishino, Yasuhisa Yoshida, Jun Suzuki, Norihito Yasuda, and Masaaki Nagata.

[56] Proceedings of the 8th International Natural Language Generation Conference, pages 45–53, Philadelphia, Pennsylvania, 19-21 June 2014. c 2014 Association for Computational Linguistics, A Template-based Abstractive Meeting Summarization: Leveraging Summary and Source Text Relationships Tatsuro Oya, Yashar Mehdad, Giuseppe Carenini, Raymond Ng Department of Computer Science University of British Columbia, Vancouver, Canada.

[57] Procedia Computer Science 87(2016)25-31,Fourth International Conference On Computer Science & Engineering,A Study On Abstractive Text Summarization Techniques in Indian Languages,SunithaC[a*],Dr.A.Jaya[b*],AmalGanesg[a.]

[58] Procedia Computer Science 89 (2016) 404 – 411,Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016),ATSSI: Abstractive Text Summarization using Sentiment Infusion,Rupal Bhargava∗, Yashvardhan Sharma and Gargi Sharma.Birla Institute of Technology & Science, Pilani, Pilani Campus 333 031, India

.