RESEARCH ARTICLE OPEN ACCESS

A Brief Introduction To Malayalam Text Summarization

Anju Gopan [1], Haripriya K.S [2], Jeethu Joshy [3], Sheena Kurian K [4]

[1], [2], [3] UG Scholar, Dept. Of Computer Science and Engineering, KMEA Engineering College, Kerala - India
[4] Assoc. Prof., Dept. Of Computer Science and Engineering, KMEA Engineering College, Kerala - India

ABSTRACT

As the textual data is increasing rapidly, it becomes more complex to identify the relevant information from that. Text summarization is the process of shortening long text into a short text by identifying the relevant information from the given document and discarding the redundant information. On the basis of this, we propose a text summarization which generates a summary by rewriting the contents in the given textual document by identifying the similarity between the words. Here we are introducing a model which generates the summary by extracting relevant sentences.

Keywords: Natural Language Processing, Text Summarization, Abstractive Text Summarization, Extractive Text Summarization, Lemmatization, Pos Tagging.

I. INTRODUCTION

Nowadays, large number of text documents are being published online. Numerous documents in Indian languages like Hindi, Malayalam, Tamil etc are also available online. But reading all these documents to understand a concept and identify relevant information is time consuming and tedious. So it would be of great help if we get a summary of the published material. Generating a summary of documents by a human is also nearly impractical. There comes the relevance of automatic text summarization. Due to this reason text summarization has been widely discussed and explored in recent times. The relevance of text summarization is also increased due to the abundance of documents on the internet. Text summarization is a technique that can be used to generate a summary from a given document.

There are mainly two methods for text summarization: extractive text summarization and abstractive text summarization. Extractive method summarizes a given document by identifying the important sentences from the original text. This method does not identify the similarity and word meanings between the words. It mainly points out the relevant sentences from the given sentences that are meant to produce a condensed summary. It will not generate a fluent summary as the summary generated by humans.

In case of abstractive text summarization it summarizes the text by identifying the relationship between words and substituting the words having similar meanings with the other. It generates the summary as a human generated summary. On the basis of this, we propose a model for text summarization in malayalam. Here we are introducing a model which generates the summary by extracting relevant sentences. When it comes to Indian language like malayalam, it is a bit complex to generate summary due to rich morphological words. Vocabulary of Malayalam is borrowed from other languages like Sanskrit and Tamil. The word meanings used in Malayalam may vary in different situations.

TABLE 1. Truncated sentences.

S1	"എന്തിനു പറഞ്ഞു?"
S2	"എന്തിനു പറഞ്ഞു എന്ന് ആലോചിച്ചു"

For example, in Table.1, S1 is the sentence which is depicted as a question, "why did he said so?" and S2 means "wondered why he said that".

Even though S1 is a truncated version of sentence S2, S1 also depicts a sentence that "wondered why he said that". In both the cases the situations that used for,
"എന്തിനു പറഞ്ഞു" is different.

Challenges in malayalam summarization:

- Ambiguity- Different words may have same meanings which makes it difficult to understand the exact meaning used in that passage.
  Eg: സംശയം , സന്ദേഹം

Here, both the words 'സംശയം' and 'സന്ദേഹം' have the same meaning as"doubt."

- Difficult to find resources: It is difficult to find good resources for training, because there exist different sentences which contribute to the same meaning to the text.
- Availability of tool: There is a lack of availability of appropriate tools for pre-processing of malayalam dataset.
- Difficulty in finding root word: It is quite difficult to map words with appropriate root word, because several words have the same meaning and some of them can be represented in different words with the same meaning. In order to overcome all this, we need to design an architecture for summary generation which generates accurate results. Table 1. shows the example for the same word with different meaning.

TABLE 2. Same word with different meaning

S1 S1'	"പുലി ആടിനെ ഓടിച്ചു" Tiger hunted the goat
S2 S2'	"രാമു ദാമുവിൻറെ സൈക്കിൾ ഓടിച്ചു " Ramu rode Damu's bicycle

Table.2. shows the same word with different meanings. It does not have predicate agreement between subject or object and verb for gender, number and person. Hence summarization is not very easy in this language.

II. RELATED WORKS

For approaching Malayalam text summarization in NLP, the first step is to eliminate punctuations, hyphens and redundant white spaces etc. This is called pre-processing of the input text. A pre-processing phase, the linguistic techniques are used to pre-process input text documents using some important techniques such as sentence segmentation, punctuation marks removal, filtering stop-words, stemming (reducing common root words), etc[1]. After that the next phase is tokenizing the data, which reduces the complexity of sentence structure. Tokenization is the process of splitting up a sentence into small pieces known as tokens or words. Larger sentences can be divided into pieces of words by performing tokenization. By performing tokenization, instead of processing any task in an entire sentence at a time, the task can be done in pieces of a sentence[2].

Based on the LDA model and word2vec weight matrix, the similarity analysis of natural language processing text is designed[3]. Word2vec is an unsupervised learning mechanism which uses a large number of data samples. This is implemented by using two mechanisms called CBOW (continuous bag of words) and skip gram. CBOW aims to predict the probability of current words according to the context. Skip gram is the opposite, the probability of predicting the context based on the current word. Both methods use artificial neural networks as their classification algorithm[4]. LDA is a generative model that learns the representation of documents from frequencies of each word used in that document using the Bag-of-Word representation where bag-of-words is a representation of text that describes the occurrence of words within a document[3].

Stop words are frequently occurring words in natural language which are considered as unimportant in natural language generation tasks such as text summarization etc. A dictionary based approach is being utilized to remove stopwords from the input document. A generic stop-word list containing approximately 75 Malayalam stopwords are created manually and that stopword is compared with the tokens of the input document using sequential search technique. If it matches, the word is removed completely and the comparison process continues till all the stop words are compared. The input tokens devoid of stopwords are thus left out in the array and are stored back into the corpora[5].

Text summarization technique is based on a ranking algorithm which gives the collection of sentences having the highest ranks from the input text. The method extraction of sentences, which gives the idea of the input text in a short form, is presented. Sentences are ranked by assigning weights and they are ranked based on their weights. Highly ranked sentences are extracted from the input document so it extracts important sentences which directs to a high-quality summary of the input document and stores the summary as audio[6].

Sequence to sequence model is a special type of recurrent neural network which is used for natural language generation tasks like text summarization, translation etc. This discovers various inherent semantic and compositional aspects as part of document-independent features. The summary of the document is built using the abstract sentences generated from a model generated using a neural network. Sequence to

sequence learning technique is used to build the model from the raw data. The required dataset was built by translating the publicly available news repository and its summary from BBC into Malayalam language[7].

In contrast to the stemming method, lemmatization which uses a method for extracting the root word from a Malayalam word. One is clustered indexed dictionary based, which uses a suffix replacement method by considering nearly thousand rules which are identified by several test cases done during the development of the project[8].

III. PROPOSED WORK

Text summarization can be seen as the process of pointing out the relevant sentences from a passage. It can be achieved through a series of steps. The first step towards the work starts from data preprocessing. Data preprocessing is the process by which the data is transformed into the form actually needed. The input text may contain unnecessary. Here we are removing commas, hyphens, double quotes/single quotes and punctuations from our dataset.

In the preprocessing stage the data is refined. It includes extracting words from sentences, making word count, finding the stem word, ranking sentences, deleting unnecessary sentences based on sentence ranking etc. Data preprocessing are achieved through several steps.

In the tokenization process the words are extracted from the sentence. It breaks the text into small chunks. Tokenization is considered as the primary step towards text processing. Tokens can be words, special characters, subwords etc. Usually tokenization is operated using the delimiters in the sentences. In the case of Malayalam, most of the words are splitted using white space between them.

TABLE 3. Tokenization

S3	“അയാൾ പറഞ്ഞു കൊട്ടത്തു.”
S4	“അയാൾപറഞ്ഞുകൊട്ടത്തു.”

In Table 3, it will be easy to split the sentence S4 since there are enough white space between them. When tokenized, we will get it as ‘അയാൾ’, ‘പറഞ്ഞു’, ‘കൊട്ടത്തു’. Thus we get three tokens.

But in the case of sentence S4 in Table 3, there is no white space to separate them. Thus we will get a single token as

“അയാൾപറഞ്ഞുകൊട്ടത്തു”

even though it is a combination of three separate words.

In the stop word removal the words having no meanings are removed.

TABLE 4. Example Sentence

S5	“അദ്ദേഹം നല്ല ഭംഗിയുള്ള തൊപ്പികളുണ്ടാക്കി അത് ചന്തയിൽ കൊണ്ടുപോയി വിൽക്കുമായിരുന്നു”

TABLE 5. Stop words of S5

“നല്ല”, “ഭംഗിയുള്ള”, “അത്”

In Table 5, the stopwords of sentence S5 in Table.4 is identified.

The sentence without these words will also convey the same meaning. It is shown in Table.6.

TABLE 6. S5 after stopword removal

“അദ്ദേഹം തൊപ്പികളുണ്ടാക്കി ചന്തയിൽ കൊണ്ടുപോയി വിൽക്കുമായിരുന്നു”

Then word frequencies of each of the sentences are calculated.

Cosine distances are used to identify the similarity between the words using the word embedding replacement model. Thus by using word embeddings, words that are close in meaning are grouped near to one another in vector space.

For example, while representing a word such as മാതാവ്, the nearest neighbour of മാതാവ് would be “മാതാശ്രീ”, “അമ്മ”. After ranking the sentence, the sentence with high ranking is collected for generating the summary.

Ranking is based on the length of the sentence. Stemming is the process of finding the root word. During stemming the

correct root word will be identified. It will help to find out the prefix, suffix and others from the sentences.

Finding the correct root word is necessary for replacing the word with simple synonyms. Malayalam being ambiguous languages, many words seem to be similar, but conveying different meaning.

TABLE 7. Root word

“പുറപ്പെടുന്നു” “പുറപ്പെട്ടുകഴിഞ്ഞു” “പുറപ്പെടുകയാണ് “പുറപ്പെട്ടിട്ടുണ്ട് “	“പുറപ്പെടുക”

Table 7. shows the different word suffix that give meaning to the root word “പുറപ്പെടുക”. This word is used to indicate the departure, but when suffix are used, it also indicates the timing

TABLE 8. Root word with different suffix

പുറപ്പെടുക + ന്നു	“പുറപ്പെടുന്നു”
പുറപ്പെടുക +കഴിഞ്ഞു	“പുറപ്പെട്ടുകഴിഞ്ഞു”
പുറപ്പെടുക +ആണ്	“പുറപ്പെടുകയാണ്
പുറപ്പെടുക +ഉണ്ട്	“പുറപ്പെട്ടിട്ടുണ്ട് “

Table 8. shows the different words formed by adding different suffixes to the same root word.

Lemmatization is the task of finding the base word form from a given word form. It is closely related to stemming. When considering malayalam language, which is morphologically rich, it contains a lot of words which have similar meaning. Malayalam contains words which have multiple meaning considered to the situation, in order to find the correct dictionary meaning used is a task. It needs to evaluate the sentence correctly to find the exact dictionary word.

"രാമു ബാലന് പണി കൊടുത്തു "

This implies two meanings, one it can be interpreted as Ramu is employed because of Balan, second meaning is that Balan has done something which made Ramu in trouble. Thus without proper understanding of the sentence meaning, we cannot generate the desired output.This implies two meanings, one it can be interpreted as Ramu is employed because of Balan, second meaning is that Balan has done something which made Ramu in trouble. Thus without proper understanding of the sentence meaning, we cannot generate the desired output.

TABLE 9. Lemmatization on a word example

Input	Lemma	Rules
പാട്ടം	പാട്ടുക	ട്ടം -> ട്ടുക
നേടി	നേടുക	ടി ->ടുക
പോരും	പോരുക	രും ->രുക

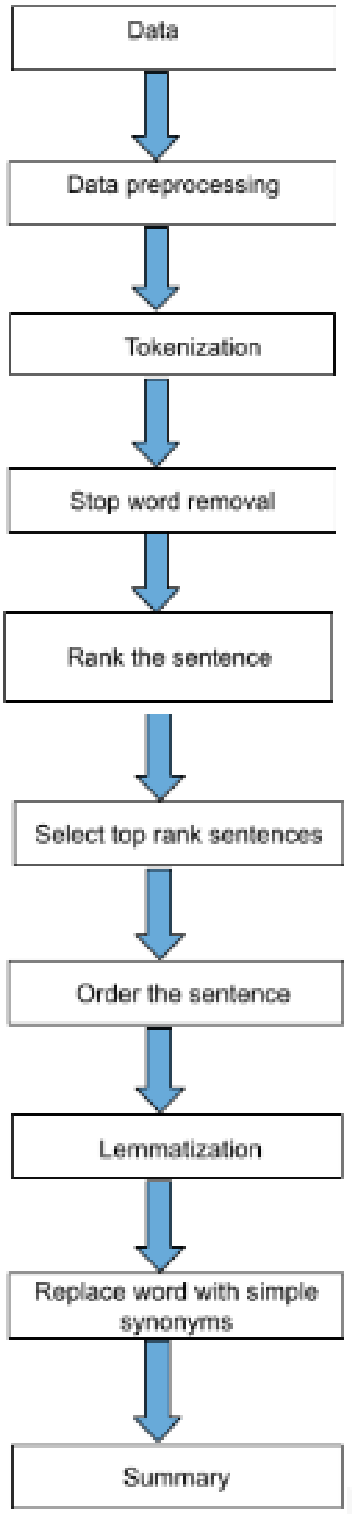
Table 9. depicts the lemma rule associated with different words ending with the same alphabet in malayalam.

POS tagging is the part-of-speech tagging it is also called grammatical tagging. Here the words in the sentences are classified according to the corresponding grammar categories- subject, object and verb.

For example, consider the malayalam sentence,

TABLE 10. Example

“അച്ഛൻ കുട്ടിയെ എടുക്കുന്നു”	The father picks up the child



After applying pos tagging on the sentence given in table 10, the result is the subject, object and verb in it as shown in table 11.

TABLE 11. Pos tagging

അച്ഛൻ	കർത്താവ്	Subject
കുട്ടിയെ	കർമം	Object
എടുക്കുന്നു	ക്രിയ	Verb

Table 11. shows the malayalam semantic rule for sentence, "അച്ഛൻ കുട്ടിയെ എടുക്കുന്നു ".
But it is not mandatory that all the sentences in Malayalam needs to follow this order.
"കുട്ടിയെ അച്ഛൻ എടുക്കുന്നു" have the same meaning as the sentence in Table 11. We need to understand the sentence first in order to apply pos tagging.

IV. RESULT AND OBSERVATIONS

This article presents a simple approach to generate a summary from the given passage. Currently we have done it for small passages. One of the method used to evaluate the quality of machine generated summary is to compare it with human summary. Hence we created a summary for the stories as the validation data and are able to generate a summary with the sentences given. The input data given is a short story extracted from the google documents. First the input is given as a text file which is the original text and the human summary. The vector value assigned to each sentence gives them a rank according to the length of the sentences. The sentences with the higher ranks are considered in the summary. The summary generated doesn't contain the entire word. It may broke when ranking is applied to it. It generated a summary of almost half of the original input text document.

TABLE 12. Input

പണ്ടു പണ്ടൊരിടത്ത് ഒരു തൊപ്പി വില്പ്നക്കാപ്പരൻ ഉണ്ടായിരുന്നു. അദ്ദേഹം നല്ല ഭംഗിയുള്ള തൊപ്പികളുണ്ടാക്കി അത് ചന്തയിൽ കൊണ്ടുപോയി വിൽക്കുമായിരുന്നു. തൊപ്പി വിറ്റുകിട്ടുന്ന കാശു കൊണ്ടായിരുന്നു അയാൾ ജീവിച്ചത്. പല പലസ്ഥലങ്ങളിൽ പോയായിരുന്നു അയാൾ തൻ്റെ തൊപ്പികൾ വിറ്റിരുന്നത്. ഒരു ദിവസം തൊപ്പി വിൽക്കാനായി പോയ കച്ചവടക്കാരൻ ഒരു കാട്ടിലെത്തി. ആ കാട് കടന്നു വേണമായിരുന്നു ചന്തയിലെത്താൻ. അന്ന് പതിവിലും കൂടുതൽ ചൂടുള്ള ദിവസമായിരുന്നു. തൊപ്പി സഞ്ചിയുമായി കാട്ടിലൂടെ നടക്കുന്നതിനിടയിൽ അയാൾക്ക് നല്ല ക്ഷീണം അനുഭവപ്പെട്ടു . ഇന്നു നല്ല ചൂട് ഉണ്ടല്ലോ. മാത്രമല്ല നടന്നും ക്ഷീണിച്ചു. എന്തായാലും ഒരു മരച്ചുവട്ടിൽ ഇരുന്നു വിശ്രമിക്കാം. എന്നും പറഞ്ഞു അയാൾ തൻ്റെ തൊപ്പികൾ സൂക്ഷിച്ചിരുന്ന സഞ്ചിയും താഴെ വച്ച് ഒരു മരച്ചുവട്ടിൽ ഇരുന്നു. എന്നാൽ ക്ഷീണം കാരണം അയാൾ അവിടെയിരുന്നു ഉറങ്ങിപ്പോയി. ഉറങ്ങി എണീറ്റ തൊപ്പി വിൽപ്പനക്കാരൻ ഞെട്ടിപ്പോയി. തൻ്റെ സഞ്ചിയിൽ വിൽക്കാൻ വച്ചിരുന്ന തൊപ്പികൾ ഒന്നും കാണാനില്ല. അദ്ദേഹം ചുറ്റും നോക്കി. അപ്പോഴാണ്മരത്തിൻ്റെ മുകളിൽ അയാൾ ആ കാഴ്ച കണ്ടത്. ആ മരത്തിൽ നിറയെ കുരങ്ങന്മാർ. അവരുടെ തലയിൽ അയാൾ വിൽക്കാനായി ഉണ്ടാക്കിയ മനോഹരമായ തൊപ്പികളും. തൊപ്പി വില്പ്പനക്കാരൻ വിഷമിച്ചു. ഇനി എന്ത്ചെയ്യും ഈ കുരങ്ങൻമാരുടെ കയ്യിൽ നിന്ന് എങ്ങനെയാണ് തൊപ്പികൾ തിരിച്ചു വാങ്ങുന്നത് ആദ്യം അയാൾ ആ കുരങ്ങന്മാരെ വിരട്ടി നോക്കി. അപ്പോഴല്ലേ രസം അവർ തിരിച്ച് അയാളെ വിരട്ടി. അടുത്തയാൾ ഒരു കമ്പെടുത്ത് അവരെ പേടിപ്പിക്കാൻ നോക്കി. അപ്പോൾ കുരങ്ങന്മാരും തിരിച്ചു കമ്പെടുത്തു ആ തൊപ്പി വില്പനല്പക്കാരനെ ആക്രമിക്കാൻ ഒരുങ്ങി. ഇത്തരത്തിൽ പലതരത്തിലും അയാൾ ആ കുരങ്ങന്മാരിൽ നിന്നും തൊപ്പി വാങ്ങാനായി ശ്രമിച്ചു. എന്നാൽ എല്ലാ ശ്രമവും പരാജയപ്പെട്ടു . അയാൾ തലപുകച്ച് ആലോചിച്ചു. എങ്ങനെയും കുരങ്ങൻമാരിൽ നിന്നും തൊപ്പികൾ വാങ്ങിയേ തീരൂ. അപ്പോഴാണ് അയാൾക്കൊരു കാര്യം മനസ്സിലായത്. ഞാൻ എന്തു ചെയ്താലും കുരങ്ങന്മാരും അതുപോലെ ചെയ്യുന്നു. പെട്ടെന്ന് അയാൾക്ക് ഒരു ബുദ്ധി തോന്നി. ആ തൊപ്പി കച്ചവടക്കാരൻ തൻ്റെ തലയിലിരുന്ന തൊപ്പിയെടുത്തു താഴെയിട്ടു . കുരങ്ങന്മാർ മറ്റൊന്നും ചിന്തിക്കാതെ അവരുടെ തലയിലിരുന്ന തൊപ്പികളും താഴേക്കിട്ടു .ഇതു കണ്ട തൊപ്പി വില്പ്പനക്കാരൻ ഉടൻതന്നെ തൻ്റെ തൊപ്പികളെല്ലാം നിലത്തു നിന്നും എടുത്തു. പിന്നെ അയാൾ ഒരു നിമിഷം പോലും അവിടെ നിന്നില്ല. അയാൾ തൻ്റെ തൊപ്പികളും സഞ്ചിയിലാക്കി സ്ഥലംവിട്ടു .

The input text given is free from all the punctuations except the full stop. Hence that made it easy to tokenize each of these sentences given.

TABLE 13. Expected output

“പണ്ട് ഒരു തൊപ്പി വിൽപ്പനക്കാരൻ ഉണ്ടായിരുന്നു. അയാൾ തൊപ്പി ഉണ്ടാക്കി ചന്തയിൽ വിൽക്കുമായിരുന്നു. പല സ്ഥലങ്ങളിൽ പോയി തൊപ്പി വിറ്റ് ഉണ്ടാക്കുന്ന കാശുകൊണ്ട് ആയിരുന്നു അയാൾ ജീവിച്ചത്. ഒരു ദിവസം അയാൾ കാട്ടിലെത്തി. കാട് കടന്നുവേണം ആയിരുന്നു അയാൾക്ക് ചന്തയിൽ എത്താൻ. അന്ന് കൂടുതൽ ചൂടുള്ള ദിവസമായിരുന്നു. അയാൾക്ക് നല്ല ക്ഷീണം തോന്നിയത് കൊണ്ട് അയാൾ വിശ്രമിച്ചു. അയാൾ തന്റെ തൊപ്പി സൂക്ഷിച്ചിരുന്ന സഞ്ചി ഒരു മരച്ചുവട്ടിൽ വെച്ചിരുന്നു. ക്ഷീണം കാരണം അയാൾ ഉറങ്ങിപ്പോയി. ഉറങ്ങിയെണീറ്റപ്പോൾ അയാളുടെ തൊപ്പി ഒന്നും കാണാനില്ല. അപ്പോഴാണ് തന്റെ തൊപ്പി വെച്ച് കുറെ കുരങ്ങന്മാർ മരത്തിൽ ഇരിക്കുന്നത് കണ്ടത്. എന്ത് ചെയ്യുമെന്ന് ആലോചിച്ച് അയാൾ കുരങ്ങന്മാരെ വിരട്ടി നോക്കി. അപ്പോൾ അവർ അയാളെ തിരിച്ച് വിരട്ടി. പിന്നീട് അയാൾ കമ്പെടുത്ത് വിരട്ടി, തിരിച്ചു കുരങ്ങന്മാരും അയാളെ വിരട്ടി. പെട്ടെന്ന് അയാൾക്ക് ബുദ്ധി തോന്നി. അയാൾ തലയിൽ ഇരുന്ന തൊപ്പി എടുത്ത് താഴോട്ടു. കുരങ്ങന്മാരും അവരുടെ തലയിൽ ഇട്ടുന്ന തൊപ്പികൾ താഴെയിട്ടു. അയാൾ ഉടൻ തന്നെ തൊപ്പികൾ എല്ലാം നിലത്തുനിന്നും എടുത്തിട്ട് സ്ഥലംവിട്ടു.”

Table 14. Generated output

“ആതൊപ്പി കച്ചവടക്കാരൻ തൻ്റെ തലയിലി രുന്ന തൊപ്പി യെടുത്തു താഴെയിട്ടു . ഒരു ദിവസം തൊപ്പി വിൽക്കാനായി പോയ കച്ചവടക്കാരൻ ഒരു കാട്ടിലെത്തി. തൊപ്പി വിറ്റുകിട്ടുന്ന കാശു കൊണ്ടായിരുന്നു അയാൾ ജീവിച്ചത്. ഒരു ദിവസം തൊപ്പി വിൽക്കാനായി പോയ കച്ചവടക്കാരൻ ഒരു കാട്ടിലെത്തി. ഇതുകണ്ട തൊപ്പി വില്പ്പനക്കാരൻ ഉടൻതന്നെ തൻ്റെ തൊപ്പികളെല്ലാം നിലത്തു നിന്നും എടുത്തു. ആതൊപ്പി കച്ചവടക്കാരൻ തൻ്റെ തലയിലി രുന്ന തൊപ്പി യെടുത്തു താഴെയിട്ടു.”

There are lots of differences in the expected output and generated output. This system is working on to handle more ambiguous data efficiently. The future scope of this work could be able to handle more complex datasets.

V. CONCLUSION

The paper presented a simple approach towards malayalam text summarization. Abstractive summarization was the main focus area. Hence it improved the efficiency and accuracy of the generated summary. An important factor in this experiment is the amount of data it can handle and the sentences from which it generates a summary. It mainly focuses on the sentence ranking, where each sentence is compared with one another. It is not completely independent of the extractive way, it includes the extractive approach also. As a part of future experiment work, the dataset used for training will be able to handle more in an abstractive way.

ACKNOWLEDGEMENT

We express our profound gratitude and deep regards to Associate Prof. Sheena Kurian K, of the Department of Computer Science and Engineering for all her effort, time and patience in helping us to complete the paper with all her

suggestions and ideas leading to the successful completion of the paper.

REFERENCES

1. M. F. Mridha ,Kamruddin Nur , Mahmud Hasan and Muhammad Mohsin Kabir , “A survey of automatic text summarization: progress, process and challenges”,IEEE, 2021.
2. Sindhya K Nambiar, David Peter S, Sumam Mary Idicula, ”Abstractive Summarization of Malayalam Document using Sequence to Sequence Model”, 2021 7th International Conference on Advanced Computing & Communication Systems (ICACCS). IEEE, 2021.
3. Xiaolong Wang,Xingtong Dong,Shuxin Chen. “Text Duplicated-checking Algorithm Implementation Based on Natural Language Semantic Analysis”. IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC 2020).
4. Limwattana, Siriwat, and Santitham Prom-on. Topic Modeling Enhancement using Word Embeddings. 18th International Joint Conference on Computer Science and Software Engineering (JCSSE). IEEE, 2021.
5. Deepa Mary Mathews , Sajimon Abraham , Effects of Pre-processing Phases in Sentiment Analysis for Malayalam Language, International Journal of Computer Sciences and Engineering ,2018.
6. J.N Madhuri, Ganesh Kumar.R, Extractive Text Summarization Using Sentence Ranking, International Conference on Electronics and Sustainable Communication Systems ((ICESC) IEEE, 2020.
7. Dhanya P.M, Dr. Sreekumar A, Dr. Jathavedan M, “Vriksh: A Tree Based Malayalam Lemmatizer Using Suffix Replacement Dictionary”International Journal of Emerging Technologies in Engineering Research (IJETER)2018.
8. Andres Arellano Government of Chile, Santiago, Chile, Edward Carney Lockheed Martin and Mark A. Austin “Natural Language Processing of Textual Requirements”, The Tenth International Conference on Systems, 2020.
9. Nallapati. R., Zhou. B., Santos. C., Gulcehre. C and Xiang. B. (2016). “Abstractive text summarization using sequence-to-sequence and beyond”. The SIGNLL Conference on Computational Natural Language Learning, 2016.
10. Kanitha D K, D. Muhammad Noorul Mubarak and S. A. Shanavas, “Malayalam Text summarization Using Vector Space Model”, International Journal of Engineering and Techniques - Volume 4 Issue 2, April-2018.
11. Renjith S R, Sony P,“An Automatic Text Summarization For Malayalam Using Sentence Extraction”, International Journal of Advanced Computational Engineering and Networking, Volume-3, Issue-8, Aug.-2015.
12. Syed Muqtadir Uddin Hussaini, Faraaz Mohd Khan, Faisal Khan, Dr. Abdul Subhan, “Text Summarization using Natural Language Processing”, Journal of Engineering Science, Vol 11, Issue 4 , April, 2020.
13. Zhang, Fan, ”Enhancing Short Text Topic Modeling with FastText Embeddings, International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE) IEEE , 2020.
14. Prabhudas Janjanam, CH Pradeep Reddy, Text Summarization: An Essential Study. Second International Conference on Computational Intelligence in Data Science (ICCIDS-2019). IEEE, 2019.