

## Multi-Document Text Summarization using Semantic Link Network

**Dipeksha Suryawanshi**  
MSC (Computer Science)  
Indira College, Malegaon

**Rutuja Ahire**  
MSC (Computer Science)  
Indira College, Malegaon

**Prof. Gokul B. Hiray**  
Indira College, Malegaon

---

**Abstract-:** *Large papers and texts must be manually summarised, which is tiresome and error-prone. Additionally, the outcomes of this type of summary might be different for a certain article. The enormous expansion of information and data has made automatic text summarization crucial. It selects the text passage that is the most instructive and creates summaries that highlight the primary goal of the provided content. It provides a summary generated by a summarising algorithm that enables users to understand the content of documents without having to read each one individually. Therefore, the main goal of Text Summarizer is to convey a text's content in fewer words and phrases. Abstractive and extractive summarization are two different types of summarization. This case study is built on an extractive idea that was applied to the models being looked at. There are several helpful automated text summary tools available now for English and other languages.*

*However, there aren't enough automated summarizers for languages spoken in India. Our major focus in this approach is creating an automated text summarizer for Marathi. We anticipate utilising the ROUGE measure to assess the obtained summary.*

**Keywords—** Summarization, Neural machine translation ,NLP

---

### I. INTRODUCTION

The summarization of texts is a method of reducing actual text to an abstract form while preserving all of its information and significance. It selects the text passage that is the most instructive and creates summaries that highlight the primary goal of the provided content. It provides a summary generated by a summarising algorithm that enables users to understand the content of documents without having to read each one individually. Therefore, the main goal of a text summarizer is to convey a text's content in fewer words and phrases. Abstraction-based summarising and Extraction-based summarization are two categories into which summarization systems may be divided.

Creating extractive summaries entails selecting pertinent sentences in a sequential order from the original material. Applying statistical and linguistically solid characteristics to the input text allows for the extraction of the pertinent sentences. But extraction has a limit. The extracted words and phrases are arranged chronologically. Conversely, abstractive text summaries are created by applying NLU principles.

Typically, this type of summarizer includes phrases that are absent from the original text. It attempts to mimic techniques employed by people, such as better and more thorough representation of a topic that is contained in the original piece. Although it is efficient, implementing it is highly challenging.

## II. LITERATURE SURVEY

These are the literature survey which is based on different research paper that related to our own project

Various fields make use of text summarization systems like, education field, social media (news articles, twitter, facebook messages), search engines, bio-medical field, government offices, researcher, etc [2].

Virat V. Giri and et al. reviewed text summarizers based on various Indian languages and their performances. They studied and proposed summarization method for marathi in detail wherein marathi stemmer, marathi proper name list, EnglishMarathi noun list, marathi keywords extraction, marathi rule based named entity recognition etc. for pre-processing of text followed by processing of text [1].

Sheetal Shimpikar and et al. studied various techniques of text summarization for various Indian languages [2]. Sunitha C and et al. worked on Abstractive summarization methods that are used for Indian languages. They explained Abstractive summarization technique, classified in two approaches such as structure based approach and semantic based approach [3]. Hamzah Noori Fejer and et al. gave a major contribution by proposing a combined approach of clustering technique and extracting keyphrases. They have proposed a new approach of clustering which combines hierarchical and k-means clustering. The results obtained from their experiments proved the proposed model gives better performance when compared with existing ones [4]. An unsupervised approach for marathi stemmer has been discussed by Mudassar Majgaonker and et al. [6]. The present work on text summarization of marathi text with question based system using rule based stemmer technique or generating question, we used rule based approach of abstractive text summarization and POS tagger, NER tools and rule based stemmer. Here marathi text is taken as input, on it POS tagger is applied and then questions are generated for the given input as per marathi language rules by Deepali K. Gaikwad and et al. At this stage they have framed rules of stemmer only for who 'type questions [5]. Thus it can be extended to learning all What type questions too.

Mangesh Dahale proposed text summarizer using inverted indexes [9]. Jayshri Patil and et al. reviewed different approaches of Named Entity Recognition (NER) and discussed issues and challenges arising in Indian languages [8]. Pooja Pandey and et al. discussed extraction of root words using morphological analyzer for devanagari script [11]. Aishwarya Sahani and et al. contributed to automatic text categorization of marathi language documents [7]. Rafael Ferriera et. Al used four dimensional graph based model for text summarization which relies on four dimensions(similarity, semantic similarity, co-reference, discourse information) to create the graph [16]. Federico Barrios et. al used variations in similarity measures along with TextRank for summarization [15]. Our work includes use of TextRank along with positional distribution of sentence scores and considering thematic similarity which gave promising results.

## III. PROPOSED SYSTEM

Text summarization is a method of reducing actual text to an abstract form while preserving all of its information and significance. It selects the text passage that is the most instructive and creates summaries that highlight the primary goal of the provided content. It provides a summary generated by a summarising algorithm that enables users to understand the content of documents without having to read each one individually. Therefore, the main goal of a text summarizer is to convey a text's content in fewer words and phrases. Abstraction-based summarising and Extraction-based summarization are two categories into which summarization systems may be divided. vindicated by each bumps on the block chain. To make any sale valid, the bumps of the block chain will have to give the evidence of the processing it has done in order to corroborate the sale. That evidence will be taken in terms of the quantum of processing done. The below mentioned sale system has two benefits. this system would be suitable to apply a distributed system as well as the banking bumps could be semi-automatized so as to reduce work.

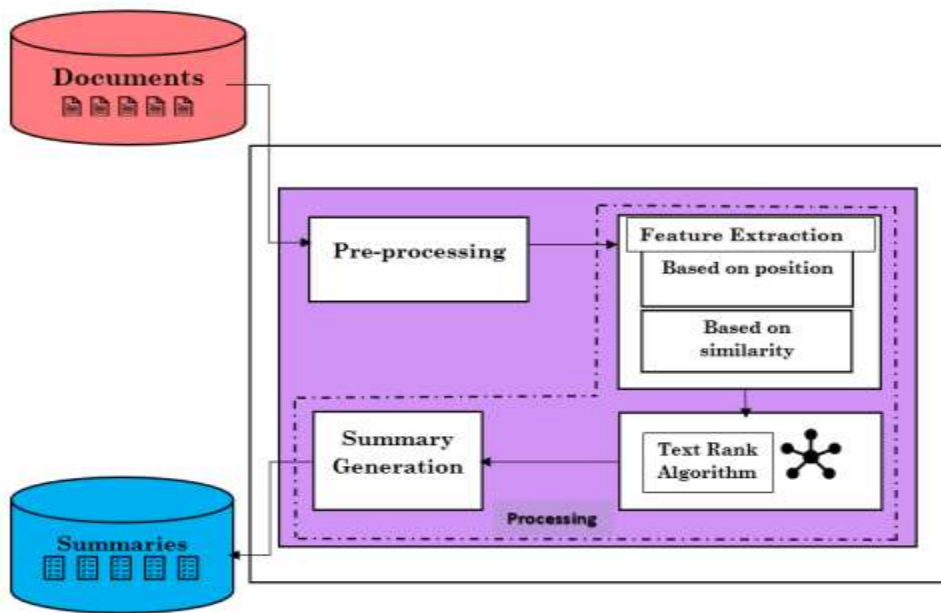


Figure 1: Architecture diagram

A: DataSetText will be extracted from documents on a variety of topics, including news items, politics, and many documents from the collection. We have 634 papers that are based on news stories that explore various topics in Marathi. For getting many documents, one can utilise the EMILLE (Enabling Minority Language Engineering), which provides monolingual, parallel, and annotated corpora for Asian languages, including marathi. Pre-Processing and processing stage are the two major steps that make up the system. Stage

B: Pre-processing-:The pre-processing phase of text summary is crucial. It produces pre-processed data that is excellently suited for the processing step. Pre-processing stages often involve processes to eliminate punctuation, tokenization, stop words, stemming, etc. Various preliminary processing stages will be covered in this section.

C. Feature extraction-:The features like SOV (Subject Object Verb - Experimental) verification, sentence positional value (POS tagging), TF-ISF (Term Frequency/ Inverse Sentence Frequency) or TF-IDF (Term Frequency/ Inverse Document Frequency) are extracted from pre-processed sentences. Sentences are further ranked on basis of features extracted.

Class Diagram-:

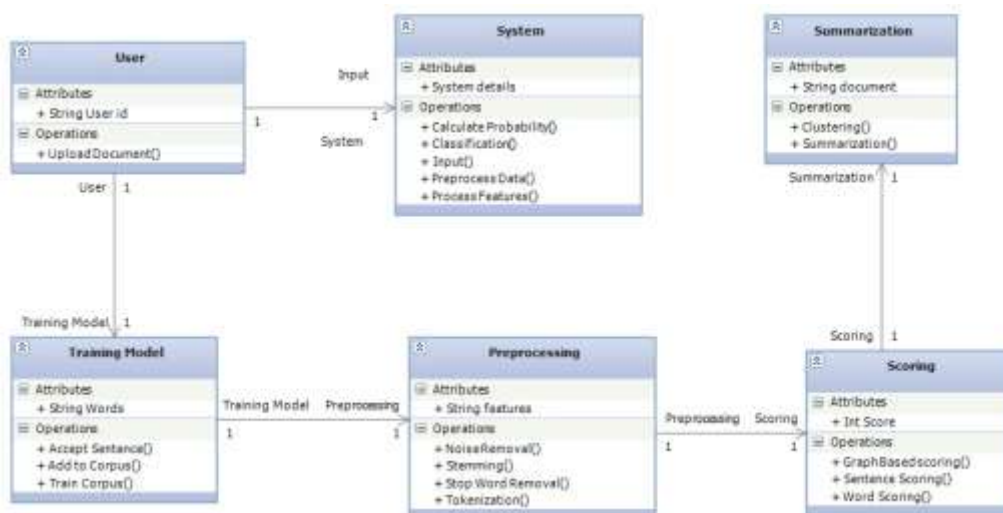


Figure 2: Class diagram

## CONCLUSION

A quick and efficient automatic summarising method is required since there is such a huge growth in the volume of material that is available online. The phases of feature extraction, scoring, and graph construction in this system method are crucial. This approach is applicable to a number of industries, including education, search engine performance improvement, and Marathi media clustering. In future we can add multi document summarization

## REFERENCES

- [1] Ms. DeepaliKadam, "International Journal of Innovationsand; Advancement in Computer Science" IJIACS ISSN2347 – 8616 Volume 4, Special Issue March 2015.
- [2] Ramesh Nallapati, FeifeiZhai, Bowen Zhou. 2016.SummaRuNNer: A Recurrent Neural Network-basedSequence Model for Extractive Summarization of Documents arXiv: 1611.04230v1 [cs.CL] 14 Nov 2016
- [3] Ramesh Nallapati, Bowen Zhou, Cicero dos Santos.Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond.ArXiv: 1602.06023v5[cs.CL], 2016.
- [4] D. K. Kanitha, D. Muhammad Noorul Mubarak & S.A. Shanavas. 2018. COMPARISON OF TEXTSUMMARIZER IN INDIAN LANGUAGES"International Journal of Advanced Trends in Engineeringand Technology (IJATET)" Volume 3, Issue 1
- [5] Virat V. Giri, Dr.M.M. Math, and Dr.U.P.Kulkarni, ASurvey of Automatic Text Summarization System forDifferent Regional Language in India, BonfringInternational Journal of Software Engineering and SoftComputing, Vol. 6, Special Issue, October 2016.
- [6] AliakseiSeveryn and Alessandro Moschitti, Learning to Rank Short Text Pairs with Convolutional Deep Neural Networks, <http://citeseerx.ist.psu.edu>.
- [7] Tarasov D. S., Natural Language Generation, Paraphrasing and Summarization of User Reviews with Recurrent Neural Networks, ReviewDot Research, Kazan, Russia.
- [8] Tian Shi, YaserKeneshloo, NarenRamakrishnan, Chandan K. Reddy, Senior Member, IEEE, Neural Abstractive Text Summarization with Sequence-toSequence Models, arXiv:1812.02303v2 [cs.CL] 7 Dec 2018.
- [9] RupalBhargava, SukrutNigwekar, Yashvardhan Sharma, Catchphrase Extraction from Legal Documents Using LSTM Networks, <http://ceur-ws.org/Vol-2036/>