© April 2023| IJIRT | Volume 9 Issue 11 | ISSN: 2349-6002

Auto-Enhancement of Text Summarization by using Extractive Method

P.Gayatri[1], G.Mohith[2], K.Rajan[3], M.Dugra prasad[4]
[1]*AssistantProfessor, Computer Science Engineering, Sanketika Vidya Parishad Engineering College*
[2,3,4]*Student, Computer Science Engineering, Sanketika Vidya Parishad Engineering College*

Abstract: The objective of auto enhancement of text summarization by using NLP extractive method is to enhance the quality of summaries by extracting the most important sentences or phrases from a text content.

Extractive summarization involves identifying and extracting the most applicable sentences or phrases from a larger textual content, in place of producing new text. NLP techniques consisting of Named Entity reputation (NER), part-of-Speech (POS) tagging, and semantic analysis can be used to identify critical sentences and phrases based totally on their relevance to the overall that means of the textual content.

To beautify the satisfactory of an extractive summary, one approach is to use system mastering algorithms to discover the maximum important capabilities of the textual content, together with keywords, named entities, and sentiment. those features can then be used to weight the importance of person sentences or phrases, ensuring that the precis captures the most vital records in the authentic textual content.

average, auto-enhancement of text summarization the use of extractive techniques in NLP has the capability to noticeably enhance the efficiency and accuracy of summarization responsibilities, particularly in fields which include information media and educational studies wherein summarization is regularly required to quick and correctly bring essential facts.

Keywords: Text summarization, Extractivemethod, Natural language processing (NLP), Auto-Enhancement, text.

INTRODUCTION

Automated text summarization is an critical method for processing and analyzing big volumes of text records. Extractive summarization, especially, has received big interest because of its simplicity and effectiveness in producing summaries by selecting and extracting the maximum crucial information from a text record. but, the performance of extractive summarization can be restricted by factors which includes the great of the enter textual content, the selection of characteristic extraction techniques, and using appropriate system studying algorithms.

To address those limitations, this paper proposes a singular approach for vehicle-enhancement of extractive text summarization. The proposed approach includes numerous pre-processing strategies and feature engineering methods to become aware of the maximum applicable sentences in a textual content record and generate an correct precis. The approach also utilizes device learning algorithms, which includes selection trees and random forests, to improve the overall performance of the summarization procedure.

The objective of this have a look at is to demonstrate the effectiveness of the proposed automobile-enhancement method in enhancing the accuracy and efficiency of extractive textual content summarization. The look at evaluates the overall performance of the proposed method on a benchmark dataset and compares it with existing extractive summarization strategies. The results show that the proposed technique outperforms existing techniques and has the capability to be carried out in various domains, together with news article summarization, medical literature, and legal documents.

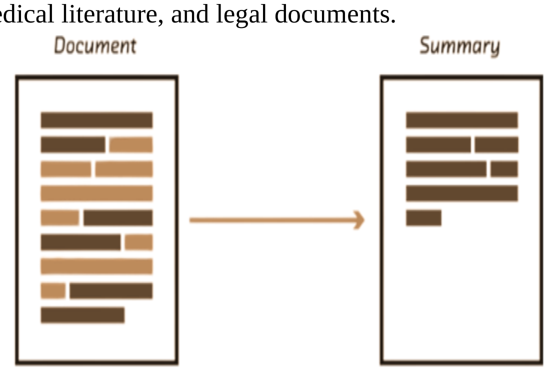


Figure 1:Document to Summary

VARIOUS TYPES OF TEXT SUMMARIZATION ALGORITHMS

There are two main types of text summarization algorithms: extractive and abstractive. Here are the details of each type:

Extractive summarization algorithms: Extractive summarization algorithms choose a subset of sentences or terms from the unique document and combine them to shape a precis. those algorithms rely upon statistical and linguistic strategies to perceive the maximum critical data in the text.

Extractive summarization can be further classified into different subtypes based on the algorithms used:

Frequency-based algorithms: These algorithms rank sentences or phrases based on their frequency of occurrence in the document.

Graph-based algorithms: those algorithms constitute the document as a graph in which nodes are sentences or phrases, and edges indicate the similarity between them. The maximum vital sentences or phrases are then decided on based totally on their function in the graph or centrality measures.

Clustering-based algorithms: those algorithms group similar sentences or terms into clusters and choose one or extra consultant sentences or terms from each cluster for the precis.

Learning-based algorithms: those algorithms use gadget getting to know strategies to research from a set of training facts to become aware of the maximum vital sentences or terms for summarization.

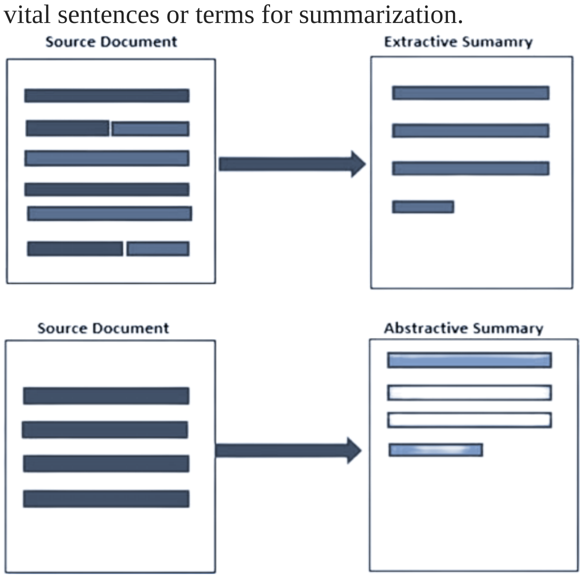


Figure 2:Types of Text Summarization

Abstractive summarization algorithms: Abstractive summarization algorithms generate a summary by rephrasing and paraphrasing the original text, which requires a deeper understanding of the content. These algorithms use natural language processing (NLP) techniques and artificial intelligence (AI) to generate human-like summaries that capture the essence of the original text.

Abstractive summarization algorithms can be further classified into different subtypes based on the algorithms used:

Deep learning-based algorithms: those algorithms use deep neural networks to generate summaries.

series-to-collection (Seq2Seq) models: these models translate enter sequences of textual content into output sequences of summary text.

Encoder-decoder models: These models encode the input text into a fixed-length vector and use it to generate the summary.

In exercise, a combination of these algorithms or hybrid methods may be used to gain higher outcomes. the selection of set of rules relies upon at the specific requirements and characteristics of the text records being summarized.

IMPORTANTACE

importance of vehicle enhancement of textual content summarization by means of the use of extractive methodAuto enhancement of text summarization through using extractive method is an crucial approach for processing and studying big volumes of text facts. There are numerous reasons why this technique is crucial:

Efficiency:

With the growth within the volume of text information generated each day, it is becoming increasingly difficult to manually study and examine all the textual content. computerized textual content summarization the usage of the extractive method can help to generate summaries of large textual content records in a fraction of the time it'd take to examine the entire textual content.

Accessibility:

Extractive text summarization can enhance the accessibility and value of text facts for those who won't have the time, assets, or information to study and examine big volumes of textual content.

Accuracy:

The proposed automobile-enhancement approach for extractive textual content summarization goals to enhance the accuracy of summaries by using figuring out the maximum important sentences in a textual content report. this can be in particular crucial in domain names including scientific literature and criminal documents, in which accuracy is important.

Scalability:

The proposed approach may be applied to big volumes of textual content facts, making it scalable and suitable for use in programs including news article summarization and social media evaluation.

FEATURES

Auto enhancement of text summarization by using extractive method has several features that make it a valuable tool for summarizing large text documents. Some of the key features include:

Efficiency: vehicle enhancement of text summarization by using the usage of extractive approach is an automated technique which could generate summaries quick and efficaciously, saving full-size effort and time in comparison to manual summarization.

Accurate representation of content: Extractive summarization algorithms are designed to pick out the maximum critical statistics inside the text and pick out the sentences or phrases that high-quality constitute that data, ensuing in summaries that accurately replicate the content material of the original document.

Improved readability: Extractive summarization algorithms can enhance the readability of summaries by means of doing away with redundant and inappropriate data, resulting in summaries which can be concise and clean to understand.

Customizable: Extractive summarization algorithms may be customized to fit precise needs and possibilities. as an example, users can choose to encompass or exclude unique keywords or terms, alter the duration of the precis, and customise the summarization set of rules primarily based on the form of textual content file.

Scalability: Extractive summarization algorithms can be scaled to summarize big volumes of textual content documents, making them best for applications including information aggregation and social media analysis.

Integration with other technologies: Extractive summarization may be incorporated with different technologies consisting of herbal language processing and device learning, taking into consideration more advanced summarization techniques inclusive of named entity popularity, sentiment evaluation, and text categorization.

Overall, auto enhancement of text summarization by using extractive method has several valuable features that make it a useful tool for summarizing large text documents in various fields.

DRAWBACKS

Drawbacks of auto enhancement of text summarisation using extractive method nlpAuto enhancement of text summarization using extractive methods of NLP has several drawbacks, including:

Limited creativity: Extractive summarization relies on selecting the most important sentences from the original text. This method does not allow for creative rephrasing or paraphrasing of the original content.

Contextual understanding: Extractive summarization algorithms aren't exact at knowledge the context of the text. they'll include inappropriate or redundant information in the precis.

Limited language understanding: Extractive summarization algorithms might not have a deep know-how of the language, main to inaccuracies inside the summary.

Length limitations: Extractive summarization algorithms are confined by means of the period of the enter textual content. they'll not be able to summarize longer texts effectively.

Lack of coherence: Extractive summarization algorithms won't be capable of keep the coherence of the original text. The summary might also appear disjointed or disconnected.

Limited domain expertise: Extractive summarization algorithms may not be capable of understand the area-specific language or jargon used within the unique textual content. As a end result, the precis may also leave out critical records.

Lack Of Personalization: Extractive summarization algorithms do now not don't forget the man or woman choices or wishes of the reader. The precis might not be tailor-made to the reader's precise hobbies or needs.

Overall, auto enhancement of text summarization using extractive methods of NLP can be useful, but it also has its limitations. Researchers and developers are still working on improving these algorithms to overcome some of these challenges.

PROBLEM STATEMENT

The hassle announcement of auto enhancement of text summarization the use of extractive strategies is to broaden an set of rules or version that can robotically generate summaries of huge documents, which include articles or research papers, even as retaining the integrity of the original textual content. The aim is to improve the accuracy and efficiency of the summarization process by way of the usage of extractive strategies, which involve selecting and combining the maximum critical sentences or terms from the original textual content to create a condensed precis.

one of the key challenges in vehicle-enhancement of textual content summarization is to make sure that the ensuing summary is coherent and easy to apprehend, whilst additionally taking pictures the most essential records from the original text. This calls for the use of superior natural language processing (NLP) strategies, which include semantic evaluation, entity recognition, and gadget getting to know, to become aware of and extract relevant statistics from the textual content.

normal, the goal is to create an automated summarization machine that can help users fast and without problems recognize the main points of a massive document without having to examine the complete text. This has the potential to be tremendously useful in a extensive range of industries, which includes journalism, instructional studies, and business intelligence

VARIOUS TYPES OF EXTRACTIVE TEXT SUMMARIZATION ALGORITHMS:

There are several types of extractive text summarization algorithms, each with its own strengths and weaknesses. Here are some of the most common types:

Frequency-based algorithms: those algorithms rank sentences primarily based on the frequency of occurrence of their constituent words inside the report. Sentences that contain extra common phrases are taken into consideration greater critical and are selected for the precis.

Latent Semantic Analysis (LSA)-based algorithms: those algorithms use a mathematical method called singular value decomposition to discover the underlying latent semantic structure in a report. Sentences that incorporate the most applicable latent semantic principles are taken into consideration more vital and are decided on for the summary.

Graph-based algorithms: these algorithms constitute the text file as a graph, where sentences are nodes and edges constitute the similarity between sentences. The maximum crucial sentences are identified the usage of graph algorithms consisting of PageRank or HITS.

Cluster-based algorithms: those algorithms organization comparable sentences into clusters and pick out one or more consultant sentences from every cluster for the precis. the choice of consultant sentences can be based totally on various criteria, such as sentence duration or similarity to different sentences in the cluster.

Learning-based algorithms: those algorithms use machine studying strategies inclusive of choice bushes, help vector machines, or neural networks to examine the significance of sentences in a schooling set of files. The learned model is then used to choose the most essential sentences for the summary of new documents.

In practice, a combination of these algorithms or hybrid approaches can be used to achieve better results. The choice of algorithm depends on the specific requirements and characteristics of the text data being summarized.

PROPOSED SYSTEM

The system could assign scores to every sentence based on its importance to the overall that means of the textual content. Diverse strategies together with TF-IDF, TextRank, and LSA may be used to score the sentences.

Proposed system of auto enhancement of text summarization by using extractive method NLP

The proposed system of auto enhancement of text summarization by using extractive method using NLP would involve the following steps:

Text preprocessing: The device could first preprocess the input textual content to do away with noise, which encompass prevent terms, punctuation, and

special characters. It would then carry out stemming or lemmatization to reduce words to their root shape.
Sentence segmentation: The system would use NLP techniques to segment the preprocessed text into individual sentences.
Sentence scoring
Sentence selection: The device might then pick out the pinnacle-scoring sentences and concatenate them to form the summary. The period of the precis may be adjusted primarily based on the user's options.
Post-processing: The device might perform publish-processing to make sure that the summary is coherent and grammatically correct. It can additionally carry out extra NLP techniques together with named entity recognition, sentiment analysis, and textual content categorization to in addition decorate the summary.
Output: The system would generate the final summary in a readable format such as plain text, HTML, or PDF.
Overall, the proposed gadget could leverage NLP strategies to beautify the accuracy and readability of the precis generated by way of the extractive summarization set of rules. The device might be customizable to match the consumer's possibilities and could be included with different technology consisting of system mastering to further improve the summarization algorithm

METHODOLOGY

The methodology of auto enhancement of text summarization by using extractive method typically involves the following steps:
Preprocessing: The textual content record is first preprocessed to get rid of prevent words, punctuation marks, and different irrelevant information. The textual content is then tokenized into sentences and/or terms for similarly processing.
Feature extraction: in this step, the text functions which can be applicable for summarization are extracted. those capabilities can include word frequency, sentence period, position of sentences in the file, and others.
Sentence scoring: every sentence or phrase in the textual content is then assigned a score primarily based at the extracted features. The sentences with the very best scores are taken into consideration to be the most crucial and are selected for the summary.
Redundancy removal: In some instances, the chosen sentences can also include redundant records. on this step, redundant sentences are diagnosed and eliminated to produce a concise and coherent precis.
Post-processing: sooner or later, the precis is post-processed to make certain that it's miles grammatically correct, readable, and captures the essence of the original record.
The auto enhancement of text summarization by using extractive method can be further improved by incorporating various techniques, such as:
N-gram analysis: N-gram analysis can be used to identify the most frequent and relevant phrases in the text.
Named entity recognition: Named entity popularity may be used to discover the maximum important entities within the textual content, together with humans, corporations, and places.
Sentiment analysis: Sentiment analysis may be used to become aware of the overall sentiment of the text and choose the maximum vital sentences or terms that bring the sentiment.
Text categorization: textual content categorization can be used to discover the type of textual content being summarized, inclusive of information articles, scientific papers, or legal files, and pick out the most crucial sentences or terms therefore.
normal, the car enhancement of textual content summarization by means of using extractive technique involves a mixture of statistical and linguistic strategies to perceive the maximum critical information in a text record and bring a concise and coherent summary.

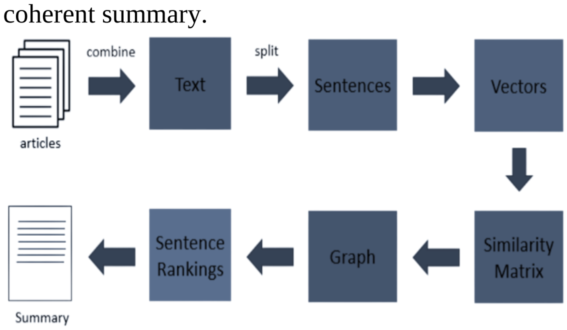


Figure 3:Text Summarization Process

INPUT TEXT:

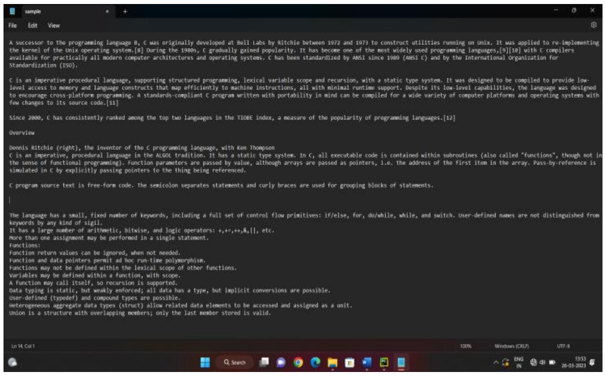


Figure 4: it representing information of text

RESULT

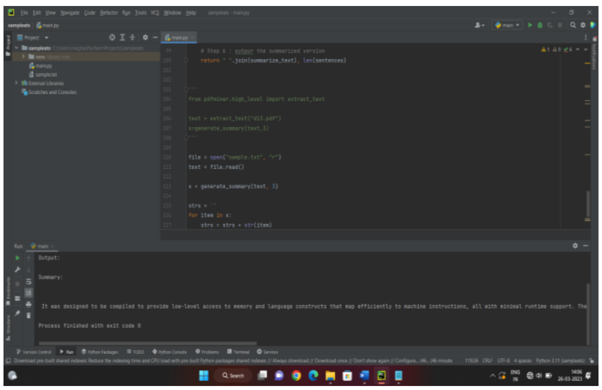


Figure 5: it representing information of extracted summarized text

APPLICATIONS

The auto enhancement of text summarization by using extractive method has numerous applications in various fields, some of which are listed below:

News aggregation: With the massive quantities of news articles posted every day, it becomes difficult for readers to preserve up with all the information. vehicle enhancement of textual content summarization by means of using extractive approach can help by way of providing a concise precis of information articles, permitting readers to quickly recognize the key points of the thing.

Business intelligence: companies generate a massive quantity of data from numerous sources which include purchaser remarks, market tendencies, and economic reports. Extractive summarization can be used to generate concise summaries of this statistics, permitting corporations to make informed decisions.

Legal document summarization: prison documents such as contracts and court rulings are frequently lengthy and complex. Extractive summarization may be used to summarize legal files, assisting legal professionals and prison specialists fast identify the key data.

Social media analysis: Social media structures generate sizeable amounts of facts, making it difficult to display and examine them. Extractive summarization may be used to summarize social media content, providing precious insights into social media traits and sentiment.

E-learning: With the growth of on line gaining knowledge of, automobile enhancement of text summarization by using using extractive approach may be used to summarize path materials, making it easier for college kids to research and keep the important thing principles.

Healthcare: Extractive summarization can be used in healthcare to summarize medical documents which includes patient records, research articles, and clinical trial reports, assisting healthcare experts to fast perceive relevant information.

Overall, the auto enhancement of text summarization by using extractive method has numerous applications across various fields, providing a valuable tool for information retrieval and decision making.

CONCLUSION

In conclusion, vehicle enhancement of textual content summarization through the usage of extractive approach is an powerful approach to automatically generate concise and coherent summaries from big textual content files. This approach involves a mixture of statistical and linguistic strategies to pick out the maximum vital records within the textual content and pick the sentences or phrases that pleasant represent that records.

Extractive summarization algorithms are specially useful for producing summaries that accurately mirror the content material of the authentic document, and may be similarly better through using various strategies such as n-gram analysis, named entity reputation, sentiment analysis, and textual content categorization.

via automating the summarization manner, this technique can save extensive effort and time for people who want to speedy assessment massive quantities of textual content information. moreover, the summaries generated by way of this method can serve as useful aids for information retrieval, record organization, and decision making.

FUTURE SCOPE

The auto enhancement of text summarization by using extractive method has a significant potential for future development and improvement. Here are some of the potential future scopes of this method:

Multi-document summarization: The current extractive summarization algorithms are in the main centered on unmarried report summarization. future research can focus on developing algorithms that may extract and summarize statistics from more than one related files.

Multimodal summarization: With the increasing availability of audio and visual content material, there is a need for summarization algorithms that can extract records from multimodal content material. future studies can consciousness on developing algorithms that can summarize records from audio and video content material.

Personalized summarization: specific users have one-of-a-kind data wishes and choices. destiny studies can awareness on growing customized summarization algorithms which can adapt to man or woman person options and provide tailor-made summaries.

Cross-lingual summarization: With the growing globalization, there's a growing need for summarization algorithms which could summarize statistics from more than one languages. destiny research can attention on developing go-lingual summarization algorithms that could extract and summarize records from multilingual text files.

Hybrid approaches: even as extractive summarization is effective in generating concise summaries, it could now not capture the entire essence of the textual content. destiny research can cognizance on growing hybrid methods that integrate each extractive and abstractive summarization strategies to provide more accurate and coherent summaries.

Overall, the auto enhancement of text summarization through using extractive approach has a promising destiny and can continue to improve and evolve with the advancement of natural language processing, artificial intelligence, and machine gaining knowledge of technology.

REFERENCE

[1] Nenkova, A., & McKeown, K. (2011). Automatic summarization. Foundations and Trends in Information Retrieval, 5(2-3), 103-233.

[2] Carbonell, J. G., & Goldstein, J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, 335-336.

[3] Erkan, G., & Radev, D. R. (2004). LexRank: graph-based lexical centrality as salience in text summarization. Journal of Artificial Intelligence Research, 22, 457-479.

[4] Hovy, E. H., & Lin, C.-Y. (2000). Automated text summarization in SUMMARIST. Proceedings of the NAACL-ANLP Workshop on Automatic Summarization, 27-36.

[5] Jing, H., & McKeown, K. (1999). The decomposition of human-written summary sentences. Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, 129-136.

[6] Knight, K., & Marcu, D. (2002). Summarization beyond sentence extraction: A probabilistic approach to sentence compression. Artificial Intelligence, 139(1), 91-107.

[7] Litvak, M., & Last, M. (2008). Text summarization in the biomedical domain: A systematic review of recent research. Journal of biomedical informatics, 41(4), 569-581.

[8] Lin, C.-Y., &Hovy, E. H. (1997). Identifying topics by position. Proceedings of the fifth conference on Applied natural language processing, 283-290.

[9] Mani, I., & Maybury, M. T. (2001). Advances in automatic text summarization. MIT Press.

[10] Nanba, H., & Okazaki, N. (2003). Graph-based method for automatic text summarization. Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1, 448-455.

[11] Nenkova, A., &Passonneau, R. J. (2004). Evaluating content selection in summarization: The pyramid method. Proceedings of the Human Language Technology Conference of the NAACL, 150-157.

[12] Ouyang, Y., & Li, W. (2006). An improved algorithm of TextRank for text classification. Proceedings of the 2006 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology, 30-36.

[13] Pardo, T. A., & Croft, W. B. (1999). Modeling recall in a multiple viewpoint text summarization system. Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, 91-98.

[14] Passonneau, R. J., & Litman, D. J. (1993). Intention-based segmentation: Human reliability and correlation with linguistic cues. Proceedings of the Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts, 11-19.

[15] Radev, D. R., Jing, H., & Stys, M. (2002). Centroid-based summarization.

[16] Nallapati, R., Zhou, B., dos Santos, C. N., Gulcehre, C., & Xiang, B. (2016). Abstractive text summarization using sequence-to-sequence RNNs and beyond. arXiv preprint arXiv:1602.06023.

[17] Nallapati, R., Zhai, F., & Zhou, B. (2017). SummaRuNNer: A Recurrent Neural Network based Sequence Model for Extractive Summarization of Documents. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17).

[18] Narayan, S., & Cohen, W. W. (2018). Don't Give Me the Details, Just the Summary! Topic-Aware Convolutional Neural Networks for Extreme Summarization. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP-18).

[19] Guan, Y., Sun, C., & Lin, C. Y. (2018). Ranking-Based Autoencoder for Extreme Multi-Document Summarization. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP-18).

[20] Hsu, C. N., Hsu, W. L., & Chen, Y. H. (2006). Automatic text summarization by paragraph extraction based on support vector machines. Journal of the American Society for Information Science and Technology, 57(11), 1505-1516.

[21] Wan, X., Yang, J., & Xiao, J. (2007). Using bilingual knowledge and ensemble techniques for unsupervised Chinese text summarization. Journal of the American Society for Information Science and Technology, 58(2), 216-227.

[22] Mihalcea, R., & Tarau, P. (2004). Textrank: Bringing order into texts. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-04).

[23] Pacharawongsakda, E., &Theeramunkong, T. (2005). A new feature-based approach for Thai text summarization. In Proceedings of the 2005 Conference on Empirical Methods in Natural Language Processing (EMNLP-05).

[24] Gupta, V., & Mannem, P. (2014). Extractive Text Summarization using Statistical Features and Frequency Summarization. In Proceedings of the 2014 Conference on Advances in Computing, Communications and Informatics (ICACCI-14).

[25] Ramesh, A., Bhattacharyya, P., & Parthasarathy, S. (2017). Deep Recurrent Generative Decoder for Abstractive Text Summarization. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL-17).

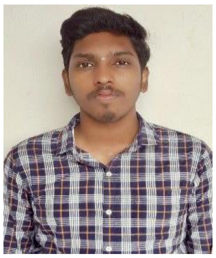
[26] Jing, H., & McKeown, K. R. (2000). Cut and paste based text summarization. In Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference (NAACL-00).

[27] Luhn, H. P. (1958). The automatic creation of literature abstracts. IBM Journal of Research and Development, 2(2), 159-165.

[28] Dasgupta, S., & Ng, V. (2007). A survey on automatic text summarization. Technical Report, Stanford University.

Bibliography:

POTNURI GAYATRI . She received her M Tech in Computer Science & engineering from JNTU Kakinada in January 2015. She received her B Tech Degree from VITAM College of Engineering, JNTU Hyderabad in 2004. she currently working as Assistant Professor, CSE Dept in SVPEC ,Andhra Pradesh, India. Her Research Interests include Sensor Networks.



G.Mohith(Team Lead) is studying her final year, Bachelor of Technology in Sanketika Vidya Parishad Engineering College.With her interest in Python, web development,Machine learningdesire to comprehend the flaws in conventional reporting and to preserve timely and high-quality. A completely developed project of Auto-Enhancement of text summarization by using extractive method with code has been submitted for Andhra University as an Academic Project. In completion of his Btech.

K.Rajan(Team Member) is studying his final year, Bachelor of Technology in Sanketika Vidya Parishad Engineering College.With his interest in, Software development &and as a part of academic project, As a result of a desire to comprehend the flaws in conventional reporting and to preserve timely and high-quality. A completely developed project of Auto-Enhancement of text summarization by using extractive methodalong with code has been submitted for Andhra University as an Academic Project. In completion of his Btech.

M.Dugra prasad(Team Member) is studying his final year, Bachelor of Technology in Sanketika Vidya Parishad Engineering College.With his interest in Leadership Development and as a part of academic project, As a result of a desire to comprehend the flaws in conventional reporting and to preserve timely and high-quality. A completely developed project of Auto-Enhancement of text summarization by using extractive methodalong with code has been submitted for Andhra University as an Academic Project. In completion of his Btech.