



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 5, May 2023

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

[ijrset@gmail.com](mailto:ijrset@gmail.com)

[www.ijrset.com](http://www.ijrset.com)



# Data Summarization Using NLP

Vaishnavi S. Masale<sup>1</sup>, Onkar S. Kore<sup>2</sup>, Pravin R. Narone<sup>3</sup>, G. M. Kosgikar<sup>4</sup>

U.G. Student, Department of Electronics & Telecom Engineering, Brahmaveddada Mane College, Belati,  
Solapur, India<sup>1234</sup>

Associate Professor, Department of Electronics & Telecommunication Engineering, BMIT College, Belati,  
Solapur, India<sup>4</sup>

**ABSTRACT:** The amount of text data available has increased dramatically in recent years from a variety of sources. This large volume of literature has a wealth of information and knowledge that must be adequately summarized in order to be useful. Text summarizing is one of the jobs that has received a lot of attention, but there hasn't been any practical application for it aside from tasks based on Extractive summarization, such as news or book summarization. Text summarization creates an automatically generated summary that includes all pertinent significant information from the original material. Observing the summary results, extractive and abstractive techniques are one of the most common. Extractive summarizing is maturing, and research is currently shifting to abstractive summarization and real-time summarization. The results of the analysis provide an in-depth explanation of the trends that are the focus of their research in the field of text summarization, pre-processing and features that have been used, describes the techniques and methods that are often used by researchers as a comparison and means for developing methods. In this study, various recommendations are made regarding opportunities and problems in text summarization research, and why abstraction and hybrid summarization methods produce the best results.

**KEYWORDS:** Machine Learning, NLP Concept etc.

## I. INTRODUCTION

The technique, where a computer program shortens longer texts and generates summaries to pass the intended message, is defined as Automatic Text Summarization and is a common problem in machine learning and natural language processing (NLP).

Text summarization is the process of creating a short, coherent, and fluent summary of a longer text document and involves the outlining of the text's major points.

Text identification, interpretation and summary generation, and analysis of the generated summary are some of the key challenges faced in the process of text summarization. The critical tasks in extraction-based summarization are identifying key phrases in the document and using them to discover relevant information to be included in the summary.

The Main Purpose of data summarization is to get meaningful summarization with using abstractive extraction method. The big data sets, Documents or Articles are very hard to read every time so we are going to give solution called data summarization.

## II. BLOCK DIAGRAM

Data summarization is the process of generating a concise and meaningful summary of a large body of text or data. Natural Language Processing (NLP) techniques can be used to automatically generate summaries of text documents, news articles, scientific papers, or any other type of textual content.

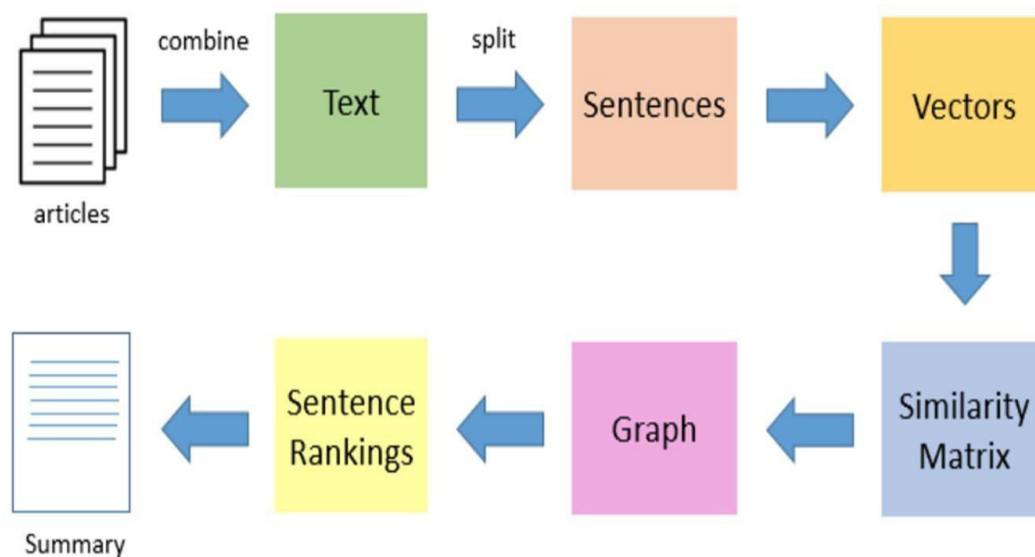
NLP-based data summarization typically involves several steps. The first step is to preprocess the text by removing stop words, punctuation, and other noise. The next step is to identify the most important sentences or phrases in the text, which can be done using various methods such as frequency analysis, graph-based algorithms, or machine learning models.

Once the important sentences or phrases have been identified, they can be combined into a summary using various techniques such as sentence fusion, extraction-based methods, or abstractive methods. Extraction-based methods



involve selecting the most relevant sentences from the original text and concatenating them to form a summary, while abstractive methods involve generating new sentences that capture the key information from the original text. NLP-based data summarization has many applications in various fields such as news summarization, scientific literature summarization, social media summarization, and business intelligence. It can help users quickly understand the key points of a large amount of text, identify trends and patterns, and make informed decisions based on the summarized information.

### III. EXPERIMENTAL RESULTS



### IV. CONCLUSION

In conclusion, data summarization is a technique used to extract the most important information from a source document and present it in a condensed form. This process is essential for dealing with large volumes of text and can save a significant amount of time compared to manually summarizing text.

The proposed project aims to provide a software solution for automating the text summarization process, which will allow users to quickly generate summaries of lengthy documents. By reducing the amount of text and presenting only the most important information, the software can improve efficiency and productivity in a variety of fields, including business intelligence, news and media, legal industry, and healthcare.

Overall, data summarization is an important tool for managing large volumes of text, and the development of automated solutions like the proposed project can have significant benefits in terms of accuracy, productivity, and cost-effectiveness.

### REFERENCES

1. Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing order into texts. Proceedings of EMNLP, 2004, 404-411.
2. Erkan, G., & Radev, D. (2004). LexRank: Graph-based lexical centrality as salience in text summarization. Journal of Artificial Intelligence Research, 22, 457-479.
3. Hovy, E., & Lin, C.-Y. (1998). Automated text summarization in SUMMARIST. Proceedings of the 2nd Conference on Language Resources and Evaluation (LREC), 881-888.
4. Nallapati, R., Zhai, F., & Zhou, B. (2016). Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL), 2016, 214-223.



**|| Volume 12, Issue 5, May 2023 ||**

**| DOI:10.15680/IJIRSET.2023.1205354 |**

5. See, A., Liu, P. J., & Manning, C. D. (2017). Get to the point: Summarization with pointer-generator networks. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL), 2017, 1073-1083.
6. Lin, C. Y. (2004). ROUGE: A package for automatic evaluation of summaries. Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004), 74-81.
7. Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: A method for automatic evaluation of machine translation. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), 2002, 311-318.
8. Barzilay, R., & Lapata, M. (2005). Modeling local coherence: An entity-based approach. Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL), 2005, 141-148.
9. Conroy, J. M., Schlesinger, J. D., & O'Leary, D. P. (2007). Topic-focused summarization of chat conversations. Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL), 2007, 1-8.
10. Das, A., Datar, M., & Garg, A. (2007). Google news personalization: Scalable online collaborative filtering. Proceedings of the 16th International World Wide Web Conference (WWW), 2007, 271-280.



Impact Factor: 8.423



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details