



**IJIRCCCE**

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 8.379**



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

# Cross-Cultural Lexical Distillation System

Dr. G. Maria Priscilla, Ashvanth.V.N

Associate Professor & Head, PG & Research Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore, India

UG-Student, PG & Research Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore, India

**ABSTRACT:** This project “Cross-cultural lexical distillation system” deals with the Natural Language Processing (NLP) which summarizes the large document into a shorter version of the text. This automatic extractive text summarization is done by using Extraction based text summarization which involves selecting sentences of high relevance (rank) from the document based on word and sentence features and put them together to generate a summary. It summarizes the text faster and effectively as this algorithm uses Graph-based sentences ranking. In this new era, where remarkable information exists on the Internet, it is most important to provide an improved mechanism to extract the information quickly and most efficiently. There is a plenty of text material available on the Internet. So there is a setback of searching for significant documents from the number of documents available and absorbing appropriate information from it. To solve the above two problems, the text summarization is very much necessary. This summarizer can reduce the paperwork and time, as the summarized text is converted into audio format and even share the summary as they wish by just one tap. The final performance of the application is validated using ROUGEN value which gives precision value of 0.602 which is better than any other extraction based text summarization.

**KEYWORDS:** Cross-cultural lexical distillation system, Natural Language Processing (NLP), Extraction-based text summarization, Graph-based sentences ranking, ROUGEN value

## I. INTRODUCTION

An Extractive Text Summarizer for Multilingual Purpose is the GUI based text summarization that uses the Text Rank algorithm to summarize the massive texts into a straight meaningful summary for multiple languages. The significance of Natural Language Processing (NLP) is ruling the AI industry for making humans to perform complex tasks with ease. As humans should not waste much time in reading the entire large text to recognize the subject of the larger document, just by one tap they could save their valuable time. In today’s world, the Internet is the platform where all the study materials, news articles, even textbooks, etc., are searched. But the problem arises when grasping relevant information from the massive materials on the internet. So this application will provide the best-consolidated results will be given as a summary and extracts keywords for making humans understand the shortest version of the articles or text that describes the subject. This summarizer automatically summarizes the larger texts of any language and converted into audio for reducing the time for reading the subject. And also the summary can be easily shared by anyone by just scanning the QR code of the summary. The Complete application is shown in appendix.

## II. RELATED WORK

The explosion of textual data online has fueled the development of multilingual text summarization systems, providing a foundation for your project. Existing solutions offer a range of functionalities, but some limitations exist. TEE (Text Summarization Engine) by Google AI utilizes machine learning for summaries in various languages, but its user interface might lack customization options. DoSummarizer, another online platform, focuses on extractive summarization for over 20 languages, though it may not have features like audio conversion or QR code generation.

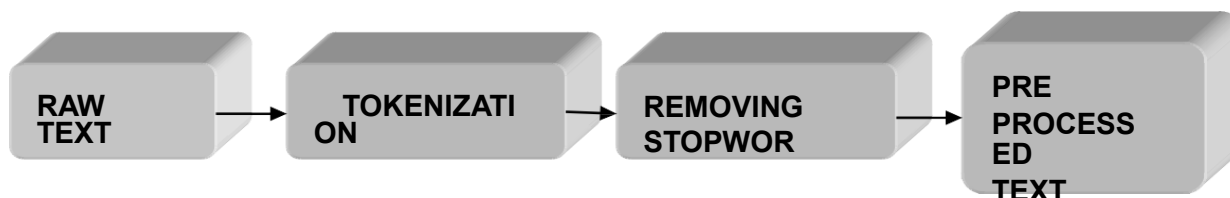
For developers, open-source libraries like Martης (Mars) offer multilingual summarization using various algorithms, including TextRank. While Martης allows more control, it necessitates coding knowledge for integration into a user interface. Similarly, Sumy, a Python library, provides various summarization algorithms with multilingual support, but requires coding expertise for practical use.

Commercial options like Amazon Comprehend and Summarization.com offer powerful multilingual summarization APIs, but often come with subscription fees. While these systems showcase advancements, your project, the Extractive Text Summarizer for Multilingual Purpose, has the potential to carve its own niche by offering a unique combination of features.

### III. PROPOSED METHODOLOGY

In today's digital age, information reigns supreme. However, the sheer volume of textual data online – research papers, news articles, and educational materials – can be overwhelming. Sifting through lengthy documents to grasp key concepts can be time-consuming and frustrating. Traditional methods like skimming or highlighting may not be effective for large texts. This is where the Extractive Text Summarizer for Multilingual Purpose steps in. Imagine a user-friendly application that tackles information overload by automatically generating concise and informative summaries of large texts. This application leverages the power of Natural Language Processing (NLP) and the TextRank algorithm to identify the most relevant sentences within a document, ensuring a factually accurate representation of the original content.

The multilingual capabilities cater to a diverse user base, allowing summaries to be generated for a wide range of languages. This makes the application a valuable tool for students, researchers, professionals, and anyone seeking to navigate the vast ocean of textual information in a globalized world. But the functionalities don't stop there. Recognizing the value of time and convenience, the application offers an audio output option. This allows users to consume the summarized information hands-free while commuting, exercising, or completing other tasks. This caters to auditory learners and those who prefer a hands-free approach to information consumption.



### IV. DATAFLOW DIAGRAM





V. RESULT

To evaluate the results the system generated summary is cross- validated with the human-generated summary. The humangenerated summaries are taken as reference summaries. ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is the package for evaluating the automatically generated summary and gives the precision and recall values. These values will range from 0 to 1. Higher the Precision higher the accuracy of the summary. The precision is the fraction of the total number of overlapping instances among the system generated summary, where recall is the fraction of the total number of overlapping instances among the reference summary. Using this precision and recall, F-score is generated which is also known as Dice similarity coefficient (DSC) by using the following formula,

$$F_{Score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

ROUGE – N	PRECISION	F –SCORE	RECALL
ROUGE – 1	0.602	0.517	0.452
ROUGE – 2	0.275	0.236	0.206
ROUGE – L	0.500	0.457	0.421

The precision (0.602) and recall (0.452) obtained is higher than any other text extractive summarization algorithm. Thus, the usage of text rank algorithm in the system application works efficiently and gives more accurate summaries. Hence this system application generates a summary more likely human-generated summary.

VI. CONCLUSION

The GUI-based Automatic extractive Text summarization was designed and coded which emerges trends in education domains, biomedicine, product review, emails and blogs. This is because there is information overload in these areas, especially on the World Wide Web. Automated summarization is an important area in NLP research. The purpose of extractive document summarization is to automatically select several indicative sentences, passages, or paragraphs from the original document. The objective of text summarization is to present the source text into an appropriate version with semantics. The primary plus of using summarization is to reduce reading time. Most summarization techniques are based on extractive methods. Abstractive summarization requires complex code for language generation and it is difficult to make a model into the domain-specific areas. So using the extractive method of summarization, the Text rank algorithm made the summarizer application perform faster than any other extractive method and even it has the feature of converting text into audio and also into QR code for making efficient sharing of content to others by just scanning the QR code. The sentences that are highly relevant to other sentences in the source text are likely to be more informative for the given text, and will be therefore given a higher score as being useful for the overall understanding of the text. As a future enhancement, a simple neural network will be modeled for ranking sentences through encoding and decoding, which can make the application system to perform as like as human beings.

REFERENCES

1. Alvee, R. Fahim, M.D. Ramkrishna, S. Ruhit, R. and Hossain, A. ‘Bengali Text Summarization using TextRank, Fuzzy C- Means and Aggregate Scoring methods’, IEEE Region 10 Symposium (TENSYP), pp. 331-336. (2019)
2. Akshi, K. Aditi, S. Sidhant, S. and Shashwat, K. ‘Performance Analysis of Keyword Extraction Algorithms Assessing Extractive Text Summarization’, International Conference on Computer, Communications and Electronics, pp. 408 – 414. (2017)
3. Devika, R. and Subramaniyaswamy, V. ‘A semantic graphbased keyword extraction model using ranking method on big social data’, Wireless Networks, Springer, pp. 1-13. (2019)



4. Dijana, K. and Vlado, K. 'Automatic Text Summarization of News Articles in Serbian Language', 18th International Symposium INFOTEH-JAHORINA(INFOTEH), pp. 1 – 6. (2019)
5. Madhuri, J.N. and Ganesh, K.R. 'Extractive Text Summarization Using Sentence Ranking' International Conference on Data Science and Communication (IconDSC), pp. 1-3. (2019)
6. [Mona, T.K. Mohan, S.S. H S S S Raviteja, D. Sai, K.P.V.](#) 'Smart Summarizer for Blind People', 3rd International Conference on Inventive Computation Technologies (ICICT) , pp. 15-18. (2018)
7. Mudasir, M. Rafiya, J. and Muzaffar, S. 'Text Document Summarization using Word Embedding', Expert Systems with Applications, Vol.143, No.1, p. 112958. (2019)
8. Reda, E. Gamal, B. and Akram, E.K. 'Extractive Arabic Text Summarization Using Modified Page Rank Algorithm', Egyptian Informatics Journal, Vol.21, No.2, pp. 73-81. (2019)
9. Suzan, V. Emiel, K. Sander, W. and Antal, V.B . 'Query- based summarization of discussion threads', Natural Language Engineering, Vol.26, No.1, pp. 3-29. (2019)
10. Umadevi, K.S , Romansha, C. Nivedita, S. Likitha, A. and Jagadeesh, K.R 'Text Summarization of Spanish Documents', International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1793-1797. (2018)
11. Yogesh, K.M. Dinesh, G 'Analysis of Sentence Scoring Methods for Extractive Automatic Text Smmarization', International Conference on Information and Communication Technology for Competitive Strategies (ICTCS), pp. 1-6. (2014)



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

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



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

Scan to save the contact details