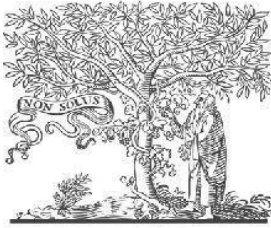


COPY RIGHT



ELSEVIER
SSRN

2024 IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper; all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 8th Aug 2024. Link

<https://ijiemr.org/downloads.php?vol=Volume-13&issue=issue08>

DOI: 10.48047/IJIEMR/V13/ISSUE 08/7

Title Optimizing Telugu Language Summarization Through Algorithms

Volume 13, ISSUE 08, Pages: 50 - 58

Paper Authors

Dr. G.V. Ramesh Babu , Dasari Uday Kiran



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper as Per **UGC Guidelines** We Are Providing A Electronic Bar code

Optimizing Telugu Language Summarization Through Algorithms

Dr. G.V. Ramesh Babu

Associate Professor, Department of Computer Science, Sri Venkateswara University, Tirupati
gvrameshbabu74@gmail.com

Dasari Uday Kiran

Master of Computer Applications, Sri Venkateswara University, Tirupati.
udaykirandasari1@gmail.com

Abstract

In this work, we the summary of the text has been clearly stated. Reading a big document takes time, at the same time, if we summarize the same document, it is easy to understand and saves time. Here we show a text summary in Telugu. First, documents must go through a preprocessing process, including tokenization, stop word removal, radicalization, and N-gram analysis. After that, it uses Histo-Fuzzy C-means clustering to achieve the clustering, and a weight-based sentence ranking technique. Finally, the sentences are combined into clear and concise summaries using the median support-based Grasshopper Optimization Algorithm (MSGOA). The performance of this strategy is evaluated using an online research dataset. Compared to previous text synthesis methods, the proposed method outperforms them. Compared to the existing accuracy, the proposed method surpasses and achieves an accuracy of 84%.

Introduction

Text summarization is the process of reducing a text file to those parts that convey the main meaning of the text. It is difficult for users to find the main points in a large document. Text summarization methods can be divided into two types, namely

1) extractive summarization

2) abstract summarization Here we only use extractive summarization. The purpose of the summary is to cover sets of sentences and summarize key points. TS is a modern solution to classification problems in machine learning. This extractive to approach tries to understand the importance of individual sentences in a document and selects the most important ones. Why since machines do not create sentences and form sentences, why is it not involved, the grammar is different from each language. Determining rank is an ML problem best handled by supervised learning. The government and the public

have turned to Twitter and Sinan Weibo for information on many issues. However, with the rapid expansion of social media platforms, consumers are finding it increasingly difficult to immediately find the information they need. Extractive summarization involves finding important phrases in documents and summarizing them in a short form. Telugu is also one of the most popular regional languages on the internet. The problem now is that the data has a regional language, but the data is very large, and sometimes there may not be much time to read and understand. Not only is the data sometimes irrelevant to us, but we will know it after reading it. precious time. To avoid this, it is useful to provide a summary of the textual data in the regional language. The Internet now allows users to access large amounts of data. As popularity increases, it becomes more and more difficult to extract the most important information from such a large amount of data. In the case of textual documents, collecting and understanding

raw information from many sources is a complicated and time-consuming process for people who are short on time. Information retrieval algorithms have automated these tasks for decades. However, as the amount of data increases, several performance issues arise, including insufficient solutions and unmanageable applications for information retrieval tasks. Using high-tech machinery can help mitigate the damage caused by these issues. However, the price may be higher. Reducing the dimensionality of raw data may be a more appropriate choice to address these challenges and speed up these tasks. Enumerations are one of the most important applications in the computing world.

Text summarization is the process of producing a simplified version of a single document or set of documents. Automatically summarizing text documents is a difficult process because the generated summaries must contain as much of the original document as possible. The text summary saves us time and effort by giving us a quick summary of all content. Initially, these text summaries were handwritten, but as data volumes increased and automation became more common, automatic summarization methods became increasingly important. Depending on their value in the extraction enumeration, certain sentences are immediately selected and included in the enumeration. Abstract summaries, on the other hand, adapt or paraphrase terms so that they have the same meaning as the original text. Extraction summarization has been performed on many Telugu documents.

TS is a modern solution to classification problems in machine learning. This extractive ts approach tries to understand the importance of individual sentences in a document and selects the most important ones. Why because machines do not create sentences and do not form sentences. Why it does not imply that syntax varies for each language. Extractive ts has steps like ranking sentences based on score and sentence selection. According to the ranking, the top

n will select the most sentences to summarize. This process is inevitably associated with classification challenges in the area of ML. Rank determination is an ML problem best handled by supervised learning. The government and the public have turned to Twitter and Sina Weibo for information on many issues. However, with the rapid expansion of social media platforms, consumers are finding it increasingly difficult to immediately find the information they need. For online search users, the large amount of data is likely to cause problems.

Query-based summarization is another well-known automatic text summarization method. This method uses the input query to calculate a summary of the text content. The approach focuses on the user query parameters and thus on the user queries provided by the user to the system. Text summarization is used to shorten input documents while preserving their overall meaning and informational value. Thus, text summarization is the act of reducing data so that users can assimilate it more quickly.

The main contribution of this paper is a preprocessing step using tokenization, stop word removal, voting, and N-gram methods. Ratings include Histo Fuzzy C-means clustering and sentence ranking. Phrase selection using MSGOA sorting. Finally, these sentences are compiled into a short and informative summary. The rest of the article is as follows: Section 3 explains text summarization techniques in Telugu well. Section 2 includes an overview of similar studies, while Section 3 presents the proposed method for text synthesis in Telugu. Publish the results in Section 4 and complete the work in Section

Literature Review

[1] Reddy Naidu et al present a method for automatically extracting text summary keywords from of a Telugu email newsletter data set. The technique of compressing textual material into a summary that retains important ideas is called summarization. The extractive summary uses the information they get and extracts

the sentences that best reflect the hidden information. Most extractive synthesis techniques are based on the concept of keyword detection, and the extraction is usually performed by extracting related words that appear more frequently than other words, emphasizing the most more important. Manually extracting and annotating keywords is a time-consuming and error-prone process.

[2] Yan Du et al. A new automatic news text summary model with fuzzy logic, multi-features and genetic algorithm is proposed. Word features are the most requested features.

The keywords are selected from the extracted words with a higher score than the predefined words. They applied a fuzzy logic framework to calculate the final score. Since news text is a single type of text with several different place and time aspects, it is sometimes possible to detect these different news items as keywords. Each feature is weighted using a genetic algorithm. Linear combinations reveal the meaning of each text.

[3] Angel Hernandez-Castaneda et al propose a technique to improve keyword recognition that uses semantic information from ATS. The Automatic Text Synthesis (ATS) task is to synthesize a document to create an abbreviated version. Summarizing forces you to choose not only the main ideas of the sentences but also their key relationships. Related Works uses a ranking system to choose which units of text (mostly sentences) should be included in the summary. However, as important information may have been omitted, the resulting summary may not cover all the topics discussed in the source text. By grouping sentences to locate major themes in the original manuscript, the technique increases coverage and accuracy.

[4] Rana Alqaisi et al present an automatic, general and extractive method to summarize Arabic documents. Due to the increasing use of the Internet and social media, a large amount of textual material is now available online. This online textual data leads to abundance

and redundancy. Reducing information redundancy and saving time is crucial when reading text data online. Therefore, there is a constant need for an automatic text summarization system capable of extracting relevant and salient information from a set of texts on the same or related topics. In the proposed system, a multi-objective optimization method based on clustering and evolution is applied. Clustering-based methods identify the most important topics in the text, while scalable multi-objective optimization methods prioritize three objectives: coverage, diversity/redundancy, and relevance.

[5] Fu Chengyou et al. The use of Refined BERT presents the merging of topic information and semantic relevance for text summarization with Refined BERT (TIF-SR). A high-quality summary system should focus on the document and the similarity between the summary and the source document. They extract the keywords from the topic and integrate them into the source documents as part of the input, because the topic information is very important in the executive summary [27-32]. Second, the semantic similarity between the generated summary and the original material is calculated to improve the quality of the summary.

Proposed Methodology

Reading many long documents is a long and difficult process. The summary of the same document gives an overview of the content. For imported documents, summaries can be generated. This work presents a text summarization in Telugu using a new hybrid approach combining fuzzy organizational enthalpy C-means clustering and median-support based Grasshopper Optimization Algorithm (MSGOA). Initially, input documents are fed into the preprocessing stage for text summarization. After merging all the papers, process the output document. Tokenization, stop word removal, voting, and N-grams are some of the features available. Sentences are sorted by assigning weights and sorted by their weights in the Histogram Fuzzy C-means clustering and sentence ranking. Phrases with high rankings are extracted from input documents, resulting in document-quality summaries. The MSGOA ranking is used to select

the phrases that appear in the summary. Select the phrase with the highest rank. **Preprocessing**

Text summary development requires preprocessing of the input document (D) first to detect plain text reports. Tokenization, stop word removal, voting, and N-grams are all examples of basic preprocessing.

Stop word removal

Most frequently used words such as 'the', 'a', 'and' and 'this' are removed from the textual data at this stage, these words have no semantic information. They are inefficient when it comes to classifying documents. Therefore, they will have to leave. On the other hand, creating stop-word lists is difficult and not consistent across text sources. This approach also reduces the amount of textual data and increases the efficiency of the system. These words appear in every text document, although they are not required for text mining applications.

Stemming of Telugu language

Agreement in Telugu Agreement in Telugu is somewhat difficult. By removing suffixes and prefixes from words, it transforms them into the basic form of English. The phrases:

['రాయటం', 'రాయడం', 'వ్రాయటం', 'రాసిన', 'వ్రాసిన']

can all be reduced to a single symbol Not only do these phrases have so many variations in Telugu.

'వ్రాయడం'.

N-Gram stemmers

Adamson and Boreham proposed this technique. The shared graph approach is what is known as. Several letters are shown in the figure. Single shared graphs are used to calculate measures of association between pairs of terms in this method. To illustrate: studying and learning are two words that come to mind. le earnni in ng learning learner le earn ne ererer

In this example, learn contains six separate graphs, while learner has five; the two words share three different patterns: le, ea, and rafter determining the number of different graphs, the Dice coefficient is used to calculate the similarity measure based on the different graphs. [26], which is the definition of the dice coefficient.

$$S=2C/ (A+B)$$

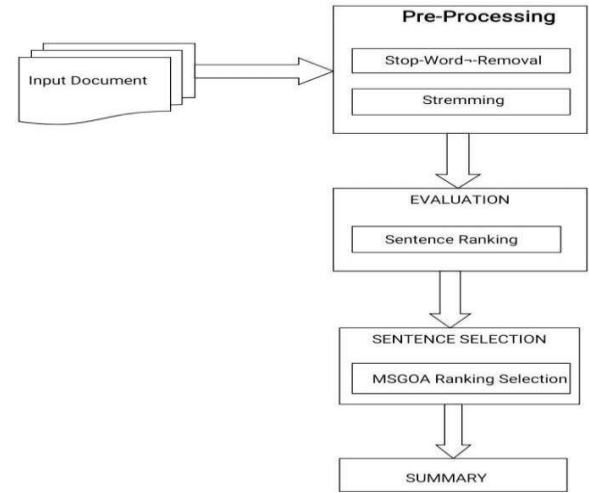


Figure 1: MSGOA Sentence selection based block diagram

Preprocessing

The method of pre-processing is taking the input as raw data and giving output as text data

We require some techniques to do the summarization that are Tokenization, stop word removal ,stemming and N-gram these are require for text summarization.

Tokenization

Tokenization is the symbols and some other meaningful items and these tokenization are Used to break down the sentence orA paragraph text into words to select a relevant sentence it requires the weight of the paragraph text.

Stop word removal

This method also the decrease the amount of text data and boost the system efficiency, here the main important some common words are deleted from the text data

EX: A, THE, AND, THAT, THIS, THEY

Stemming of Telugu language

It removes the prefix and suffix words and translate them into basic English language

N-Gram Stemmers

Adamson and Borcham this both are come up with one technique the above diagram method is called pair of letters the above diagram is used to calculate pair of terms in this method.

In general example eating and eaten are both words comes to mind by using one formula we can find that is
 $S=2C/(A+B)$

Based on Fuzzy C-means clustering

Here we are giving the rank to words and same rank will be replaced to words. When words are repeated, rank also repeated.

$$H = \sum_{i=1}^n H_d$$

Sentence Ranking

It will give ranking to the sentence and the sentence will be repeated same ranking will be given. The weight of the sentence will be calculated by using (1) and (2).

$$C_{wht} = 10 * (5 * ((n-1)/N)) \quad (1)$$

$$S_{wht} = \sum M_m = 0 C_{wht} / M \quad (2)$$

Median Support Value Based Grasshopper Optimization Algorithm

It select the high ranking sentence and removing the unimportant sentence or words, after removing unimportant sentence it will make it as group

$$S = (p+Q) / P * Q$$

Distance is calculated as $d_{ij} = |X_j - X_i|$

Algorithm

- Step 1: Taking input from the input document
- Step 2: creating a pre-processing
- Step 3: un-wanted words will be removed by using stop-word-Removal
- Step 4: Here same ranking will be given to same words by using stemming
- Step 5: After step-4 Here unwanted words will be removed by using sentence ranking and grouped the remaining word
- Step-6 By using MSGOA algorithm high rank words are selected

Results

The proposed methodology approach deals with Telugu text summarization based on machine learning .here we are used to algorithms Fuzzy-c, Median support value based Grasshopper Optimization Algorithm by using these two we calculated the performance of the proposed approach is evaluated .

Here all algorithms are implemented in java language and executed on a core i5 processor, 2.1MHZ, 4GB RAM

Performances measures

The statistical metrics of accuracy, F-measure recall and Precision can be implemented.

Accuracy

Accuracy by using this we can find the correctness of our work and it mentioned the number of correction predictions/the total number of predictions .Number of correct predictions (TN+TP). Total no. of Predictions (TN+TP+FN+FP)

$$\text{Accuracy} = \frac{(TN+TP)}{(TN+TP+FN+FP)}$$

Where, TN is true negative, TP is the true positive, FP is the false positive, and FN is the false negative.

Recall

Precision is a metric for how relevant the extracted phrases are, whereas recall is a metric for how many truly relevant results are produced when utilizing equations.

$$\text{Re call} = \frac{\text{Extracted Summary Provided Summary}}{\text{Provided Summary}}$$

Precision

Precision is the proportion of the predicted positive instances that were correct text size, as calculated using equation.

$$\text{Precision} = \frac{\text{Extracted Summary Provide Summary}}{\text{Extracted Summary}}$$

F-measure

It is a measure of a test's accuracy. Maintain a balanced state among the Recall and Precision is given in equation,

$$F_{\text{measure}} = 2 \text{Precision} * \text{Recall} / \text{Precision} + \text{Recall}$$

Identifying verbs

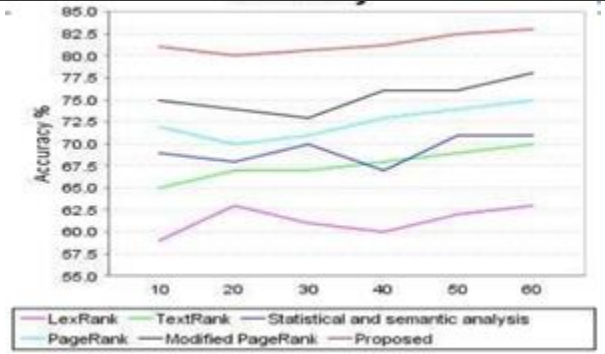
S.NO	sentences	verbs
1.	నేనుఉరులోనుండివెళ్ళివేరేరాష్ట్రంలో నివాసిస్తునాను	నివాసిస్తునాను
2.	నేనుబాగుణను	బాగుణను
3.	నాకుసహాయంచేసినందుకుధన్యవధములు	ధనివధములు
4.	సాధారణంగారైతులుతమస్వంత భూమిలోనేసాగుచేస్తారు	సాగుచేస్తారు'
5.	మేముపరిగెడుతునము	పరిగెడుతునము
6.	మీరుఎన్నిమాట్లాడతారు	మాట్లాడతారు
7.	మీరుతెలుగులోబాగావ్రాయగలరు అనుకుంటా	వ్రాయగలరు
8.	కళాకారుడుతనపెయింటింగ్గురించివివారిస్తునాడు	వివారిస్తునాడు
9.	ఆకుక్కమనిషికాలనుకొరరికింద	కొరరికింది
10.	కవితపాటపాడుతుంటేవినడానికిచాలాబాగుంది	పాడుతుంటే
11.	మాస్నేహితురాలుకొత్తగాబట్టలువ్యాపారంప్రారంబించారు	ప్రారంబించారు
12.	పండగరోజుదేవునియొద్దవార్తానముతీసుకుంటారు	వార్తానము
13.	నగుండెవేగంగాకొట్టుకుంటుంది	కొట్టుకుంటుంది
14.	రాజితకిపెయింటింగ్చేయడంచాలాఇష్టం	పెయింటింగ్
15.	ఉపాధ్యాయులుపిల్లలనిప్రతివిషయములోప్రోత్సహించారు	ప్రోత్సహించాలి
16.	ఆఅమ్మాయిఆకలితోవుంటుందిరోజు	ఆకలితో
17.	మాకాలేజీలోపిక్సిక్కిఎప్పుడుతీసుకువెళ్లలోమాస్టర్నిర్ణయిస్తారు	నిర్ణయిస్తారు
18.	జైలులోకైధిలుపారిపోయారు	పారిపోయారు
19.	పిల్లవాడుతల్లినిచంపేశాడు	చంపేశాడు
20.	నదైర్యంమాతల్లితండ్రులు	తల్లితండ్రులు
21.	మాతముడుచాలాపిరికివాడు	పిరికివాడు
22.	మాచెల్లిపరీక్షలలోఫస్కాస్వచ్చింది	పరీక్షలలో
23.	నేనుచైన్లోబాగానిద్రపోయాను	నిద్రపోయాను
24.	మామాస్టరునాకురామయననిబోధించారు	బోధించారు
25.	మాతముడుకష్టపడిఆటలోమొదటిస్థానంసంపాదించారు	సంపాదించారు

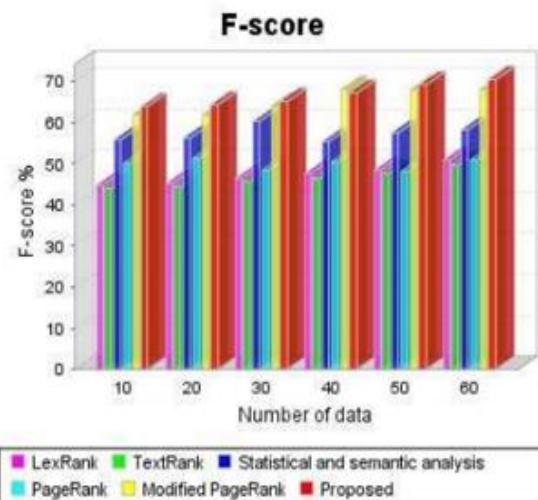
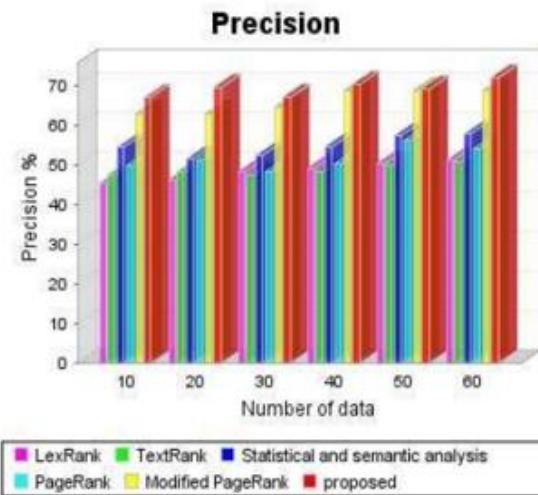
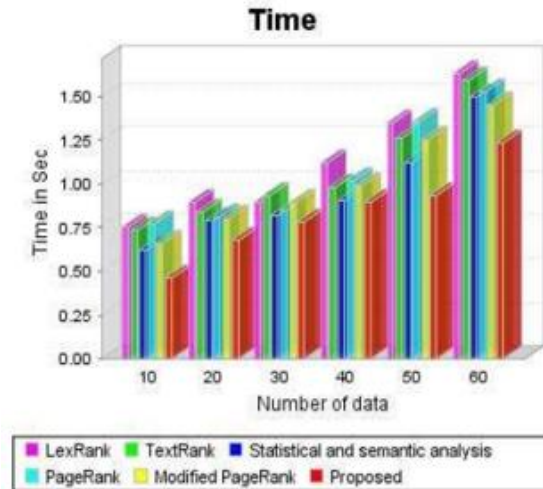
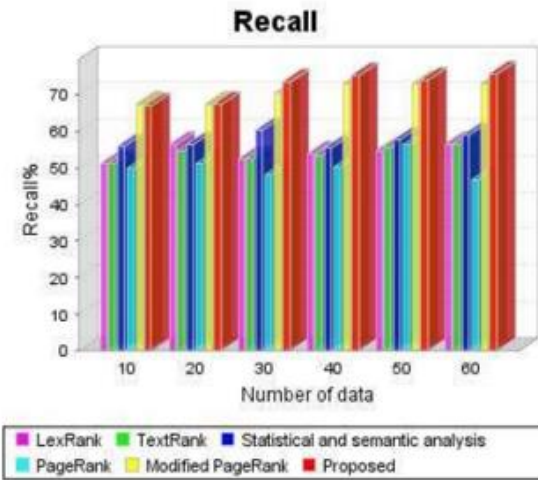
```

Final Project
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
No of Words after removing stop words=136
No of Stop Words= 4
{'భారీ': 1, 'పెత్తన': 1, 'మొండిబకాయిలు': 1, 'పెరిగిపోవడంతో': 1, 'బడిబిబ': 5, 'వనవనాలపై': 1, 'భారత': 1, 'లిబియ': 1, 'భ్యాంకు': 2, '(': 2,
max_score2.75
-----summary-----
గతంలో ఓవర్సీస్ భ్యాంక్, యునైటెడ్ బ్యాంకుకు సంబంధించి కూడా ఆరేబిబ పీసీఎ చేపట్టింది.
1
No of Words after removing stop words=147
No of Stop Words= 10
{'నన్నాడిల్లి': 1, '': 1, 'అర్కి': 2, 'మండ్రి': 3, 'అరుణి200c': 1, 'జైజీ': 3, 'కొమ్మవారం': 1, 'ప్రభుత్వ': 1, 'రంగి': 1, 'భ్యాంకుల': 1, '(': 3, 'మీ
max_score2.1538461538461537
-----summary-----
ఈ ఆర్కి సంవత్సరం పీసీఎ బిల చీఫ్ అట్ జైజీ బిరుదును తొలి సమావేశం ఇదే.
2
No of Words after removing stop words=50
No of Stop Words= 0
{'కీటక్': 1, '': 1, 'ఇంగ్లండ్200c': 1, 'బిరుదును': 1, 'సెకండ్': 1, 'వన్': 1, 'మ్యాచ్200c': 1, 'భారత': 1, 'బిబ్బ': 1, 'మూడో': 1, 'వికెట్
max_score1.5555555555555556
-----summary-----
27వ ఓవర్ మొదటి బంతికి క్లీన్ బౌల్ చేశారు. మ్యాచ్ గెలిచేందుకు ఇంగ్లండ్ ఇంకా 299 పరుగులు చేయాల్సి ఉంది.
3
No of Words after removing stop words=66
No of Stop Words= 0
{'ఇస్లామాబాద్': 1, '': 1, 'పాకిస్తాన్': 3, 'అంతర్జాతీయ': 2, 'ఉగ్రవాది': 1, 'పాపీజ్': 1, 'సయ్యాద్200c': 1, 'స్వేచ్ఛనిచ్చి': 1, '': 1, 'భారతదేశానికి':
max_score1.7777777777777777
-----summary-----
కళ్ళిరు అంపై అంతర్జాతీయ న్యాయస్థానాన్ని ఆశ్రయిస్తామని పాకిస్తాన్ ప్రకటించింది. కళ్ళిరులో నిత్యం మానవ హక్కుల ఉల్లంఘన జరుగుతోంది అంటుంది.
4
No of Words after removing stop words=41
No of Stop Words= 3
    
```

```

Final Project
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
Add text cell
పదేశాల్లో కలుసుకున్నప్పుడు అంగంలోనే మళ్ళాడుకోవాలని సామాజిక మార్గమాల ద్వారా చెప్పకుండున్నారు. బహిరంగ ప్రదేశాల్లో
7
No of Words after removing stop words=215
No of Stop Words= 11
{'అడు': 2, 'క్రికెట్200c': 3, '': 9, 'ఇబు': 2, 'హాకీ': 7, 'రిండింటిలో': 1, 'దాయకులు': 1, 'సమరంస్కాధిల్లి': 1, '': 1, 'అదివారం': 3, 'ఇతర'
max_score3.6666666666666665
-----summary-----
అటల్ పై అమీతానక్కి ఉ న్న వాల్లెతే మరీను..!
8
No of Words after removing stop words=1525
No of Stop Words= 120
{'నిర్మాణ': 1, 'సంస్థ': 3, '': 30, 'డిజిటల్': 3, 'ఇండియా': 3, 'లిమిటెడ్200c': 3, '': 87, 'త్రావర్': 6, 'ఫిలిమ్మార్గా200c': 3, 'పెన్200c:
max_score46.22222222222222
-----summary-----
కెమెరాఫోన్లను మెప్పించారు. కెమెరాఫోన్లను మెప్పించారు. కెమెరాఫోన్లను మెప్పించారు.
9
No of Words after removing stop words=41
No of Stop Words= 1
{'స్కాథిల్లి': 1, '': 1, 'బిధిర': 1, 'యువతి': 1, 'గీతను': 2, 'తల్లిదండ్రుల': 1, 'చెంతకు': 1, 'చేర్చిన': 1, 'రూ.లక్ష': 1, 'నగదు': 1, 'అందజేస్తామని
max_score1.2
-----summary-----
భారతీకు చెందిన గీతను పదిహేనేళ్ల క్రితం లాహారలో ఆగి ఉన్న సంద్యుల ఎక్స్‌ప్రెస్‌లో పాకిస్తాన్ సైనికులు గుర్తించారు.
10
    
```





Conclusion

Deleting words and votes using fuzzy C-means combined with grouping method based on histo-enthalpy. After ranking the selection text optimization, the technique used is based on the Grasshopper Supported Median Optimization Algorithm (MSGOA) to obtain summaries successfully. The results are finally obtained with the proposed method to accurately predict the text summary. The results show that the Grasshopper optimization algorithm (GOA) based on the median support value produces better results as the number of iterations increases. Based on previous research, the results of this survey were better than others. The final accuracy rate for this study was 84%.

References

- [1] Nawaz A, Bakhtyar M, Baber J, Ullah I, Noor W, Basit A. Extractive Text Summarization Models for Urdu Language. Information Processing & Management. 2020 Nov 1; 57(6):102383.
- [2] Mutlu B, Sezer EA, Akcayol MA. Candidate sentence selection for extractive text summarization. Information Processing & Management. 2020 Nov 1; 57(6):102359.
- [3] Tanwi, Satanik Ghosh, Viprav Kumar, Yashika S Jain, Mr. Avinash B, "Automatic Text Summarization using Text Rank",

(IRJET), Volume: 06 Issue: 04, April 2019, e-ISSN: 2395-0056.

[4] K. U. Manjari, S. Rousha, D. Sumanth and J. Sirisha Devi, "Extractive Text Summarization from Web pages using Selenium and TF-IDF algorithm," 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), Tirunelveli, India, 2020, pp. 648-652, doi: 10.1109/ICOEI48184.2020.9142938.

[5] Sana Shashikanth, Sriram Sanghavi, "Text Summarization Techniques Survey on Telugu and Foreign Languages", International Journal of Research in Engineering, Science and Management, Volume-2, Issue-1, January-2019. ISSN: 2581- 5792.

[6] Rumagit, R. Y., Setyawati, N., & Bangkalang, D. H. (2019). Comparison of graph-based and term weighting method for automatic summarization of online news. *Procedia Computer Science*, 157, 663–672.

[7] Qaroush, A., Farha, I. A., Ghanem, W., Washaha, M., & Maali, E. (2019). An efficient single document Arabic text summarization using a combination of statistical and semantic features. *Journal of King Saud University-Computer and Information Sciences*.

[8] Noor, F., Bakhtyar, M., & Baber, J. (2019). Sentiment analysis in e-commerce using svm on roman urdu text. *International conference for emerging technologies in computing*. Springer 213–222.

[9] Miao, L., Cao, D., Li, J., & Guan, W. (2020). Multi-modal product title compression. *Information Processing & Management*, 57(1), 102123. <https://doi.org/10.1016/j.ipm.2019>.

[10] Mutlu, B., Sezer, E. A., & Akcayol, M. A. (2019). Multi-document extractive text summarization: A comparative assessment on features. *Knowledge-Based Systems*, 183, 104848.

[11] Mutlu, B., Sezer, E. A., & Akcayol, M. A. (2020). Investigation of text summarization features by fuzzy systems and convolutional neural networks. Submitted: *Semantics conference 2020*.

[12] Joshi, A., Fidalgo, E., Alegre, E., & Fernández-Robles, L. (2019). Summccoder: An unsupervised framework for extractive text summarization based on deep auto-

encoders. *Expert Systems with Applications*, 129, 200–215.

[13] Nasar, Z., Jaffry, S. W., & Malik, M. K. (2019). Textual keyword extraction and summarization: State-of-the-art. *Information Processing & Management*, 56(6), 102088. <https://doi.org/10.1016/j.ipm.2019.102088>.

[14] Yasunaga, M., Kasai, J., Zhang, R., Fabbri, A., Li, I., Friedman, D., & Radev, D. (2019). ScisummNet: A large annotated corpus and content-impact models for scientific paper summarization with citation networks. *Proceedings of AAAI 2019*