



V P Dr. P. G. Halakatti College of Engineering and Technology, Vijayapura

IEEE NKCon 2022

Kannada Text Summarization

Name : Isha Ankalikar
Affiliation: CSE Department
BLDEACET
Email:ishaankalikar@gmail.com

Name : Kruthika Pujari
Affiliation : CSE Department
BLDEACET
Email:kruthikamp31@gmail.com

INTRODUCTION

Automatic text summarization enables us to access the most important content in the shortest possible time. In this text, a summarization technique is proposed by extracting the most informative sentences.

To this end, a number of features are extracted from the sentences, each of which evaluates the importance of the sentences from an aspect. This project has shown that the use of more suitable features leads to improved summaries generated, in order to evaluate the automatic generated summaries.

The planned summarization system considers 5 options to work out the necessary sentences within the document. The options used are, term frequency, inverse sentence frequency, keyword feature, sentence length, and sentence position.

RESULTS

Kannada text is taken as input on which following procedure is carried out:

1. Input text is tokenized into words.
2. Stop words are excluded.
3. Root word is extracted.
4. Frequency of words are calculated.
5. Sentencing score is assigned to each sentence based on frequency of words.
6. Sentence rank is allocated based on which it is selected as a part of summary.

Percentage of meaning between system summary V/S manual summary is been calculated by comparing the summary that is been generated through the experts of kannada literature (professor's) and system generated summary. The same kannada expert verify the system summary to the summary generated by them, thus they come up with the percentage of match between both the summaries which is called as percentage of meaning between system V/S manual summary.

DISCUSSION / FUTURE SCOPE

Initially the data is taken from the input text then it is tokenized into individual words. Words may contain the some meaning less words which are been removed using stop words removal method. The next process is to extract the root word of the given input word which is done through stemming. All the above steps clean the input text so the actual calculation process get executed. Those are term frequencies and keywords. finally the words/sentences need to be placed in correct place this is done through sentence length and positioning, sentence scoring and ranking.

APPLICATIONS :

- Notable social media platforms use this process to generate summaries for posts that are grouped based on the content called topics.
- These topics are used to engage users online. Google's home feed for example generates summaries based on the user's preferences.
- Search engines today directly answer the provided query rather than just providing links.
- Text is extracted from ranked and credible websites and the summary is generated for this text which is returned as an answer.

METHODOLOGY

In the process of Kannada document summarization, some pre-processing operations are carried out before the sentence scoring algorithm is executed. The pre-processing function prepares the document for the ranking of sentences and the generation of summary

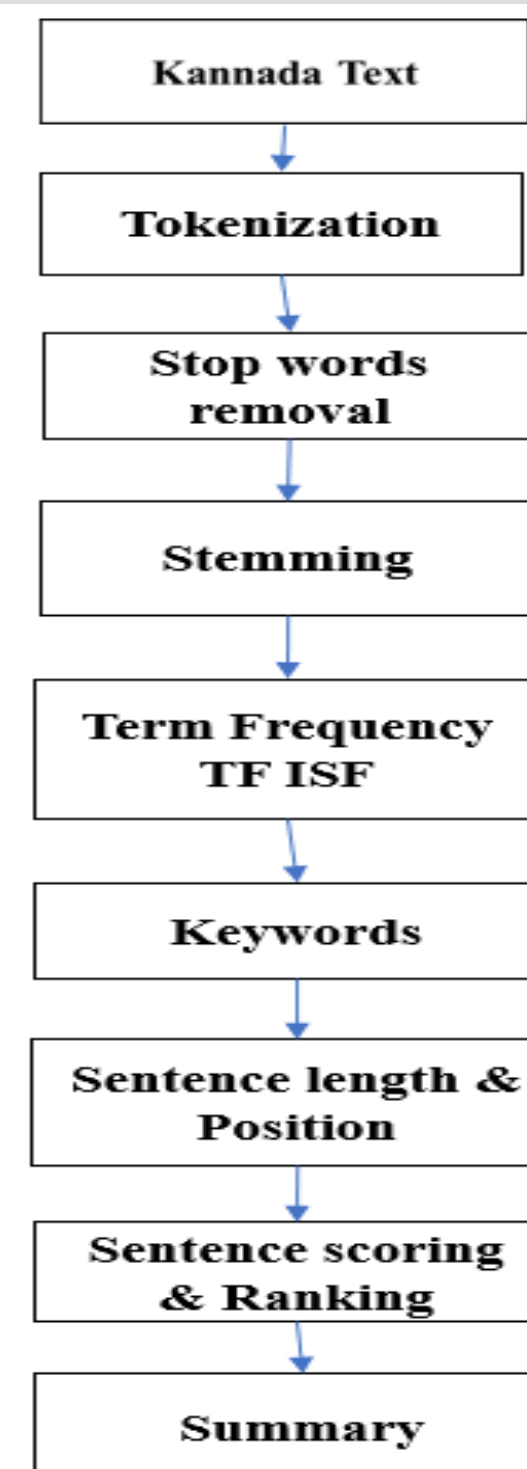


Fig 1: Extracted summary of kannada text.

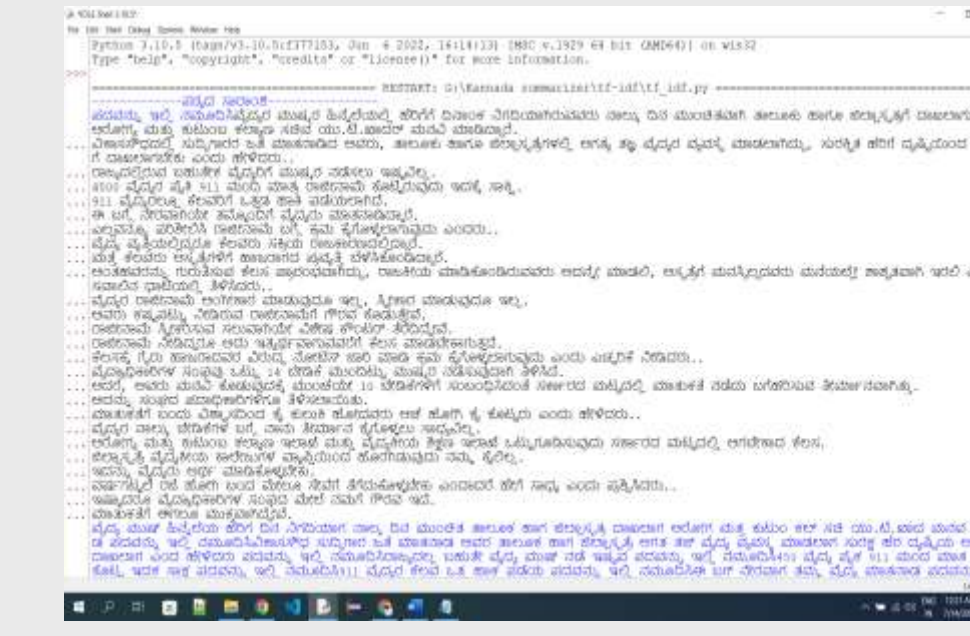


Fig 1.1: Extracted summary of kannada text.

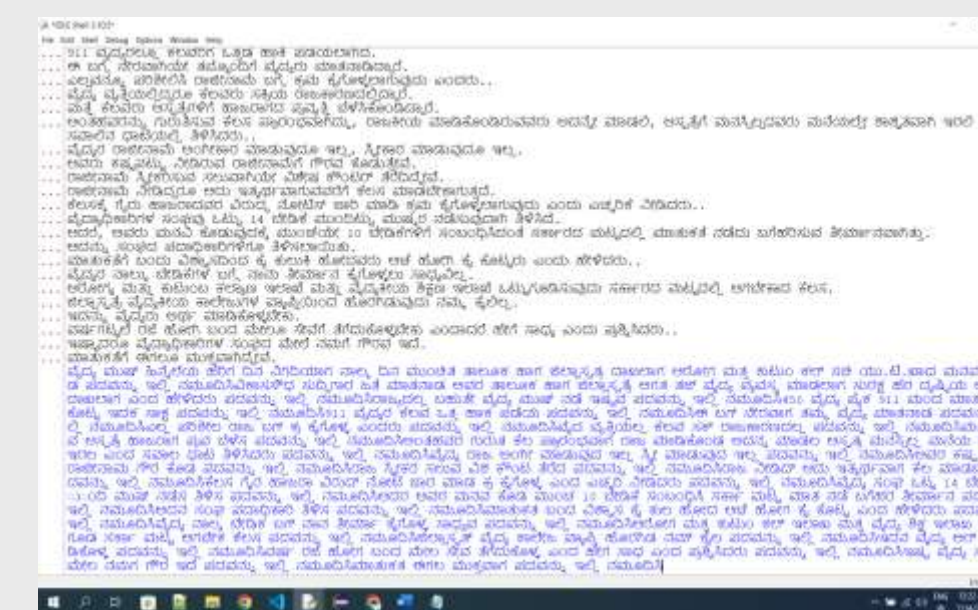


Fig 1.2: Extracted summary of kannada text.

Paragraph	No of sentences	System summary (SS) with no of sentences	Manual summary (MS) with no of sentences	% of meaning of SS V/S MS
1	50	36	28	72
2	72	54	48	75
3	84	58	54	69.04
4	96	66	62	68.75
5	102	70	75	68.62
6	115	68	67	59.13
7	97	82	73	84.53
8	77	57	52	74.02
9	65	42	40	64.61
10	80	50	45	62.5

Fig 2: Analysis of summary generation.

CONCLUSION

Kannada Text Summarization is a demanded application for users to obtain the list of information in a short time of the search. Kannada Text summarization automatically produces a summary containing important sentences and include all relevant important information from the original Text or document.

Percentage of meaning of system V/S manual summary is compared to check the proper sentence format. We have considered 105 paragraphs average result of system V/S manual summary is 65.69%.

REFERENCES

1. A context best text summarization from By Rafael ferreira and Luciano D'souza from IEEE.
2. A sentence selection module and HLO algorithm from extractive text summarization by Russian Rameez and Nijat from IEEE.
3. A survey of extraction and abstraction automatic text summarization technique by Vipul Dalal and doctor lathesh Malik from IEEE..