Janatha Education Society (R)

VIVEKANANDA INSTITUTE OF TECHNOLOGY

(Affiliated to VTU and Approved by AICTE)

Gudimavu, Kumbalagodu Post, Kengeri Hobli, Bengaluru – 560 074

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

PROJECT EXHIBITION – 2019

HYBRID DOCUMENT SUMMARIZATION

Student's Name	USN:	
1. Mohammed Abu	1VK15CS030	**Internal Guide(s):** Prof. Chandramma R
2. Naveen P	1VK15CS034	
3. Nikhil Chandran	1VK15CS038	
4. Srivatsa V J	1VK15CS054	

Abstract

- Document Summarization is the technique by which the huge parts of content are retrieved.
- An online semantic lexicon WordNet is utilized, Word Sense Disambiguation (WSD) is a critical and testing system in the territory of characteristic dialect handling (NLP).
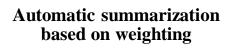
Methodology

- The significance of a sentence in info content is assessed by the assistance of 3 algorithms.
- Document Summarization assesses the weights most sentences utilizing the Simplified Lesk calculation and arranges them in decreasing order of their weights.
- The words to added to sentences, reducing the sentences based on score etc. is done in this step to produce summary.

Conclusion

- The weight-based algorithm gives the summery according to the score of the sentence. The information extraction algorithm uses parts of speech tagging which gives the summary having proper meaning for the sentences. Hence, we like to conclude that information extraction algorithm is more accurate than query and weight-based algorithms.

TABULATIONS OF RESULTS / GRAPHS

Summarization using query-based algorithm
↓
Automatic summarization based on weighting
↓
Information Extraction algorithm
↓
Automatic summarization based on weighting
↓
Test results
↓
Accuracy and graphs based on summirized document

Flow Diagram/ Circuit Diagram