Abstract

A title is a short summary that represents document’s main theme. Title can help the reader to have the main idea without reading the entire document. To generate a title for a document, we have to select appropriate words as title words and put them in sequence. The process of generating title for a given document by using machine, can be done by using summarization
approaches or by using Statistical approaches or by combing both. For a given document, selecting appropriate words for generating a title by using any available approach mainly depends on the characteristics of the language. In this paper, we have examined the influence of the language characteristics in the process of title word selection by using the Naïve Bayes probabilistic approach (called BMW Model) on the documents which are available in the language 'Telugu'. And also we have investigated the influence of word weight for the selection of title words in BMW Model. By using F1 metric, we have evaluated the title word selection process.

Reference

- Term-weighting approaches in automatic text retrieval, Salton and Buckley Information Processing & Management Vol. 24, No. 5, pp. 513-523, printed in Great Britain, 1988
- MUC-6 (1995), Proceeding of The Sixth Message Understanding Conference, 1995

Index Terms

Computer Science Natural Language Processing
Key words

BMW Model
Indic Script
Title Word Selection
F1 measure
Statistical Approach