

## **AN APPROACH FOR EXTRACTING THE KEYWORD USING FREQUENCY AND DISTANCE OF THE WORD CALCULATIONS**

**ASHWINI C. MADANE<sup>1</sup>, J. NAVEENKUMAR<sup>2</sup>  
AND DEVENDRA M. THAKORE<sup>3</sup>**

<sup>1</sup> <sup>2</sup>Research scholar, Bharati Vidyapeeth Deemed University,  
College of Engineering Computer Department, Katraj, Pune, India  
<sup>3</sup>Professor, Bharati Vidyapeeth Deemed University,  
College of Engineering Computer Department, Katraj, Pune, India.

### **Abstract**

A significant word used in indexing or cataloguing is regarded as a Keyword. Keywords provide a concise and precise high-level summarization of a document. They therefore constitute an important feature for document retrieval, classification, topic search and other tasks even if full text search is available. Keywords are useful tools as they give the shortest summary of the document. A keyword is identified by finding the relevance of the word with or without prior vocabulary of the document or the web page. Extracting keywords manually is an extremely difficult and time consuming process, therefore it is almost impossible to extract keywords manually even for the articles published in a single conference. Therefore there is a need for automated process that extracts keywords from documents. This paper concentrates on the extracting the keywords by understanding the linguistic, non-linguistic and various other approaches but applying the simple statistics approach.

-----  
**Keywords:** Keyword Extraction methods, Keyword Frequency Count, Stemming, Tokenization. **1**