Abstract

Traditional information retrieval methods become inadequate for increasing vast amount of data. Without knowing what could be in the documents; it is difficult to formulate effective queries for analyzing and extracting useful information from the data. This survey focused on some of the present strategies used for filtering documents. Starting with different types of text features this paper has discussed about recent developments in the field of classification of text documents. This paper gives a concise study of methods proposed by different researchers. Here various pre-processing steps were also discussed with a comprehensive and comparative understanding of existing literature.

References

System Based on Genetic Algorithm and Taxonomy*. accepted March 9, 2018, date of publication March 15, 2018, date of current version May 9, 2018.

**Index Terms**

Computer Science

Information Sciences

### Keywords

Content filtering, Fake Profile, Online Social Networks, Spam Detection.