



International Journal on Recent Researches In Science, Engineering & Technology

(Division of Computer Science and Engineering)

A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. It is an absolutely free (No processing charges, No publishing charges etc) Journal Indexed in JIR, DIIF and SJIF.

Research Paper

Available online at: www.jrrset.com

Chief Editor : Dr. M.Narayana Rao, Ph.D., Rtd. Professor, NIT, Trichy.

ISSN (Print) : 2347-6729

ISSN (Online) : 2348-3105

Volume 3, Issue 4,
April 2015.

JIR IF : 2.54

DIIF IF : 1.46

SJIF IF : 1.329

To Upgrade the Quality of Search Results By Concept Based Model

W.Rubavathy

UG Scholar, Department of Information Technology
Jeppiaar Engineering College, Anna University
Chennai (Dt), Tamil Nadu, India.
rupawilliams03@gmail.com

P.Sharmila

UG Scholar, Department of Information Technology
Jeppiaar Engineering College, Anna University
Chennai (Dt), Tamil Nadu, India.
sharmila2771994@gmail.com

R.Thilagavathy

Assistant Professor, Department of Information Technology
Jeppiaar Engineering College, Anna University
Chennai (Dt), Tamil Nadu, India.
thilaga1025@gmail.com

Abstract: Text mining methods are based on the vector space model (VSM) which is a widely used data representation for text classification and clustering. VSM mainly depends up on the single term analysis (term frequencies). due to this, numerous keywords will be generated. some keywords are unrelated to the topic. so, this model does not give refined keywords. The objective of this paper is to propose concept based mining model. This model mainly depends up on three levels sentence -based concept analysis, document -based concept analysis and corpus or domain based analysis. sentence based concept analysis gives the conceptual term frequency (ctf) value, document based concept analysis gives the term frequency (tf) value and corpus based concept analysis gives the document frequency (df) value. with the help of calculated frequencies, we are determining the concept based similarity measure. hence matched concepts are found out and clustering is performed. clustering of web documents can be effectively done by the algorithms such as suffix tree and k means. lot of surveys prove that suffix tree algorithm is executing the clustering in a effective manner. This model produces high quality results than VSM.