AN EFFICIENT WEB CONTENT MINING USING RELEVANCE ANALYSIS APPROACH

M. GIRI[a] AND AKASH KUMAR[b]

[a]Associate Professor, Department of Computer Science & Engineering,
Sreenivasa Institute of Technology & Management Studies, Chittoor, Andhra Pradesh, India.
[b]Professor, Department of Computer Science & Engineering,
Bundelkhand University Jhansi, Uttar Pradesh, India.

Abstract

In context of web mining, large collection of web documents are used in the process of mining to extract more useful information. Most of the web information is irrelevant. Web document presents 10-15% of data using 85-90% of tags. The previous researchers on web mining proposed many methods, for mining web documents, but all these mining methods process all documents, statements and words without checking its relevance. The discovered patterns by using all these methods may be correct or may not. In this paper we proposed new web mining method called Web Mining using Relevance Analysis (WMRA). It consists of four phases: document selection phase (list of documents selected), relevance analysis phase (document relevance analysis, statement relevance analysis and word relevant analysis), web mining phase (apply our designed algorithm to identify patterns), presentation phase (presentation of discovered patterns**)**. Experiments are conducted on various web documents that are related to one topic. Experimental results of proposed system produce better, correct and quality patterns compare with existing methods on web document mining.