Chapter 95
Applications of Web Usage Mining across Industries

A. V. Senthil Kumar
Hindusthan College of Arts and Science, India

R. Umagandhi
Kongunadu Arts and Science College, India

ABSTRACT

Web Usage Mining (WUM) is the process of discovery and analysis of useful information from the World Wide Web (WWW) by applying data mining techniques. The main research area in Web mining is focused on learning about Web users and their interactions with Web sites by analysing the log entries from the user log file. The motive of mining is to find users' access models automatically and quickly from the vast Web log data, such as similar queries imposed by the various users, frequent queries applied by the user, frequent web sites visited by the users, clustering of users with similar intent etc. This chapter deals with Web mining, Categories of Web mining, Web usage mining and its process, Applications of Web usage mining across the industries and its related works. This Chapter offers a general knowledge about Web usage mining and its applications for the benefits of researchers those performing research activities in WUM.

WEB MINING

Definition of Web Mining

Web mining is a technique used to automatically discover and extract the interesting and potentially useful patterns and implicit information from the web documents and services (Etzioni, O. 1996). Exploring and extracting precisely pragmatic knowledge from web data is also called as web mining. Web mining is indispensable to enhance the utility of web. Application of data mining techniques in the World Wide Web is called as web mining (Srivastava, T. *et al*., 2005).

DOI: 10.4018/978-1-5225-1837-2.ch095

Web is the largest and voluminous data source in the world. The plentiful unstructured or semi-structured information on the web leads to a great challenge for the users, who hunt for prompt information. The scenario grows pathetic and distressing to provide personalized service to the individual users from billions of web pages. The unpredictable amount of web information available becomes a menace of experiencing ambiguity in the web search. To prevent the web users from getting overwhelmed by the quantity of information available in the web, search engines are used.

The massive utility of web resources in recent scenario has turned to be an essential commitment for numerous reasons. Clinging on to the web information from a microcosmic level to the macrocosmic level has been growing over the last three decades. At the same time, the inconceivable boom of information available in the websites simultaneously throws the challenge of retrieving the precise and appropriate information at the time of need. To state the precise statistics of active websites, the March 2012 survey of Netcraft (http://news.netcraft.com/archives/2012/01/03/january-2012-web-server-survey.html; March 2012) figures around 644,275,754 websites may be quoted. This survey aids to comprehend how the web appears to be a panacea due to its inevitable applications in several facets of life. Moreover, the web information is the mostly sought after powerful platform for working, studying, searching information, besides, being in touch with our friends. Apparently, the unpredictable amount of web information available becomes a menace of experiencing ambiguity in the web search. To prevent the web users from getting overwhelmed by the quantity of information available in the web, several strategies are proposed. These strategies attempt to solve the tedious information exploration process of the user, through Information System, Information Filtering and Recommendation Systems.

Applications of Web Mining

Web mining is used in four significant fields namely, Resource finding, Information selection and Pre-processing, Generalization and Analysis. Retrieving the anticipated web resource through exploration is called Resource finding. Information selection and Pre-processing is the process of making automatic choices while pre-processing to obtain a definite data from the retrieved web resources. Automatic method to examine general patterns at individual web sites as well as across multiple sites is called Generalization. Analysis is a method of validation and/or interpretation of the mined patterns to reinstate the quality of results observed.

Classification of Web Mining

Web mining is categorized into, web content mining, web structure mining and web usage mining based on type of data used for mining processes. However the prevalent types are web content mining and web usage mining. In one, web structure is considered as part of web content; while in the other, web usage is considered as part of web structure. The main goal of all the three classification types is the method of knowledge discovery of inherent, unidentified and potentially valuable information from the web. Each of them focuses on varied mining objects of the web. Figure 1 describes the web categories and their objects. Subsequently, a brief introduction is given for each of the categories.