Chapter 8
Web User Behavior Analysis Using Pre-Processing of Web Documents to Create Effective Web Designs

Abhijit Dnyaneshwar Jadhav
Dr. D. Y. Patil Institute of Technology, Pimpri, India

Santosh V. Chobe
Dr. D. Y. Patil Institute of Technology, Pimpri, India

ABSTRACT

The interactions with web systems is huge because of COVID-19, where every system is run through online interactions. The world wide web is continuously expanding, and users' interactions with websites generate a vast quantity of data. Web usage mining is the use of data mining techniques to extract important and hidden information about users. It allows you to see the most frequently visited sites, imagine user navigation, and track the progress of your website's structure, among other things. The web mining techniques help us to analyze the user's behavior and accordingly create the required web designs, which will appear in the relevant searches of the users. In this scenario, one of the important processes is web document preprocessing, which will help us to extract the particular quality data inputs for analyzing the behaviors which helps in effective web design. Here, the authors discuss preprocessing of web documents. From the four different phases of the web mining, web document pre-processing is a very important phase.

DOI: 10.4018/978-1-7998-9426-1.ch008

WEB MINING

The World Wide Web is a collection of web sites that provides internet users with a wealth of information. For internet users, the knowledge available on the internet has evolved into a valuable resource. Because the number of websites available on the internet is growing and becoming more complicated, the total volume of web is enormous. A website serves as a link between the customer and the business. The corporations can update visitor's performance during web inquiry, and identify the trends. Web mining is defined as the search for and analysis of useful information on the World Wide Web. Web content mining, web structure mining, and web use mining are the three types of web mining. The extraction of useful information and online knowledge from web resources or web contents such as text, picture, audio, video, and structured data is referred to as web content mining (Mehra & Thakur, 2018). Web use mining may be defined as the discovery and analysis of user access patterns using log file mining. The WUM's output may be utilised for web personalisation, recovering system performance, site modification, and use description, among other things. Web log file is a server log file that contains access logs of the web server and is a vital data source in Web use mining. The Data Preprocessing segment is a crucial stage in the WUM. Data cleansing, session identification, user identification, and path completion are all included. Material preprocessing is used to remove unwanted data from log files so that the pattern discovery algorithm can detect the user pattern (Anand & Aggarwal, 2012).

Need of Web Mining

Web mining is the use of Data Mining methods to locate and extract information from Web publications and services automatically. Web mining's major goal is to extract relevant information from the World Wide Web and its usage trends. The focus on web mining in academics, the software business, and online-based organizations has resulted in a substantial amount of expertise. By recognizing online pages and categorizing web content, web mining helps to increase the power of web search engines. E-commerce websites and e-services benefit greatly from web mining (Chu et al., n.d.).

Types of Web Mining

As seen in Figure 1, web mining may be separated into three types.

- **Web Content Mining:**