# Web Mining Concepts, Applications and Tools:A Survey

Dr.A.Mumtaj Begam[1], B.V.Saravanan[2]
[1]Head &Assistant Professor, Department of Computer Science,
[2]Guest Lecturer, Department of Computer Science,
Government Arts College, Paramakudi

**Abstract**

**Web Mining is the application of data mining techniques to extract knowledge from web data, including web documents, hyperlinks between documents usage of websites. Web Mining helps to solve the issue of discovering how users are using a website. Large amount of text documents, multimedia files and images are available in the web and it is still increasing. Web Mining has been categorized into three distinct ways the first, called web content mining is the process of information discovery from sources across the World Wide Web. The second, called web usage mining is the process of mining for user browser and access pattern. The third, called web structure mining is the process of analyzing the relationship between web pages linked by information or direct link connection through the use of graph theory.**

**Key Terms:** Web Mining, Web Content Mining, Web Usage Mining, Web Structure Mining

## 1. Introduction

The web is the interesting area of research. Web mining is moving the World Wide Web towards a more useful environment in which users can quickly and easily find the information they need [1]. It uses document content, hyperlink structure and usage statistics to assist users in meeting their need information. The main goal of web mining is to look for useful patterns in web data by collecting and analyzing information in order to gain insight into trends, the industry and users [2]. Web mining has two views in general web mining with the user centric view allows to discovery of documents on a subject. Discovery of semantically related documents or documents segments, Extraction of relevant knowledge about a subject from multiple resources, knowledge/information filtering. Web Mining with the owner centric view allows getting increasing contact/conversion efficiency (Web Marketing), Targeted promotion of goods, services, products, ads; Measuring effectiveness of site content/structure, providing dynamic personalized services or content [2]. Basically there are three sub categories as shown in figure 1.

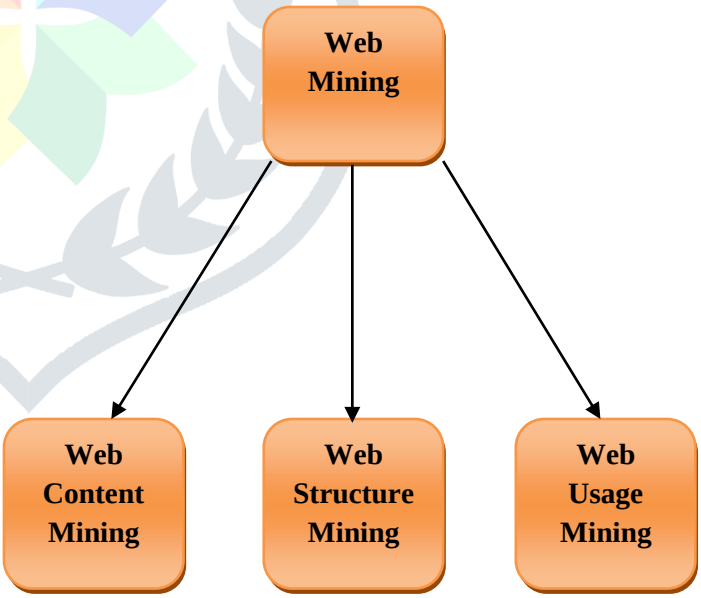


**Fig 1 Web Mining**

### 1.1. Web Content Mining

Web content mining includes the web documents which may consists of text, html, and multimedia documents, i.e. images,

audio, video etc. It deals with discovering useful information or knowledge from the web page contents than hyperlinks and goes beyond using keywords in a search engine. The web content mining has several steps such as collect the information, parse it, analyze it, and then produce the useful result. This type of web mining is also called as web text mining because text content is most popularly extracted. The two technologies are generally used for this type of mining: IR (Information Retrieval) and NLP (Natural Language Processing). The information retrieval should be performed from two views because the web page may contain semi-structured data or unstructured data.

### 1.2. Web Usage Mining

Web Usage mining also called web log mining is the process of discovering meaningful information from web data usage. It is also used to discover the interesting usage data. This includes server data (IP Address), Application server data (web logic), and application level data (events). Web usage mining is automatic discovery of user'saccess patterns. The mined data includes data logs of users web interaction, having web server logs, Proxy server log and browser logs, having data about referring page, user identification, user spent time at site and sequence of pages visit [3]. Also cookie files contain information.web usage mining has its application in improving websites design, intrusion detection, for predicting users interest, for analyzing website performance, for excellence of e-commerce, for identifying suspicious activities, to find out primary places for advertising, for social network analysis, for analysis of traffic etc.

### 1.3. Web Structure Mining

Web structure mining deals with modeling and discovering the hyperlink structure of the web pages which is based on the topology of the hyperlinks. A hyperlink is a structural unit that connects a location in the web page, either within the same web page or on a different web page. A hyperlink that connects to a different part of the same page is called as intra document hyperlink and the hyperlink that connects two different pages is called as inter document hyperlink. Hyperlink analysis can be done based on knowledge models, scope and properties of analysis and types of algorithms. Discovering hyperlinks structure within the website is also called as "Hyperlink Analysis". Two algorithms are mostly used for this type of mining that are Page Rank and HITS.

## 2. Web Mining Applications

Web Mining becomes a very hot and popular topic in web research area. Web mining it also plays an important role for E-Commerce and E-Service web site to understand their web sites and service are used and provide better service for both customer and user [5].

### 2.1. E-Commerce:

E-Commerce also known as E-Business or electronic business, is simply the safe and purchase of services and goals over an electronic medium, like the internet. It also involves electronically transferring data and funds between two or more parties. Simply put, it is online shopping as we commonly know it. Web mining techniques can support a web enabled electronic business to improve on marketing customer support and sales operations.

### 2.2. E-Learning

E-Learning or "electronic learning" is an umbrella term that describes education using electronic devices and digital media. It encompasses everything from traditional classrooms that incorporate basic technology to online universities. Web is only major choice of administer and maintain learning resources and has turn into one of the foremost choice of modern advanced distance education system. As learning becomes more technologically advanced, the difficulty of available learning resources also enlarged accordingly. Web mining can be used for improving and enhancing the process of E-Learning environments.

### 2.3. E-Governance

E-Governance means governing or administering a country /state or organization, with the help of information and communication

technology.This system may provide customized services to citizen resulting in user satisfaction and quality of services and support in citizen decision making, which leads to social benefits.

### 2.4. E-Marketing

Internet is a global computer network providing a variety of information and communication facilities. The internet or internetwork is simply the interconnection of networks globally. Internet marketing efforts are done solely on the internet. This type of marketing uses various online advertisements to drive traffic to an advertiser's website. Banner advertisements pay per click (ppc) advertisements and targeted email lists are often and used methods in internet marketing.

### 2.5. Credit Card Fraud Detection

The usage of credits cards dramatically increases day by day. Credit card becomes the most accepted popular and stylish mode of payment for both online as well as offline purchases. Because with it are also increasing. Various techniques of web mining like, classification, clustering and association will be integrated to represent the sequence of operations in credit card transaction processing and show how it can be used for the detection of frauds.

## 3. Web Mining Tools

Different web mining tools are available for example.

### 3.1. Alter Wind Log Analyzer Lite

AlterWind Log Analyzer Lite is free web site statistics software. With our free log analyzer tool help you can determine the basic characteristics of the hits on your site, learn which search engines and search phrases bring you visitors, which referring sites generate the most traffic, and discover and fix errors in the workings of your web site.

### 3.2. JWAanalytics

JWAnalytics allows real time decisioning for a website to take place (content optimization, product recommendations, FAQ search results ranking, price optimization, etc, see whole list of possible uses for details). JWAnalytics is an innovative technical solution for the storage of top quality analytical data and the real time use of this data [4]. It is a solution that will support your business strategy and the processes around it in real time (real time data mining).

### 3.3. ht: //Miner

ht: //Miner is an open source application for web usage mining and data warehousing.

Through the automatic processing of web server log files, ht://Miner is able to discover precious information regarding visitors' behavior and to store them in a PostgreSQL database for further analysis and reporting operations.

### 3.4. Web Utilization Miner (WUM)

WebUtilizationMinerWUM, a miningsystemforthediscoveryofinterestingnavigationpatterns.TheinterestingnesscriteriafornavigationpatternsaredynamicallyspecifiedbythehumanexpertusingWUM's mininglanguageMINT.

### 3.5. Web Content Extractor (WCE)

Web Content Extractor is highly accurate and efficient for extracting data from websites.

### 3.6. Screen-Scraper

Screen Scraping or Web Scraping is the process of automatically downloading text, images, and other content from websites via data extraction software.

### 3.7. Web Info Extractor (WIE)

Web Information Extractor is a powerful tool for web data mining, content extraction and content update monitor. It can extract structure or unstructured data (including text, picture and other file) from web page, reform into local file or save to database, post to web server.

### 3.8. Mozenda

Mozenda automates data collection from the web and converts unstructured data into usable data sets by using a point-and-click interface.

## 4. Conclusion

This paper describes the basic concepts of web mining. Then we discussed process of web mining, its types in detail, the different techniques used for it and applications. Finally applications are discussed which specifies a fields where actually web mining is used. There are so many tools available to work on web mining some prominent tools are discussed in this paper.

## REFERENCES

1. A. Laender, B. Ribeiro-Neto, A. da Silva, and J. Teixeira. A brief survey of web data extraction tools. SIGMOD Record, 31(2):84{93, 2002.

2. M. Eirinaki, M. Vazirgiannis, "Web mining for Web personalization", ACM Transactions on Internet Technology, 3(1), 2003, pp. 1–27.

3. Zdravko Markov, Daniel T.Larose "Webcontent structureand Usage", Wiley, 2007.

4. Bodyan G.C, Shestakov T.V, "Web Mining in Technology Management", Engineering Universe for Scientific Research and Management, Vol 1 Issue 2, April 2009.

5. V. Bharanipriya & V. Kamakshi Prasad, Web Content Mining tools: A Comparative Study in International Journal of Information Technology and Knowledge Management January-June 2011, Volume 4, No. 1, pp. 211-215.

6.Mozenda, at: http://www.mozenda. Com/web-mining-software Viewed 18 February 2013.

7. Kamika Chaudhary, Santosh Kumar Gupta, Web Usage Mining Tools & Techniques: A Survey in International Journal of Scientific & Engineering Research, Volume 4, Issue 6, June-2013 1762 ISSN 2229-5518.

8. D. Sridevi, Dr. A. Pandurangan, Dr. S. Gunasekaran, "Survey on Latest Trends in Web Mining", International Journal of Research in Advent Technology,Vol. 2, No.3, March 2014.

9. Aarti Sharma et al, "Application of Data Mining – A Survey Paper", International Journal of Computer Science and Information technologies', Vol. 5 (2), 2014.

10. Ms.Gaikwad Surekha Naganath, Ms. Mali Supriya Pralhad, "Web Mining –Types, Applications Challenges and Tools.

11. Web mining https://www. techopedia.com/definition/ 15634/ web-mining