A Survey on Web Log Mining (A case study)

Vishal Gupta[1] Mahesh Malviya[2]
[1]P. G. Student [2]Assistant Professor
[1, 2]Department of Computer Science & Engineering
[1, 2]JIT, Borawan(M.P.)India

Abstract--- In today scenario where the data mining techniques and web technologies has decreases the cost of storing and collection of the various types of data. Web mining is a type of data mining techniques to fetch data and its related information from World Wide Web. Huge amount of data that has being collected from the different web sites which has becomes a new challenge for these techniques to organize this data in a way to make use of this stored data or Knowledge to gain an edge over their competitors. With the rapid increase in the growth of the Web data, the study of data discovery in web mining and its modeling and predicting the user's access on a web has become very important. This paper deals with the preliminary discussion of various web log mining activities, classification, tasks and its application.

Keywords: Web Mining, Web usage Mining, Web Mining Structure, Web Mining Application, Web Content Mining.

I. INTRODUCTION

Web mining can be defined as the process of discovery and analysis the useful information from the web. There are various ways to obtain important information from the web [1][3].On the basis of which, web mining can be divided into three parts: Contents Mining, Structure Mining and Usage Mining. Fig 1 shows the classification of Web Mining

A. Web Content Mining:

Web Content Mining can be defined as a process of extracting useful information and data from the World Wide Web documents. Content data may corresponds to the web pages that are designed by the user for convey [2].It mainly consists of Text, Images audio or video and record like tables form or listed form. Text mining is an application to the Web content has become the widely used researched. Issues which are addressed in text mining are, topic discovery, extracting association patterns, clustering web documents and classification of Web Pages.

B. Web Structure Mining:

The web graph structure consists of Web pages as nodes, and its hyperlinks as edges connecting to the related pages. Web Structure Mining can be defined   is the process of discovering structure information from the World Wide Web. It can be further divided into two types based on their kind of structure information used.

1) Hyperlinks:

A Hyperlink is structural unit which is used to connect a location in a World Wide Web page to other location, either within the same World Wide Web page or on a different Web page. A hyperlink which connects to a different part of the same webpage is called an Intra-Document Hyperlink, and a hyperlink which is connects two different web pages

is termed as Inter-Document Hyperlink. There has been a significant body of work on hyperlink analysis, of which Desikan et al. [7] provide an up-to-date survey.

2) Document Structure:

The content of web pages can be arranged in the form of tree-structured format, based upon their HTML and Xml page tags. Web Mining efforts have focused on automatically extracting of document object model (DOM) structures out of documents [8].
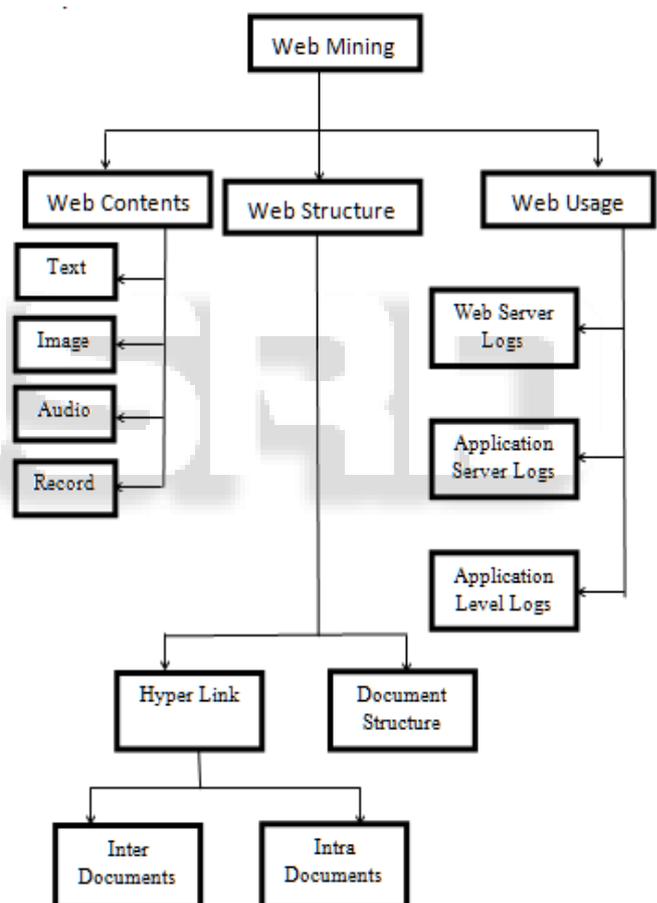

Fig. 1: Classification of Web Mining

C. Web Usage Mining:

It can be defined as the data mining techniques application to generate or discover interest usage patterns from World Wide Web data, and to understand its better serve of world wide Web-based applications [9]. Web usage mining can be classified on the kind of usage data:

1) Web Server Logs:

Web server collect all the user logs. This typical data includes IP address, page reference and its access time.

2) Application Server Logs:
Commercial application servers like Web logic [6], Story Server has a significant features to enable the application of E-commerce and build them with less effort. Main feature is to track various kinds of events related to the business and log their application in application server logs.

3) Application Level Logs:
Newly generated events are defined in an application, and its logging can be turned on for the generation of histories of all these events that are defined [9].

II. FUNCTION OF WEB MINING

Web mining mainly consists of three major tasks: crawling, topology analysis and content analysis. The art of traversing the worldwide web very efficiently and effectively is known as Crawling.

The process of exploring the relationship among the World Wide Web to find out the authoritative sources on the Internet is term as Topology analysis. Analysis the content of any web data is term as Content analysis and it is similar methods as data mining and machine learning.

III. WEB USAGE MINING AND PATTERN GENERATION

Web usage mining is an application in data mining techniques where we discover usage pattern from World Wide Web data [4].Web usage mining mainly consists of three parts, preprocessing of data, pattern discovery of pattern, and analysis of pattern. A high level Web usage mining Process of high level is presented in Figure 2 [5]. Mobasher et al. [4] has proposes the web mining process can be divided into two main parts. The first part may include the processes of transforming the World Wide Web data into its suitable transaction form. It will include preprocessing of data, transaction of data and its identification, and integration of data components. The second part may contain data mining and pattern matching techniques like association rule and sequential patterns. Firstly in the preprocessing state the user sessions is inferred by log data. In the second stage searching for appropriate patterns in the data is done by using standard data mining techniques, like association rules or mining for sequential patterns. In the third stage filter information bases on domain knowledge and the World Wide Web site structures is applied to the mining patterns for searching the interest of the patterns. The preprocessing phase allows the converting of the server sessions into episodes prior to performing discovery of knowledge.
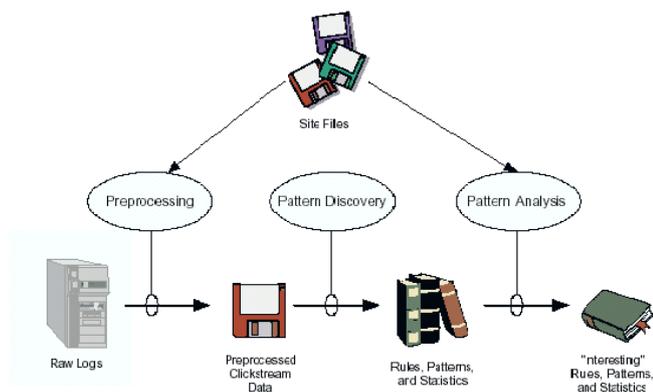

Fig. 2: A High Level Web Usage Mining process

IV. WEB MINING APPLICATIONS

Web mining analysis the corporate information more effectively by combing it with another corporate information or data with the help of web traffic data [10]. It will allows the account department, customer, inventory management and other demographic information or data to correlate with the world wide web using web browsing and will be able to answers the following complex question like-

1) How many people have hit their website and have done purchasing?
2) What is the advertising campaigns result on doing purchasing, rather than hits?
3) Web Mining can help the company for the effectiveness of their marketing or marketing research, which will help to improve their business.
4) In the today business world, structure mining is very useful in determining the connection between two or more business sites.
5) With the web mining the company can identify their strength and weakness, and can make strategic adjustments, to improve the weakness by the feedback.
6) Search engine provides an advanced and efficient searching capability to the Google. [11]. Fig 3 Shows the idea of application of web mining for the companies.
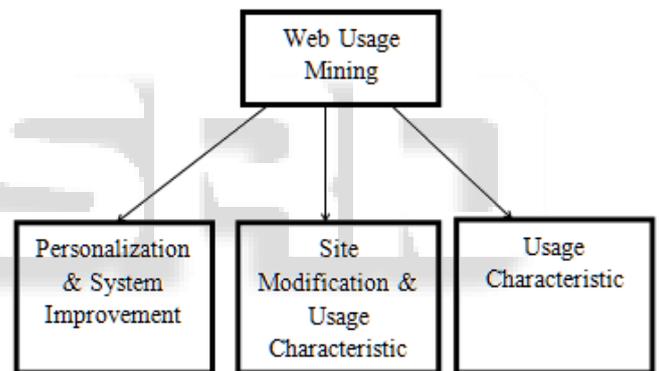

Fig. 3: Application of Web Mining

V. CONCLUSION

In today revolution where the Internet has grown from simple reach to its. In this scenario the company can find the new and better way to do their business and be up to date by E-commerce. Implement of web mining system in companies can increase customers' profiles, identify their strength and weakness. Internet is a gold mine, only for that companies which realize what the importance of Web mining and its adoption.

VI. REFERENCES

[1] Cooley, R., Srivastava, J., & Mobasher, B.. (1997) Web Mining: Information and Pattern Discovery on the World Wide Web. In Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97).
[2] B. Masand, M. Spiliopoulou, J. Srivastava, O. Zaiane, ed. Proceedings of "WebKDD2002 –Web Mining for Usage Patterns and User Profiles", Edmonton, CA, 2002.

[3] Kosala R., Blockeel H., "Web Mining Research: A Survey", ACM SIGKDD Explorations Newsletter, Vol. 2, No. 1, pp. 1 – 15, 2000

[4] Robert Cooley, Bamshad Mobasher, and Jaideep Srivastava, Web Mining: Information and Pattern Discovery on the World Wide Web (A Survey Paper) (1997), in Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), November 1997.

[5] Jaideep Srivastava,Robert Cooley, Mukund Deshpande,Pang-Ning Tan, Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data(2000). SIGKDD Explorations, Vol. 1, Issue 2, 2000.

[6] BEA Weblogic Server. http://www.bea.com/product/weblogic/server/index.shtml.

[7] P. Desikan, J. Srivastava, V. Kumar, and P.N. Tan. Hyperlink AnalysisTechniques &Applications. Technical Report 2002-152, Army High Performance Computing Research Center, 2002.

[8] C.H. Moh, E.P. Lim, and W.K. Ng. DTD-Miner: A Tool for Mining DTD from XML Documents. WECWIS, 2000.

[9] J. Srivastava, R. Cooley, M. Deshpande, and P.N. Tan. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. SIGKDD Explorations,1(2):12–23, 2000.

[10] Jaideep Srivastava, Prasanna Desikan, Vipin Kumar, "Web Mining Concepts ,Applications and Research Directions"

[11] R. Cooley, B. Mobasher, J. Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns", Knowledge and Information Systems, 1(1), 1999.