JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

The Study of Web Usage Mining For Website Personalization

Charu Sharma
Himachal Pradesh University, Shimla, India.

Abstract

Today there are approximately 5.03 billion people around the world who are having access to the internet. Each user has his own requirements and variance in accessing the data on the internet. Hence, today with the increase in competition, the concept of Web Usage Mining has evolved which is being used by many different organizations to identify the access patterns of every user accessing their website so that in future they could satisfy the requirement of the increasing number of users access and cater to their needs in a more enhanced manner. This paper focuses on the concept of Web Usage Mining and its technique. Also, the paper laid emphasis on the need of using Web Usage Mining for Website Personalization depending on the user's previous access.

Keywords: Web Mining, Web Content Mining, Web Structure Mining, Web Usage Mining, Website Personalization, Data Pre-Processing, Pattern Analysis.

1. Introduction

Web Mining is the process of mining web data. The type of data extracted for performing various data mining techniques to identify pattern of the data, categories the process of Web Mining into three broad Categories: Web Content Mining, Web Structure and Web Usage Mining[2]. Hence, it is the application of Data Mining techniques on website data.

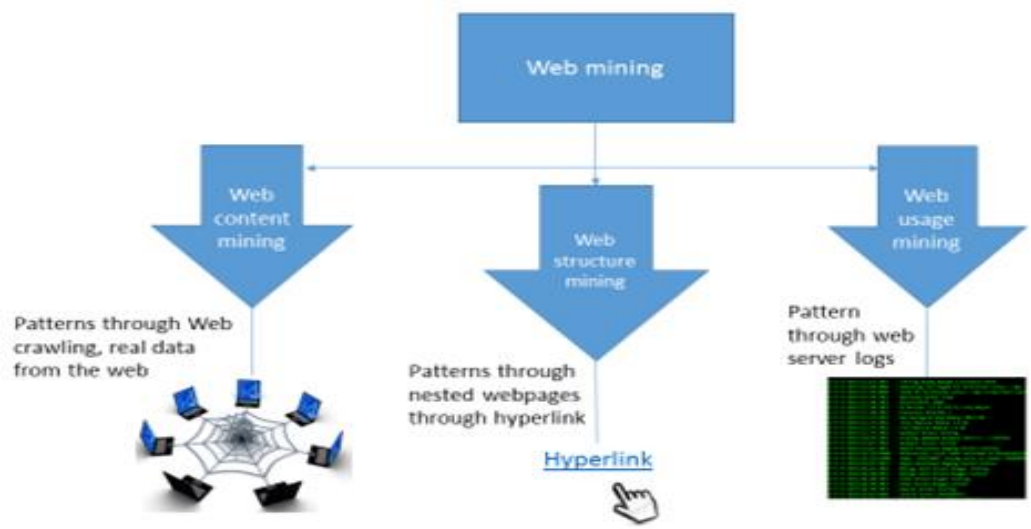

Image 1: Web Mining Categories [9]

1.1 Web Content Mining:

It is to identify pattern of the data available on the website. As the data is neither homogeneous nor structure, various tools such as web crawlers, scraper tools, etc. are used to collect and analyse the data.

1.2 Web Structure Mining:

This is the second category of Web Mining, in which the pattern of hyperlinks forming a website is identified. This is studied to identify the wrong hyperlink from a website which may result in giving irrelevant data to the user.

1.3 Web Usage Mining:

It is the process to identify the user access behaviour on the website so to identify their future access patterns. There are various phases involved and has many tools such as R, Rapid Miner, etc. to perform mining on the log file data.

2. Web Usage Mining

The paper focuses on the study of the concept of Web Usage Mining, which is the process of mining user access pattern on accessing the website. Hence, the data used in performing Web Usage Mining is the Log File Data[6]. There are broadly three type of log file data which are as follows:

- **Web Server Log File:** This log file is maintained at the server site by the website developers. The analysis of the pattern of this data would help the website developers to identify which items on the website would be accessed by the user in future access on the website
- **Proxy Log File** - Proxy Server is the server residing between the server and the client. The log file is maintained in the proxy server and the data of user access is shared to the gourp of other users sharing the proxy server.
- **Client Log File**: The log file is maintained in the client side of the user by their browswers. Client collection requires user participation as data is collected by using java applets or scripts. The main advantage of it is that itr resolves caching problem and session identification. It captures every click of the user using thebworser so that in future based on his previous clicks, the desired data is sent to the client.

2.1 Process of Web Usage Mining

Web usage mining is a process of extracting usage behaviour of the user on the web. The process uses basic data mining phases to extract the data which are:-

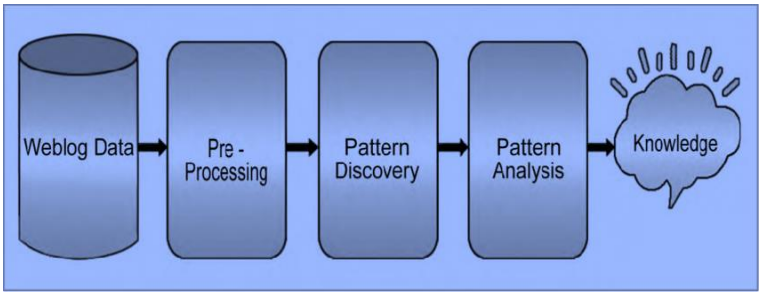

Image 2 - Phases in Web Usage Mining [10]

a. *Data Collection* – The first phase of web usage mining process is to collect data from the desired resource. The source of data in web usage are log files. The log files are divided into four types which are:- server log files, client log files, proxy server log files and browser log files.

b. *Data Pre-processing* – The second phase in web usage mining process is data pre-processing i.e. to process the data before using various pattern discovery techniques so that data can become consistent and integrated. The steps in data pre-processing are data cleansing, user identification, session identification and path completion.

c. *Pattern Discovery* – In this phase of web usage mining, the pattern of the pre-processed data is identifies. There are various statistical and machine learning techniques to extract the pattern of the data. The most common methods are clustering, classification, association rules, path analysis. The study done has used clustering technique in which K-Means Clustering Technqiue is used.

d. *Pattern Analysis* –The last phase of web usage mining process is to analyse the pattern discovered from studying the user navigation. The pattern extracted is to be tranformed to understandable form so that the pattern extracted can be analysed easily. There are methods to analyse the patter discovered which are graphics, database quering, visualization, statistics and usability analysis[10].

3. Website Personalization

Website personalization is the process of providing customized experience to the users of the website so as to cater the needs of the user. In recent times, use of internet has increased manifolds which has resulted in the increase in use of various websites for information gathering. Thus, there is a need for every website developer and designer to identify the navigational behaviour of the user in respect to some period of access on the website of the user. This will help the developer to provide the data to the user in which the user is interested rather than various irrelevant information available on the internet. Website Personalization usage has been tremendously increased in various E-Commerce websites. Therefore, the use of it is greatly increased so has become an important area of research. Therefore, Website Personalization is the most challenging and important category of Web Mining.

4. Related Works

As the user access on every website has increased with the increase in internet, there is an urgent need for every website developer to identify the user access behaviour so that the website is able to provide the desired information. Thus, Payal Sagar, Prof. A.V.Nimavat[06], in their paper has studied the entire concept of Web Usage Mining along with techniques in its three phases i.e. Data Pre-processing, Data Cleaning, Pattern Analysis and Pattern Discovery. Another scholar Tayyaba Ashraf , Imran Ashraf[01]: in their paper have studied and reviewed Web Usage Mining along with the techniques and phases involved in this mining process. Hence, Jitendra N. Shrivastava , Shivendra Pratap Singh[08], have focused on the study and evolution of Web Usage Mining by focusing on its past, future and its entire evolution. The paper also lays emphasis on the application and future scope of Web Usage Mining.

5. Conclusion

The study focused on the concept of Web Usage Mining along with its various phases and need. The paper also studies the need of Website Personalization in today's era of increasing database of customers accessing the

website. The study would helps us to apply various pattern analysis and data pre-processing techniques to identify any log file data to identify user access behaviour on a website.

References

1. Tayyaba Ashraf , Imran Ashraf: A Summarized Review on Web Usage Mining, International Journal of Current Engineering and Technology, Volume 4 No. 4, August, 2014.

2. Namdev Amrit, Varsha Patil: Survey Paper on Web Usage Mining for Web Personalization, International Journal of Innovative Research and Development, Volume 3 Issue 7, July, 2014.

3. Arjun Sidana, Dr. Himanshu Aggarwal: Review of Web Usage Mining of Web Mining, International Journal of Advanced Research in Computer Science, Volume 8 Np. 5, July 2017.

4. Rachit Adhvaryu, A Review Paper on Web Usage Mining and Pattern Discovery: Journal Of Information, Knowledge And Research In Computer Engineering, Volume 2 Issue 2, October, 2013.

5. Anupama Prasanth, Web Personalization Using Web Usage Mining Techniques: Technical Research Institute India, Volume 3 Issue , 2016

6. Payal Sagar, Prof. A.V.Nimavat: Web Usage Mining: Survey on Process and Methods, Research Hub – International Multidisciplinary Research Journal, Volme 2 Issue 5, May 2015.

7. Preeti Rathi, Dr (Mrs) Nipur Singh: A Survey of Issues and Techniques of Web Usage Mining , International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056, Volume 4 Issue 7, July 2017.

8. Jitendra N. Shrivastava , Shivendra Pratap Singh: A Survey of Web Usage Mining: Concepts With Applications and Its Future Scope, International Journal of Computer Science Trends and Technology (IJCST) – Volume 4 Issue 2, Mar - Apr 2016

9.https://www.google.com/search?q=web+mining&rlz=1C1CHBD_enIN809IN809&sxsrf=ALiCzsaK2jJuGfeVFQXVYCx0u8wyldZRbQ:1661711586036&source=lnms&tbm=isch&sa=X&ved=2ahUKEwir5LeGlur5AhXmTgGHdEXAUIQ_AUoAnoECAIQBA&biw=1536&bih=696&dpr=1.25#imgrc=69qRUVNVdFlchM, [Accessed on December, 2019].

10.https://www.google.co.in/search?q=web+usage+mining&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjY6fvg2ZLZAhUFSo8KHRjdAE4Q_AUICigB&biw=1280&bih=651#imgrc=QtgCXSwcj77iJM: [accessed on 22nd July, 2018 at 4:00 P.M.].

10. M. Aldekhail: Application and Significance of Web Usage Mining in the 21st Century: A Literature Review, International Journal of Computer Theory and Engineering(IJCTE), February 2016, Vol. 8, pp 41 – 47, February 2016.