

Role of Web Usage Mining Technique for Website Structure Redesign

Sadik Khan^{*1}, Dr. Yashpal Singh², Dr. Kalpana Sharma³

^{*1}Research Scholar, Bhagwant University, Ajmer, Rajasthan, India

²Associate Professor, B.I.E.T., Jhansi, Uttar Pradesh, India

³Assistant Professor, Bhagwant University, Ajmer, Rajasthan, India

ABSTRACT

Website is very important for any business, your website is the first thing that users see in regards to your business, so everybody want to be sure that you are giving users a better experience. That includes incorporating elements such as: A good, easy interface, Solid navigation and page structure, content presented in a easy way, check website is updated regularly. These details are very important between user staying on your website and having a great look around or leaving straight away to visit your competitor's site. Technical team of website maintenance may want to know how more visitors come to the website, which web pages are being visits mainly or lastly. Which area of website is most or least popular and need some enhancement. Web Usage Mining is a part of Web Mining. Web Usage Mining mainly serve with discovery and analyzing of usage patterns in order to concern the needs of web applications. Server log files are very useful and can provide interesting and useful information. Information provided by log data can solve user's behavior. This may improve the overall importance of the Websites by adapting the proposed structure based on users' behavior. Web Usage Mining techniques use Server log data to generate the user navigation patterns. This information is used to recommend website structure redesigning for enhancement.

Keywords : Web mining, Web usage mining, Server logs, Website redesign

I. INTRODUCTION

The world wide web contains immense amount of data. If anyhow we would be able to extract useful information from the pool of data, it will be a huge advantage for us. But this extraction task, if performed manually is almost impossible for human. This paper is divided into some sections:

In this paper we discussed about web mining, it's processing and types, current fashion of website development. Personalizing the services provided by a Web site to fit the needs of the visitors, we should personalize the web site in serving it visitors. Otherwise, users having difficulties in understanding

how the web site should be explored are disappointed and customers are lost.

Server logs of search engines keep traces of queries posted by users, which contain queries themselves along with Web pages chosen in their answers. Query mining is based mostly in the fact that user queries in search engines and Websites give beneficial information on the pursuits of people.

II. WEB MINING

Gathering of Information from WWW and integrating it to the current gathered data using traditional mining methodologies and techniques. It

also Includes Extraction of common pattern in the data using the technique of web mining.

Web mining has been categorized into following categories:

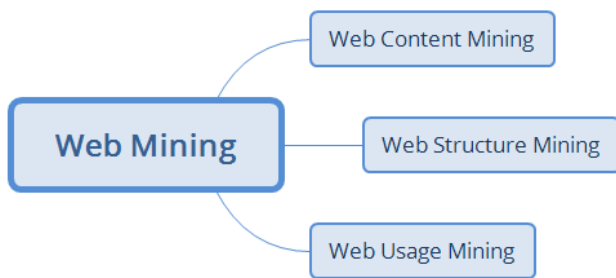


Figure 1. Web Mining

A. Web Content Mining

Web content mining is to focus on discovery of objects in the traditional collection of data, it can be any type of content like text, images, audio, video. It is difficult in mining because the content in web mining can be of semi-structured data or unstructured data.

This method can also be used to detect noise of the web pages that means the irrelevant things scattering on the web pages like ads. Can be detected by this method.

Web Content Mining has been Categorized on the basis of views one is Agent view and another is Database view. The view is to increase the rate in finding data and filtering is called as Agent views. This method is only applied to make the data in structured form. Second View aims on more structured form to apply standard data base.

B. Web Structure Mining

Analysis of web structure comes in web mining. The Hyperlink attached to one web page with the other makes a relationship between both the pages this helps in determining the structure of the web. It helps in discovering useful site for a particular topic. Applying the old traditional approach for our conclusion is wrong process. Instead of that searching for the logic and the linkage can be better than the previous one.

This method can be used for good navigational approach and facilitates the structured technique like

databases. It is feature of content pipeline for portals. The main application of Web structure mining can be used in content management, taxonomies, opinion mining and content generation.

C. Web Usage Mining

Techniques used in analysis and prediction of user behaviour while they are visiting any website of the internet. It helps in pattern of navigation from web data, and predict the important and useful data from a collection of web data.

The process of analysing the user's browsing behaviour is called Web usage mining. It can be regarded as a three-phase process, consisting of the data preparation, pattern discovery and pattern analysis phases [2]. In the first phase, Web data are pre-processed in order to identify users, sessions, pageviews, and so on. The input data are mainly the hits registered in the Web usage logs of the site, sometimes combined with other information such as registered user profiles, referrer's logs, cookies, etc [11].

Web Server Data

The user logs are accessed from Web server. Log data contains IP address, access time and page reference.

Identifying user navigational patterns

The user's activity monitored when browsing through Websites, registered in these sites' Web logs. Considering the average number of visits to a medium-sized Web site per day, we can presume that the amount of information hidden in the site's Web logs is huge, yet meaningless if they're not appropriately processed. By processing these data, either using simple statistical methods, or by using more complicated data mining techniques, we can identify interesting trends, and patterns concerning the activity in the Web site.

Web usage logs

When the Web page accessed is registered in the login log file of the hosting Web server. Records in a Web log file consist of fields that follow a predefined format. Usage data can also be obtained via proxy server logs, browser information, user pages, registration data, cookies, mouse clicks, etc.

Data Pre-processing

The initial task in the pre-processing phase is data preparation. It may be required to clean the web registration data from fields that affect pages that have returned an error or obtain to graphic files. In addition, the crawler activity can be filtered because these entries do not offer such useful information about the functionality of the site. Another problem that must be solved issues caching. Access to cached pages is not registered in the Web registry, therefore, this information is lost.

III. COMPANY WEBSITE

Most companies design and develop their own websites. Without a well-equipped website today, your business is useless. But a poorly designed website, just for the cause of having one, will not provide the purpose. Your website is deserving of all the attention of your company. A well-designed website is one that has a obvious call to action, eye-catching and high quality content, attractive graphic design, easy to use, friendly to mobile devices, etc.

Give your website the much-needed redesign, mobile responsive and search engine friendly websites generate more business. Therefore, we need a strong technical reason on which the websites should be redesigned so that they are more users friendly and efficient. This redesigning is done on the basis of visited web site pages. More important pages must be much easier to locate in comparison to pages having less importance. The importance is calculated in terms of number of page visited during a time period using server log files.

IV. EXPERIMENTAL EVALUATION

In this experiment we have the server log of one month from website www.karzamtech.com for analysing users behaviour of visiting web pages. We have used a tool web log expert, for analysing server log file. Which web page is most visited in a month shows in the given figures, which are helpful determining the position rearrangement of a web site.

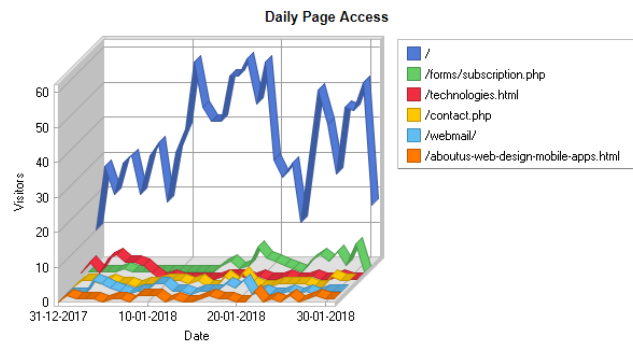


Figure 2. Daily Page access chart

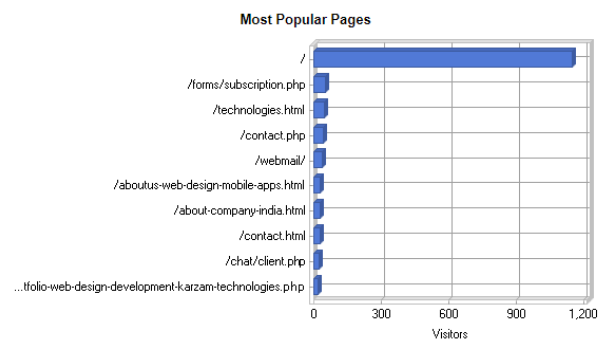


Figure 3. Most Popular Pages

V. OBSERVATIONS

Few observations were prominent from the above experiment. Every website has two parts first is header section and second footer section. Header section is the most visible section in the web site and the links shows in this section have high click chances, while footer have some less chances to click. By using above figure 3 we find out weather the most visited page is in the header section or in the footer section. If it is in the footer section then we redesign our web site structure so this web page in the header section. By this rearrangement other visitor find some easy navigation menu in our website.

HOME ABOUT US SERVICES TECHNOLOGIES PORTFOLIO CONTACT US

Figure 4. Links in Header Section

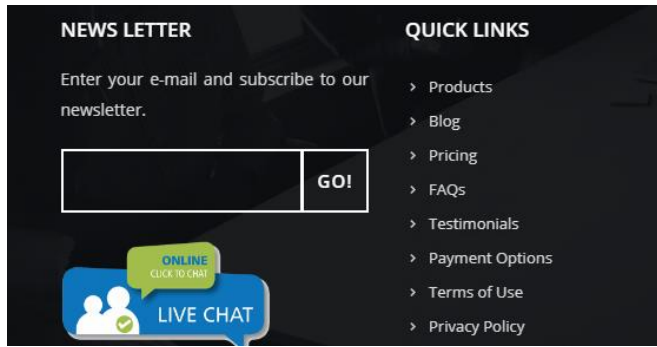


Figure 5. Links in Footer Section

VI. CONCLUSION

As the web and its usage grow continuously, so too grows the possibilities to analyse web data and extract all manner of useful knowledge from it. Using web mining techniques, we know in-depth picture at visitor behaviour on a site. This information allows businesses owners to define their Websites goals with their business objectives for the purpose of identifying areas of enhancement, promoting popular pages of the site, and ultimately increasing visitors and revenue.

With the server web log analysis and experiment we can say that since most of the important pages are designed according to their most frequently page visit. So, we have observation that website structure follows the technical aspect of website redesign.

VII. REFERENCES

- [1] Raymond Kosala & Hendrik Blockeel. Web Mining Research: A Survey. ACM SIGKDD, July 2000.
- [2] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, Pang-Ning Tan, Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data, SIGKDD Explorations, January 2000. Vol. 1, Issue 2, pp. 12-23.
- [3] Da Gomes Jr., M.G. and Z. Gong, 2005. Web structure mining: An introduction. Proceeding of the IEEE International Conference on Information Acquisition, June
- [4] A Review of Methodologies for Analyzing Websites, Danielle Booth, Pennsylvania State University, USA
- [5] P. Ravi Kumar and Ashutosh Kumar Singh, Web Structure Mining-Exploring Hyperlinks and algorithm for information Retrieval, 2010, American Journal of Applied Sciences 7 (6): 840-845
- [6] T. Vijaya Kumar, H. S. Guruprasad, Bharath Kumar K. M., Irfan Baig, and Kiran Babu S., A New Web Usage Mining Approach for Website Recommendations Using Concept Hierarchy and Website Graph, International Journal of Computer and
- [7] Electrical Engineering, Vol. 6, No. 1, February 2014 B. Mobasher, "Web Usage Mining in Data Collection and Pre-Processing," ACM SIGKDD, Ch. 12, 2007, pp. 450-483.
- [8] Gauri Jain, An Intelligent Model for Redesigning Websites using Web Mining Techniques International Journal of Computer Applications (0975 – 8887) Volume 25– No.5, July 2011
- [9] Myra Spiliopoulou, Web Usage Mining for Web Site Evaluation, August 2000/Vol. 43, No. 8 COMMUNICATIONS OF THE ACM
- [10] M Eirinaki, M Vazirgiannis, Web Mining for Web Personalization, in ACM Transactions on Internet Technology (TOIT), 3(1), February(2003).