

Customer Behaviour Analysis using Web Usage Mining

Sonia Sharma^{1*} and Munishwar Rai²

^{1*}Hindu Girls College, Jagadhri, Haryana, India

²MMICT, MMU, Mullana, Haryana, India

**Corresponding Author: soniasharma1980@gmail.com*

Available online at: www.isroset.org

Received: 07/Nov/2017, Revised: 19/Nov/2017, Accepted: 14/Dec/2017, Published: 31/Dec/2017

Abstract- This paper attempts to determine the veiled information and identifying user behavior on the web by using the web data sources. With the help of this information, Prediction of user behavior can be done very easily. The user pattern is analyzed by applying the free tool similar web and page score on the website of the particular organization. This tool tries to conduct web mining in a domain independent manner. The experimental results provide an easier way to navigate the website and improve the website design architecture. This work deliberates the detailed results of a website in a specific education domain application. The main objective of conducting the study on websites of the different universities is to discern the pattern of behavioural expressions and user characteristics so that the organization can improve the structure of website and can upload the more information in a proper way.

Keywords: Web usage mining, Consumer Behaviour, Pattern

I. INTRODUCTION

The popularity of Internet is ever growing and it provides the number of services in every field. One of the most profoundly services of the Internet is the World Wide Web or WWW which has become mainstream tool for almost every activities requiring knowledge base. It not only fulfill all kinds of requirement of different communities whatever a student wants to know more about their particular topic, any university for admission, projects or a wants to purchase something but also it provide help to Internet based business web sites which are facing competition since development of E commerce is growing rapidly and these web-portals generate large amounts of data. This data include various logs and transaction details and with these information, market analyses and predictions can be done very fairly accurately. Many organizations and corporations use web-portals as key tool for sharing and accessing information, and for creating networks/collaborations. In most cases, transactions are encouraged by online transaction process. These are now becoming the main strategy for these organizations and corporate offices. Even the top education organizations have started offering online courses and teaching methods have also been modified with webinars etc. In fact, these online opportunities have revolutionized the education system. This is also true for distance education system and given the multitasking requirement in the current situation, online opportunities meet these requirements with efficiency. Because web process is relatively easy to store and retrieve information, data mining opportunity is also enormous, and it is more beneficial when the logs could also be accessed,

which is useful for research purpose. There has been various research efforts to unearth the quantum of information and the way this could used so as to increase the e-commerce opportunities.

II. DATA MINING PROCESSES

There is always a need to develop data mining technique to extract information from the huge web-log files. These are often automated and systematic process towards of analyses of emerging information and users' response patterns. Given the voluminous data, it is highly challenging task to go into all details of website users.

Data Collection: Log files are usually accessed from various sources such as portal specific servers, open-source portals, personal sites of clients, proxy servers, etc. Web log server provides the most ready source of information relating to websites and user interface. The source files are stored in plain text format and can be accessed by the users.

Preprocessing: This is essentially for removal of click stream data and categorization of transactions in terms of visitation details. Preprocessing is very critical since all servers do not have right or required format.

Data Cleaning: Data cleaning is complex phase in the whole process of data mining from web sources. Because of such voluminous information and has unrestricted opportunities, it is important to remove the noise and misinformation or redundant information. This usually involves removal of file

such as jpeg, word, gif, sound and animation files. Thereafter, IP address is used to identify users and related log files to understand users' behaviors, which include visitation frequency and duration of website access.

Pattern Discovery: Pattern discovery is nothing but studying specific patterns of information storage and retrieval, user behaviors and specific transaction details. This involves summary statistics, pattern detection process and machine learning technique such as artificial neural networks, and clustering algorithm. The emerging pattern discovery includes data mining methods like path identification, Frequent Pattern based Association rule, clustering and classification on preprocessed log data⁷. Clustering approach is used for similarity pattern discovery.

Pattern Analysis: To isolate the predictive emerging patterns from web log files Knowledge discovery technique is used. Pattern analysis is the last stage of web usage mining process. It include the conversion between mined pattern and web log data.

III. TOOLS AND DATA COLLECTION

Number of tools are available to analyze the web log files. Data is collected by using free web analyzer tools. A comparative study is made between the Top most universities of Haryana by knowing the customer interaction between the period Dec 2016 to Feb 2017 on the web sites of the universities mentioned in table 1 by using similar web tool and page score tool. Number of experiments is available to analyze these web server access logs as an input. It generates the reports for frequent access and identifies the user access patterns.

Table 1.

Code Name	University Name
A	Maharishi Markendeshwar University, Mullana
B	Kurukshetra University, Kurukshetra
C	Ch.Devi Lal University, Sirsa
D	Guru Jambheshwar University, Hisar
E	Maharishi Dayanand University, Rohtak
F	Deenbandhu Chottu Ram, Murthal
G	Bhagat Phool Singh Vishwalyadaya, Khanpur Kalan, Sonipat
H	YMCA University of Science and Technology, Faridabad

Table 1 Shows the Universities name and its code which is temporary assigned for experiment purpose

Table 1.1

Code Name	Ranking	
	Global Rank	Country Rank
A	657947	27617
B	89270	3552
C	458553	18602
D	349861	16700
E	102399	4097
F	875610	36476
G	2539932	103937
H	690777	32654

Table 1.1 gives the comparative Global Ranking and county Rank of the Universities Analysis Report from similar web and page score.

Table 1.2

Code Name	Total Visit (On desktop and mobile Phone)
A	34.40K
B	444.60k
C	53.10k
D	104.70k
E	449.30k
F	29K
G	5.70k
H	38.70k

This table shows the how many visitor visit the websites of the Universities on Mobile phone and Desktops to find the information.

Table 1.3

Code.	Loading Time (in Sec)
A	4.46
B	23.81
C	27.52
D	80.29
E	17.01
G	73.12
H	4.24

This tables shows that when a customer browse the website of a particular universities then how much loading time it takes.

Table 1.4

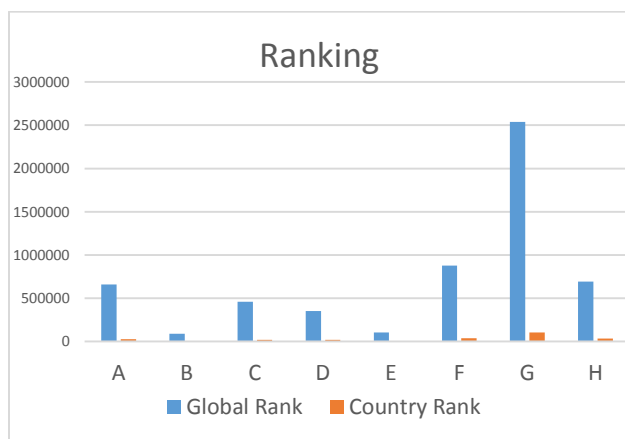
Code.	Direct	Referrals	Search	Social	mail
A	15.85%	6.36%	77.64%	0.15%	0.00%
B	31.50%	8.08%	59.57%	0.68%	0.17%
C	20.28%	4.51%	71.98%	0.00%	3.23%
D	30.38%	9.05%	60.35%	0.22%	0.00%

E	35.94%	5.72%	56.93%	0.52%	89%
F	14.64%	8.09%	76.35%	0.92%	0.00%
G	23.81%	13.89%	61.65%	0.66%	0.00%
H	21.15%	2.40%	75.65%	0.81%	0.00%

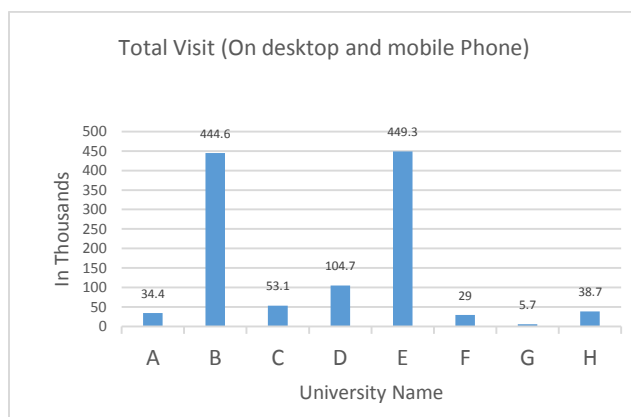
This table shows what traffic sources is used by the customer to reach a particular website of said universities.

IV. TOOLS ANALYSIS REPORT (RESULTS AND FINDINGS)

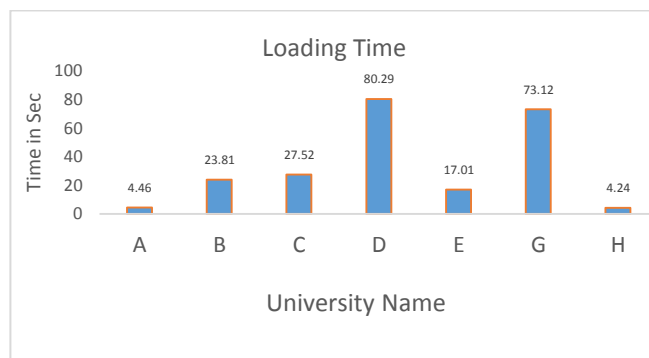
This work used the data collected from the web Sites of the Top Universities of the Haryana from 7 Dec 2016 to 31 Jan 2017. This collected data is analyzed by using Different free Tools Such as Page Score and Similar web. The complete experimental analysis was done on the basis of web log data of the educational Universities website and a comparative analysis report is prepared. The design and execution of such work is restricted and time consuming. Comparative chart of educational websites of universities is



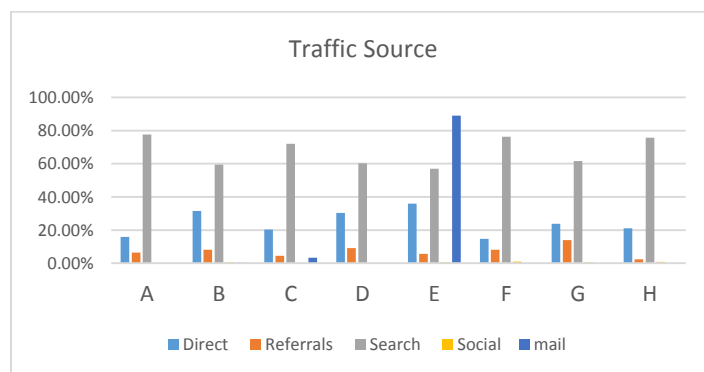
Graph 1.



Graph 2.



Graph 3



Graph 4

Findings: (7 Dec 2016 to 31 Jan 2017)

1. Fig 4 clearly shows that University code H i.e. Bhagat Phool Singh University has good Global and Country ranking approaching that maximum customer visit this university website as compared to other universities.
2. Website loading time of university A i.e. MMU Mullana is less as compared to other universities
3. 449.3K customers visit the MDU Rohtak website during Dec 2016 to Jan 2017.

V. CONCLUSION

As there is increase of tremendous use of website for online purchasing and for everything which a customer wants with the click of mouse. Web usage mining is the one of the important research areas through which browsing and navigation of data can easily be predicted. There are several techniques and tools proposed by different researchers for the web usage mining. This paper discussed about education domain web user pattern discovery and pattern analysis using similar web and page score. The results obtained should absolutely help the website Analysts, Website Maintainers, Website Designers.

REFERENCES

- [1] Li Yong-hong; Liu Xiao-liang, "*Research of data mining based on e-commerce*," Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on , vol.4, no., pp.719,722, 9-11 July 2010
- [2] Chaoyang Xiang; Shenghui He; Lei Chen, "*A Studying System Based on Web Mining*," Intelligent Ubiquitous Computing and Education, 2009, International Symposium pp.433,435, 15-16 May 2009
- [3] Li Mei; Feng Cheng, "*Overview of Web mining technology and its application in e-commerce*," Computer Engineering and Technology (ICCET), 2010, 2nd International Conference on , vol.7, no., pp.V7-277,V7-280,16-18 April 2010
- [4] Yadav, M.P.; Feeroz, M.; Yadav, V.K., "*Mining the customer behavior using web usage mining in e-commerce*," Computing Communication & Networking Technologies (ICCCNT), 2012 Third International Conference on , vol., no., pp.1,5, 26-28 July 2012 .
- [5] Zhiwu Liu; Li Wang, "*Study of Data Mining Technology Used for E-commerce*," Intelligent Networks and Intelligent Systems (ICINIS), 2010 3rd International Conference on , vol., no., pp.509,512, 1-3 Nov. 2010
- [6] S.Yadav, K.Ahmad, J.Shekhar, "*Analysis of Web Mining Applications and Beneficial Areas*", International Journal of IUM, vol.12, Issue 2, 2011.
- [7] Shen Zihao; Wang Hui, "*Research on E-Commerce Application Based on Web Mining*," Intelligent Computing and Cognitive Informatics (ICICCI), 2010 International Conference on , vol., no., pp.337,340, 22-23 June 2010
- [8] Tang, Hewen; Yan, Honglin; Zengfang Yang; Yu Ma; Chunping Li, "*Application of Data Mining in Electronic Commerce*," Information Engineering, 2009. ICIE '09. WASE International Conference on , vol.1, no., pp.631,634, 10-11 July 2009

About the Author

Sonia Sharma : is working as Asst. Prof with Dept.of C.Sc.& Applications, Hindu Girls College Jagadhri ,India. Currently she is pursuing Ph.D from Maharishi Markendeshwar University, Mullana (Ambala), Haryana ,India. She has published papers in National and International Conferences. Research Interest Include: Data Mining, ICT .e-Governance.



Dr. Munishwar Rai : is currently working as Associate professor at MMICT &BM (MCA Department) of Markendeshwar University, Mullana (Ambala), Haryana ,India. He has around 19 years of teaching experience .His area of Interest is Software Reusability, Data Mining and Computer Networks. He has published more than 20 papers in National and International level.

