RESEARCH ARTICLE   OPEN ACCESS

An Integrated Method to Improve Web Usability and Web Server Performance: A Survey

Abhishek Lavale [1], Ravi Hemraj [2]
Department of Computer Science and Engineering
Shri Balaji Institute of Technology & Management
Betul, RGPV University
M.P. - India

ABSTRACT

As per the World Wide Web turn out to be biggest today, building and ensuring easy-to-use Web based organizations is becoming a necessary proficiency for commercial existence. **In the real World Wide Web, web sites need navigation. I**t provides users an idea of the scope of the web site, it is a great fallback system, it decreases users' intellectual overhead. Though, noteworthy challenges occur, comprising correctness of problem identification because of false alarms collective in skilled evaluation, impracticable assessment of usability because of dissimilarities between the testing and actual usage environment, amplified cost because of the extended maintenance cycles and evolution distinctive for many Web based applications Log file comprise information about date, user name, time, IP address, access request and bytes transferred. Server based logs have similarly been used by most organizations to acquire knowledge about the usability of their valuable products. For instance, search based queries can be mined from server based logs to determine user info desires for usability task investigation. Server based Logs can also provide awareness into actual users executing real tasks in usual working circumstances versus in a simulated situation of a research lab. Usability is defined as the efficiency, satisfaction, and usefulness through which actual users can comprehensive specific tasks in a specific environment. The paper provides the survey to improve web usability and performance of the web server.

Keywords:- Data mining, Web Server Logs, Usage Mining, Proxy Server, Web Navigation

I. INTRODUCTION

Internet[1] has in recent times become a dominant platform for repossessing valuable information. Web can also be used for find out information from Internet based data. The concept of discovering useful pattern on the data has been given a verity of name like knowledge extraction, data mining, information discovery and data pattern processing. Mining the web data[2] is the furthermost inspiring tasks for the data management and data mining scholars because there are less structured, huge heterogeneous data available on the web. Web server maintains web log files. Log file are positioned in different locality identical to web based server, web based proxy server and user browser. Preprocessing log files have different steps including data cleaning, user and, session identification. Data cleaning in this step is removing unnecessary data and noise.Web based server log file is a simple unicode file which stored information almost every user. Log file

contain information about bytes transferred, user name, Internet Protocol address, time, date, access request.

Though users incline to circumnavigate web sites links embedded in actual content, by search mechanisms or by website navigation serves useful purposes. Users can reach at any web page inside a web site by a numeral of means: a search engine, , typing a web address directly, a link, a software program's about box, a directory and so on. Therefore, web site pages can't be isolated from the rest of the web site if the site is to be effective and have lots of valid users. In the real World Wide Web, web sites need navigation. It provides users an idea of the scope of the web site, it is a great fallback system, it decreases users' intellectual overhead.

Usability[3] is defined as the efficiency, satisfaction, and usefulness through which actual users can comprehensive specific tasks in a specific environment. Web design principles[4] are firmness, structural functional convenience, and presentational pleasure. Structural firmness relates primarily to the features that

encouragement the web site security and performance. Functional convenience mentions the accessibility of suitable characteristics, such as a web site's ease of navigation, and ease of use and that help users' communication with the GUI interface. Presentational delight talk about the website features that motivate users' wits. Client side logs and Server side based logs are used for Web usage[5] and usability analysis. Server side based logs can be automatically produced by Web based servers like Apache web server, with each entrance matching to a user demand. By analyzing these server logs, Web capability was considered and used to recommend performance heightening for Internet Web based servers. Server based logs have similarly been used by most organizations to acquire knowledge about the usability of their valuable products. For instance, search based queries can be mined from server based logs to determine user info desires for usability task investigation. Server based Logs can also provide awareness into actual users executing real tasks in usual working circumstances versus in a simulated situation of a research lab. Server based Logs also characterise the all the actions of several users over a lengthy period of time vs the minor example of users in a little time duration in typical research lab testing.

Paper is organized as follows. Section II provides background related to web mining, navigation and web usability, web server log files. Section III provides literature survey of web mining and usage mining.Section IVconcludes the paper.

II. BACKGROUND

As per the World Wide Web turn out to be biggest today, building and ensuring easy-to-use Web based organizations is becoming a necessary proficiency for commercial existence. Due of the immensely un even Web data traffic, enormous user populace, and miscellaneous usage environment, coverage constructed testing is unsatisfactory to guarantee the quality of Web based applications. By means of simple log cleanser filtered data is irrelevant and valuable, some preserve alive link add period stamp into their URL. So their precede cannot be further added openly to prefix library simply by threshold. Besides the design of threshold and estimation method of precision rate is irrelevant simple. There have different data

cleaning[6] method of server log is still very enlighten. Some researchers also focus on data cleaning method of proxy log and describe the difference between proxy log and server log. One thing is noticeable that Standard Filter is impossible to filter out any relevant item. Web design values were acknowledged to assistance progress users' online involvement. Experimental estimation by professionals and user-centered analysis are characteristically used to categorize usability matters and to guarantee agreeable usability. Though, noteworthy challenges occur, comprising correctness of problem identification because of false alarms collective in skilled evaluation, impracticable assessment of usability because of dissimilarities between the testing and actual usage environment, amplified cost because of the extended maintenance cycles and evolution distinctive for many Web based applications.Web based log file investigation began as a way for Information Technology managers to guarantee satisfactory server capacity and bandwidth on their company web sites. Organizations are nowadays looking for ways to use log based files to acquire about the usability of their web sites—that is, how successfully web visitors meet their particular transaction goals or information there.

Log[7] file comprise information about date, user name, time, IP address, access request and bytes transferred. A Website log is a text file to which the Web based server marks information to each time a user demands a resource from that certain web site. Once user give in to demand to a web based server that movement are logged in web log file. Log bases text file range 1KB to 100MB. Log based file contributes information around:

1. Which web pages were demanded in website?
2. Just how numerous bytes directed to user from web server?
3. What category of error happens?

When user give in to request to a web based server that movement are recorded in web based log text file. Log based text file used for fixing errors or debugging purpose. Investigating log based file are used to perceiving spasms on web.

Example of web log file is given below.

a) 192.168 .198.23: IP address that can be converted to host name.

b) - : The name of the remote user

c) - : Login of the remote user. Both name of remote user and login of remote user usually omitted and replaced by a dash "-"

d) [17/Dec/2016:12:23:32 -0600]:

represents Date: DD/Mon/YYYY, next Time: HH:MM:SS next Time Zone: (+|-) HH00

Relative to GMT -0500 is US EST

e) GET /jobs/ HTTP/1.1:

Method: GET and POST are methods.

URL: relative to domain,

HTTP protocol: protocol version e.g. HTTP/1.0 or HTTP/1.1

Location of Log File

Web based log file is located in three different locations.

1. **Web based server logs**: Web based log files deliver most complete and accurate usage of data to web based server like Apache. The log based file do not record user visited cached pages. Data available in the log based files are sensitive and personal information so web based server retains them closed.

2. **Web based proxy server**: Web based proxy server[8] receives HTTP request from web user, gives them to web based server, then outcome passed to web based server and return to web user. Client browser send request to web based server via web proxy server.

   The two disadvantages of web based proxy server are:

Proxy based server creation is a tough job. Most advanced network based programming, such as TCP/IP or UDP, is compulsory for this development.

The request intervention is incomplete.

3. **Client browser**: Log based file can exist in client's browser[9] window itself. HTTP cookies are used for saving logs in client browser. These HTTP cookies are sections of data or information produced by a web based server and kept in user's computer system, prepared for upcoming use.

The main objectives is to survey the different web server usability technique which can be used to create improved system with large web application.The different approach of web server usability systemto discover web usage pattern in an application and to improve the web user's overall satisfaction in web usage mining.

III. LITERATURE SURVEY

Proposed method in[8]prominence on recognizing celestial navigation related difficulties as regarded as by an helplessness to complete definite tasks or unnecessary period to complete them. The advised scheme distinguish website navigation accompanying usability complications by equating Website constructed usage patterns take out from website established server logs against predicted usage embodied in some intellectual user models. Usability engineers often use server logs to analyze users' behavior and appreciate how consumers accomplish definite tasks to expand their experience. The main steps in system is Web Data Preparation for mining and Pre-processing. Next is data cleaning i.e. removing extraneous graphics, references to style links, or audio files that might not be essential for the purpose of analysis. Then next is user identification step in which referrer fields, user agent, and IP address to identify unique users. Next process is user session identification. Path completion or misplaced references can habitually be heuristically contingent from the information of site topology and referrer info, sideways with time-based info from server constructed logs Perfect user collaborating path models detention projected Web based usage. Planned architecture consist of three modules IUIP modelling, Usage Pattern Extraction, and Usability problematic identification. Consumer pattern withdrawal component remove actual direction finding paths from server logs and find out patterns for nearly characteristic actions. In equivalent, IUIP models for the equal events. IUIP models are built on the perception of user behavior and can characterize estimated paths for detailed user-oriented tasks. The outcome examination employs the contrivance of test oracle. A revelation is commonly used to decide whether a test has failed or passed. Here, this models used the revelation to recognize the usability problems associated to the users' genuine navigation paths by analyzing the unconventionalities between the two.

Website server logs are foremost data source. Individually entry in a log encompasses the timestamp, the IP address of the inventing host, the referrer, the requested Website page, the user agent and other related data. Characteristically, the raw information

necessity to be pre-processed and transformed into user transactions and sessions to excerpt usage patterns.

A new technique to recognize navigation- related Web based usability difficulties based on show a relationship between Actual and Expectable usage data patterns. The tangible usage data patterns can be mined from Internet based server logs regularly recorded for functioning websites by first handling the log based data to user sessions, identify users, and consumer task focused transactions, and then smearing web usage mining procedure to determine data patterns amongst actual usage pathways. The expected website page usage, together with info about both the path and time essential for user focused tasks, is taken by perfect user interactive track models created by intellectual experts based on their reasoning of user behavior. The assessment is accomplished through the method of test oracle for examination of results and recognizing user navigation problems. The deviation web data produced from this assessment can help us find out usability related concerns and recommend corrective activities to increase web based usability. A program implemented to automate a noteworthy part of the actions involved. With the help of experimentation on a lesser service oriented web site, system identified web usability difficulties, which were cross validated by field specialists, and enumerated usability development by the lower time and effort and higher task success rate for given tasks after recommended improvements were implemented.

The [9] represent a new technique and tool for activity demonstrating through qualitative successive data item analysis. In specific, address the problem of creating a symbolic abstract illustration of an activity from an action trace. To use information engineering methods to help the analyst construct an ontology of the activity, i.e., a set of hierarchical semantics and graphical symbols that supports the construction of action models. The ontology building is evolutionist, pragmatic and driven by the analyst in accordance with their modeling objectives and their research based questions. The scheme helps the analyst to define transformation guidelines to process the raw data trace into abstract traces created on the ontology. The analyst imagines the abstract traces and repeatedly tests the transformation rules, ontology, and the visualization presentation to approve the models of action. With this method, and tool found pioneering ways to represent a motor car driving movement at dissimilar levels of abstraction from activity traces composed from an instrumented means of transportation. As samples, report two new policies of track changing on motor ways that modeled and found with this approach.

[10] Defines an examination to determine the technical possibility of identifying and discovering the several situations experienced by actor or human absolutely from a trace of period stamped data values of variables. More precisely, the objective of investigation was to determine the circumstances that a human actor practiced, while performing a strategic task in a simulator based situation, the arrangement of these circumstances and their time-based interval. The significant variables that were witnessed in the trace were designated apriorithru a human. The conclusion of the procedure was matched with human or manual context utilization of the similar traces. One probable usage of such automated background finding is to assist for building self-directed strategic agents proficient of performing the similar tasks as the actor or human. As such, similarly quantitatively compared the outcomes of with the help of the COPAC-derived circumstances with those achieved with human derived context utilization in constructing self-directed strategic agents.

Modern Web based applications are fully, difficult software systems. Consequently, the development of Web based applications needs a procedurally comprehensive engineering methodology called Web Engineering. It is not very clear, however, to which level present solutions from appropriate areas, most remarkably software engineering can be use again as such for the improvement of Web applications and subsequently, if Web Engineering is really a discipline on its own. [11]highlights the characteristics of Web based application improvement as found in current literature thus provided that a requirement for analyzing the suitability of present engineering solutions. The characteristics are characterized according to comprising the software product itself, four dimensions, its improvement, its use and advancement as a cross cutting concern.

[12] refer to the dissimilar features between proxy based log and server based log, thus deliver a data cleaning technique for enterprise based proxy. Though the assessment of proxy based log and server based log is considerable, they use plenty of experimental value as the threshold in the experimentation without

theoretical or description support. It also creates the experimental outcome doubtful.

A significant aspect that influences the efficiency of security schemes within an association is the web usability of security administration tools. [13]present a review of design strategies for such procedures. [13] collected recommendations and guidelines related to IT security administration systems from the literature.

Delivery of well-organized service with the help of a web site makes it essential in the restructuring phase to take into account the activities of the consumers, which can be considered by using of a web based log file that to some extent records valuable information about user visits. The re-establishment of all of the arrangements of website pages that are visited by consumers who visit a website page is known as the web session nization difficultly, and it has been expressed by means of programming; however, because a web based log can gather a great amount of valuable information, it is essential to rebuild the sessions above a duration of months, or weeks thus the resolution to this difficulty requires a long processing time. [14] Provides a heuristic methodology constructed on simulated strengthening for the session nization difficulty. Using this methodology, it has been possible to decrease the processing time up to 170 times associated to the time that is essential for the programming. Moreover, the metaheuristic explanation finds new finest values, which accomplish increases on the order of 18% in the best cases.

In [15] discuss the dissimilar components of the gaming user experience and their relation to additional conceptions within HCI. The aim is to showcase how the dissimilar essentials that form the user gaming involvement can also be used to appreciate extra concerns about user experience. The aims of games are, to deliver players with a constructive knowledge. Understanding the components that finally lead players to have a more positive experience should deliver comment about the wider user experience perception. Although non-game and video-games applications appear to be two dissimilar areas of research, in relationships of involvement, they both game to increase the individual's experience.In first reference the real usage information patterns can be mined from Web based server logs consistently recorded for operating web sites by first handling the log based data to user sessions, identify users, and web user task oriented transactions, and then applying a web usage discovering system to determine data patterns among actual web usage paths. The associated failure info is used to measure the consistency of Web based applications and the probable efficiency of statistical Web testing.

IV. CONCLUSION

Web has in recent era come to be a dominant platform for not only accessingdigital information but also determining knowledge from web data. The web data is stored as unstructured as well as structured format. The logs present in Web servers represent actual usage of the web sites. The file called log file comprise information about bytes transferred, Internet Protocol address, name of the user, date, access request time. The real usage values can be mined from different Internet server logs consistently recorded for functioning websites by primary processing the log value to identify, user sessions, users, and user task-oriented transactions. The paper review the web mining, usability and performance of web server.

REFERENCES

[1] R. Cooley, B. Mobasher, and J. Srivastava,, "Data preparation for mining World Wide Web browsing patterns," Knowl. Inf. Syst., vol. 1, no. 1, pp. 5–32, 1999.

[2] M. F. Arlitt and C. L. Williamson, "Internet Web servers: Workload characterization and performance implications," IEEE/ACMTrans. Netw., vol. 5, no. 5, pp. 631–645, Oct. 1997.

[3] C. M. Barnum and S. Dragga, Usability Testing and Research. White Plains, NY, USA: Longman, Oct. 2001.

[4] M. C. Burton and J. B. Walther, "The value of Web log data in use-based design and testing," J. Computer.-Mediated Commun., vol. 6, no. 3, p. 0, 2001.

[5] T. Carta, F. Patern`o, and V. F. D. Santana, "Web usability probe: A tool for supporting remote usability evaluation of web sites," in Human-Computer Interaction—INTERACT 2011. New York, NY, USA: Springer, 2011, pp. 349–357.

[6] R. Cooley, B. Mobasher, and J. Srivastava,, "Data preparation for mining World Wide Web browsing patterns," Knowl. Inf. Syst., vol. 1, no. 1, pp. 5–32, 1999.

[7] Hongzhou Sha, Tingwen Liu, Peng Qin, Yong Sun, Qingyun Liu, EPLogCleaner: Improving Data Quality of Enterprise Proxy Logs for Efficient Web Usage Mining, Elsevier International Conference on Information Technology and Quantitative Management, 2013, pp- 812-819

[8] RuiliGeng, Member, IEEE, and Jeff Tian, Member, IEEE, Improving Web Navigation Usability by Comparing Actual and Anticipated Usage, IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS, VOL. 45, NO. 1, FEBRUARY 2015, pp-84-95

[9] Olivier L. Georgeon, Alain Mille, Thierry Bellet, Supporting Activity Modeling from Activity Traces, 2013, pp- 201-237

[10]Viet C. Trinh and Avelino J. Gonzalez, "Discovering Contexts from Observed Human Performance ", IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS, VOL. 43, NO. 4, JULY 2013, pp-359-471

[11] GertiKappel, Elke Michlmayr, Birgit Pröll, Siegfried Reich4, Werner Retschitzegger5 Web Engineering - Old wine in new bottles?, White paper, 2—3, pp-25-31

[12] Y. Zhang, L. Dai, Z. Zhou, A New Perspective of Web Usage Mining: Using Enterprise Proxy Log, in: Proceedings of the 2010 International Conference on Web Information Systems and Mining (WISM), Vol. 1, IEEE, 2010, pp. 38–42.

[13] PooyaJaferian, David Botta, Fahimeh Raja, Kirstie Hawkey, Konstantin Beznosov, Guidelines for Designing IT Security Management Tools, pp 221-230

[14] Tec-Ed, "Assessing web site usability from server log files," White Paper, Tec-Ed, 1999.

[15] Eduardo H. CalvilloGámez, Paul Cairns, Anna L. Cox, From the Gaming Experience to the Wider User Experience, *BCS HCI* '09, 24,29