



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

A Review on Different Techniques Used for Analysis of Web Usage Logs and Web Navigation

Vrushali P. Bonde, Prof. K. K. Chhajed

Dept. of CSE, P.R.Pote (Patil) College of Engineering and Management, Amravati, India

Asst. Professor, Dept. of CSE, P.R.Pote (Patil) College of Engineering and Management, Amravati, India

ABSTRACT: As the World Wide Web becomes important part of millions of users today, building and ensuring easy-to-use Web systems is becoming a core competency for business development and survival. On any web system search, navigation, and transaction are important activities showing its usage. Navigation, especially, is very important after users identify the corresponding Website using search engines and before make transactions. Since the success of Internet enterprises depends on the usability of services offered on Websites, it is important that the Web pages follow usability principles. This paper provides the brief review on methods to identify navigation related Web usage identification and associated usability problems. There are different techniques mention in this paper that helps to identify the web usage by mining from large amount of web usage data. This identification is performed by comparing actual and anticipated usage patterns. The techniques suggested by different researchers are reviewed here all of them having their own advantages and limitations. But this paper suggest that the server logs and navigation pattern are helpful to users and business persons to provide better effectiveness and efficiency in developing their business.

KEYWORDS: Ideal user interactive path (IUIP), cognitive user model, location information, usage pattern, web server log, web and navigations.

I. INTRODUCTION

With the increasing use of Internet, people have got the platform to acquire knowledge & explore the information. In order to satisfy the increasing demands from online customers, organizations both small scale and large scale are heavily investing in the development and maintenance of their websites. With the help of Internet, people have got the platform to acquire knowledge & explore the information. E-business is gaining popularity as the feature of shopping from home concept is growing on increasing and is providing huge amount of profit to all the companies [1]. As the users of internet reaches to the millions, it becomes necessary that Web content needs to be filtered and personalized based on the particular needs of individual users. While designing attractive and personalized interactive system it is necessary to consider the user's interests, expectations and expertise, cognitive style and perception. This will help in providing useful and important services to all.

The continued growth and proliferation of e-commerce, Web services, and Web-based information systems, the volumes of clickstream and user data collected by Web-based organizations in their daily operations has reached astronomical proportions. Analyzing such data can help these organizations for determining the life-time value of their clients, design cross-marketing strategies across products and services, evaluate the effectiveness of promotional campaigns, optimize the functionality of Web-based applications, provide more personalized content to visitors, and find the most effective logical structure for their Web space [2]. This type of analysis involves the automatic discovery of meaningful patterns and relationships from a large collection of primarily semi-structured data. These data is often stored in Web and applications server access logs, as well as in related operational data sources mainly related to the users.

Web Navigation is the study and the process of monitoring and controlling the movements of the users that are moving on the web from one place to the other place on time. Web navigation refers to the process of navigating a network of information in the World Wide Web, which is organized as hypertext or hypermedia. This plays a vital role



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

for each net user. In many cases web sites are built according to the comfort of the users point of view with a good navigation access and so as to make the website looks attractive and helpful to the users. Sometimes users find navigation disturbance when they use. But as the navigation are important from both users and creators point of views, it is necessary to avoid these kinds of disturbances. For getting the important users behaviors the navigation related web usage should be identified properly. Web usage mining concept helps the user to mine the knowledge from the web log files. This mined set of information will be helpful for finding out the user navigation pattern. Extracted user navigation pattern may help the actual navigation path from the server log files. They mainly focus on the improvement of the web application and their quality.

One another concept called as Usability Engineering is a discipline that provides structured methods for achieving usability in user interface design during the development of application. User Interface Design describes any kind of interface design of an interactive system and also the information design of non-interactive systems. Usability engineering provides methods for measuring usability and for addressing usability issues. Heuristic evaluation by experts and user-centered testing are typically used to identify usability issues and to ensure satisfactory usability. However there are significant challenges exist to this as 1) accuracy of problem identification due to false alarms common in expert evaluation, 2) unrealistic evaluation of usability due to differences between the testing environment and the actual usage environment, and 3) increased cost due to the prolonged evolution and maintenance cycles typical for many Web applications [3]. On the other hand, log data routinely kept at Web servers represent actual usage. Such data have been used for usage-based testing and quality assurance, and also for understanding user behavior and guiding user interface design.

For extracting the useful and important users visiting behaviors and navigation patterns, Web Usage mining applies data mining technique to extract knowledge from these web log files [4]. Additionally, various tools and techniques can be used to extract the information from these raw log files. The extracted information can then be used for finding user navigation patterns. By finding frequent user navigation sequences or user navigation sessions from server logs, one is able to compare the actual usage behaviors of users with expected navigation trails and try to improve the interface of the site accordingly. This deviation analysis would also help us to identify some navigation related usability problems.

One another important feature is that web application can filter data location wise which will maintain the usability of user with location and which can be latterly used for business intelligence development. This method would definitely complement traditional usability practices and overcome some of the existing challenges. With this data one is able to perform usage-based testing and quality assurance and understanding user behavior and guiding user interface design, which improves the growth rate of organizations. The remaining paper is organized as, Section II gives the literature study consisting of different methods used by different researchers for identifying web navigation and mining web logs data. Section III represents and describe different techniques used for mining web server logs. Finally section IV, concludes the paper.

II. LITERATURE REVIEW

As the web logs created on the server based on the users visiting behavior are serves to be important parameter for online business development. Different researchers develop different methods and techniques for proper mining of these client and server side logs with which one can get the users visiting behaviors. Some of these work are summarized as follows:

The authors Ruili Geng and Jeff Tian have developed a new method for the identification and improvement of navigation-related Web usability problems by checking extracted usage patterns against cognitive user models. The methods suggested by author can identify areas with usability issues to help improve the usability of Web systems. Once a website is operational, our method can be continuously applied and drive ongoing refinements. the developed method is cost-

effective. Here author, currently integrating these modeling and analysis tools into a tool suite that supports measurement, analysis, and overall quality improvement for Web applications. But the limitations here that needs to overcome are necessity of validations and to explore additional approaches to discover Web usage patterns and related usability problems generalizable to other interesting domains. The model develop here need to be further enhanced and



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

optimized for more complex tasks and usability research to cover more usability aspects to improve Web users' overall satisfaction.

The authors May Wang, Benjamin Yen [7] performs study with the goal of improving Web navigation efficiency by reorganizing Web structure. For avoiding users to lose their orientation, structure stability is taken into consideration. The authors proposes a mathematical programming method to reorganize Web structure in order to achieve better navigation efficiency. Here, user can specify the requirements and how stable the website structure should be. Navigation efficiency is defined systematically. The advantage of this work is improving navigation efficiency mathematically and relieving the designer of tedious chore to modify the structure in transformation.

The authors N. Sanfia Sehnaz, Dr.I.Elizabeth Shanthi in this paper discusses strategies pertaining to identification and improvement of navigation and its related web usability [8]. It helps researchers for better effectiveness in task completion and improved efficiency in less time. The actual and anticipated usage behavior will helpful to the users for the better effectiveness of completion rate and efficiency for given tasks. This research work can be further improved using data mining techniques such as classification, clustering embedded with soft computing techniques.

The author Hitesh Hasija in this paper [9], combines the three basic fields of web mining i.e. web content mining in the form of keywords and description, web usage mining in the form of time spent by a user on web page and frequency to access it and web structure mining in the form of consideration of hyperlinks. Hence, web navigation with combination of all the three aspects is rare. According to author, implementing an optimization technique to provide user navigation is a difficult task because of the user's dynamic nature. The model develop is converted to graph form and ant colony optimization is applied over it, by defining a suitable heuristic function and optimum results are obtained within a specified time constraint. But authors said that the techniques needs to be further expand to be brought by different organizations in future to convert it into a recommendation system at a very broad level in future.

The authors T. Arce, P. E. Román, J. D. Velásquez, and V. Parada, presents a heuristic approach based on simulated annealing for the sessionization problem [10]. Using this approach, it has been possible to reduce the processing time up to 166 times compared to the time that is required for the integer programming model. But, web browser navigation is becoming hard to track the information due to the multiple ways that pages are cached, loaded, and navigated. Multitab navigation, back button browsing and history jumps are commonly not reflected on web logs.

The authors Hamilton Fernandes de Moraes Junior, Fabia Lika Nishida and Ana Cristina Vieira de Melo [11] have proposed a framework for modelling websites in order to evaluate them according to Nielsen's Heuristics. Modelling websites with the proposed framework can also help experts to evaluate the navigation issues of new websites design projects or existing websites. Here there is necessary to intend to study how to relate the corresponding complexity, the proposed framework with existing models for user's interactions and how to use existing tools to collect use of sites information and define a more formal model for web elements to automate certain usability testing evaluation.

The authors F. E. Ritter, A. R. Freed, and O. L. Haskett proposed that Server logs can also been used by organizations to learn about the usability of their products [12]. For example, search queries can be extracted from server logs to discover user information needs for usability task analysis. But they focus only on task analysis, and does not cover design elements of web sites. Also in the list of type of users that they have identified, some user will fit under multiple categories. This problem they have not been solved.

The authors Tonio Carta, Fabio Paterno, and Vagner Figueredo de Santana presented a tool that supports remote usability evaluation of Web sites [13]. The tool considers client-side data on user interactions and JavaScript events. In addition, it allows the definition of custom events, giving evaluators the flexibility to add specific events to be detected and considered in the evaluation. The tool supports evaluation of any Web site by exploiting a proxy-based architecture and enables the evaluation to perform a comparison between actual user behavior and an optimal sequence of actions. Many problems was analysed like user is not getting accurate information, there was lack of user guidance, etc. But these problems were not solved.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

The authors Shalaka S. Thorat, Priyanka More in this paper [14], proposed a mathematical programming model which is used to improve the navigation effectiveness of a website while minimizing changes to its current structure. The model develop is particularly appropriate for informational websites whose contents are relatively stable over time. It improves a website rather than reorganizes it and hence is suitable for website maintenance on a progressive basis. Optimal solutions were quickly obtained, suggesting that the model is very effective to real world websites. In addition, authors tested the MP model with a number of synthetic datasets that are much larger than the largest data set. The result obtained confirm that the improved structures indeed greatly facilitated user navigation. But this model need to be improved further by incorporating additional constraints. As by using different and useful data mining methods these constraints are identified.

The authors Luis Rivero, Marcos Kalinowski, Tayana Conte in this paper [15], presents practical findings from applying proposed Design Usability Evaluation (DUE) technologies. Such technologies allow the usability inspection of Web applications mockups to reduce the cost of identifying and fixing navigation and ease of use problems early in the development process. Here an empirical study is planned to evaluate the capability of applying these technologies without training, and the perceived ease of use and utility of the technologies, using the Technology Acceptance Model. These all problems are mainly associated with user interaction so it is necessary to avoided and there is need of improvement in Web flow.

III. IMPORTANT RELATED TECHNIQUES

A. WEB LOGS:

Web logs contains the information of visiting behavior of different web pages. These logs are commonly used for Web usage and usability analysis. There are mainly two types of logs calculated as server-side logs and client-side logs.

1. Server-side Logs: Server-side logs can be automatically generated by Web servers, with each request comes from the user to the server. A server log is a log files automatically created and maintained by a server mainly consisting of a list of activities it performed. A typical example is a web server log which maintains a history of page requests comes on it. Information about the request present in the log files are client IP address, request date/time, page requested, HTTP code, bytes served, user agent, and referrer. One important thing is these server logs typically do not collect user-specific information. These logs information is useful for analyzing these logs, Web workload was characterized and used to suggest performance enhancements for Internet Web servers [5]. Instead these log file are generating automatically based on the users request, these files are usually not accessible to general Internet users, and accessible only to the webmaster or other administrative persons.

There are many advantages to using server logs as, Server logs have also been used by organizations to learn about the usability of their products. Logs can provide insight into real users performing actual tasks in natural working conditions versus in an artificial setting of a lab. Logs also represent the activities of many users over a long period of time versus the small sample of users in a short time span in typical lab testing. Data preparation techniques and algorithms can be used to process the raw Web server logs, and then mining can be performed to discover users' visitation patterns for further usability analysis. A statistical analysis of the server log may be used to examine traffic patterns by time of day, day of week, referrer, or user agent. Marketing departments of any organization that owns a website should be trained to understand these powerful tools, leading to help in business development of that organizations.

2. Client-side logs: As the name suggest, Client-side logs can capture accurate comprehensive usage data for clients based on their visiting behaviors on the web pages. These client-side logs allow low-level user interaction events such as keystrokes and mouse movements to be recorded [5]. Such data are often used with task-based approaches and models for usability analysis by comparing discrepancies between the designer's anticipation and a user's actual usage behavior. For calculating this, the evaluator must program the UI, modify Web pages, or use an instrumented browser with plug-in tools or a special proxy server to collect such data. For the purpose security, users generally do not want any instrument installed in their computers. This client-side logs calculation and logging actual usage on the client side, can best be used in lab-based experiments with explicit consent of the participants.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

B. Web Mining

Web mining is the application of data mining techniques to extract knowledge from web data, including web documents, hyperlinks between documents, usage logs of web sites, etc. Web mining is used to discover patterns from the World Wide Web. Web mining can be divided into three different types as – Web usage mining, Web content mining and Web structure mining.

C. Web usage mining

Web Usage Mining is a part of web mining and application of data mining that discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications [6]. It refers to the automatic discovery and analysis of patterns in clickstream and associated data collected or generated as a result of user interactions with Web resources on one or more Web sites. The goal of web usage mining is to capture, model, and analyze the behavioral patterns and profiles of users interacting with the web pages. The discovered patterns are usually represented as collections of pages, objects, or re-sources that are frequently accessed by groups of users with common needs or interests.

Web usage mining itself can be classified further depending on the kind of usage data considered:

- **Web Server Data:** The user logs are collected by the Web server. Typical data includes IP address, page reference and access time.
- **Application Server Data:** Commercial application servers have significant features to enable e-commerce applications to be built on top of them with little effort. A key feature is the ability to track various kinds of business events and log them in application server logs.
- **Application Level Data:** New kinds of events can be defined in an application, and logging can be turned on for them thus generating histories of these specially defined events. It must be noted, however, that many end applications require a combination of one or more of the techniques applied in the categories above.

IV. CONCLUSION

The use of internet growing on increasing for performing any kind of activities and also transaction. This paper performs the review on different methods for the development and maintenance of different kinds of web sites and application important to the users, developers and business persons. There are different kinds of issues associated web usability problems that needs to be overcome in order to satisfy the growth and online business. This paper identifies that web usage mining as the part of data mining is the effective technique helps to identify important web usage data. The paper performs literature study on the different techniques used by different researchers in the field of web usage mining that will provide users navigation and web log information.

REFERENCES

- [1] Ruili Geng and Jeff Tian, "Improving Web Navigation Usability by Comparing Actual and Anticipated Usage", *IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS*, VOL. 45, NO. 1, FEBRUARY 2015.
- [2] R. Cooley, B. Mobasher, and J. Srivastava, "Data preparation for mining World Wide Web browsing patterns", *Knowledge Information System*.
- [3] C. P. Sumathy, R. Padmaja Valli, T. Santhanam, "An overview of pre-processing of web log files for web usage mining".
- [4] Web Server Logs [Online] available: https://en.wikipedia.org/wiki/Server_log
- [5] A. McDonald and R. Welland, "Web engineering in practice," in *Proc. 10th Int. World Wide Web Conf.*, May 2001, pp. 21–30.
- [6] Jaideep Srivastava, Prasanna Desikan, Vipin Kumar, "Web Mining — Concepts, Applications, and Research Directions", [Online] available: http://dmr.cs.umn.edu/Papers/P2004_4.pdf
- [7] May Wang, Benjamin Yen, "Web Structure Reorganization to Improve Web Navigation Efficiency", 11th Pacific-Asia Conference on Information Systems.
- [8] N. Sanfia Sehnaz, Dr. I. Elizabeth Shanthy, "A STUDY OF WEB NAVIGATION USABILITY BASED ON ACTUAL AND ANTICIPATED USER BEHAVIOUR USING DATA MINING TECHNIQUES", *International Journal of Science Technology and Management*, Vol. 5, Issue 1, January 2016.
- [9] Hitesh Hasija, "An Efficient Approach for Web Navigation Using Ant Colony Optimization", IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAI E-20 14), Jaipur, India, May 09-11, 2014.
- [10] T. Arce, P. E. Román, J. D. Velásquez, and V. Parada, "Identifying web sessions with simulated annealing," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1593–1600, 2014.
- [11] Hamilton Fernandes de Moraes Junior, Fabia Lika Nishida and Ana Cristina Vieira de Melo, "Modelling Websites Navigation Elements According to Usability Aspects", *Eighth International Conference on the Quality of Information and Communications Technology*, 978-0-7695-4777-0/12 \$26.00 © 2012 IEEE.



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

- [12] F. E. Ritter, A. R. Freed, and O. L. Haskett, "Discovering user information needs: The case of university department web sites," *ACM Interactions*, vol. 12, no. 5, pp. 19–27, 2005.
- [13] T. Carta, F. Paternò, 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.
- [14] Shalaka S. Thorat, Priyanka More, "User Oriented Approach to Website Navigation Concept Using Mathematical Model", *International Conference on Computational Intelligence and Communication Networks*, 978-1-5090-0076-0/15 \$31.00 © 2015 IEEE.
- [15] Luis Rivero, Marcos Kalinowski, Tayana Conte, "Practical Findings from Applying Innovative Design Usability Evaluation Technologies for Mockups of Web Applications", *47th Hawaii International Conference on System Science*, 978-1-4799-2504-9/14 \$31.00 © 2014 IEEE.