

## Survey Paper on Enhancing Web Navigation Usability Using Web Usage Mining Techniques

S. S. Patil<sup>1</sup> and H.P.Khandagale<sup>2</sup>

<sup>1,2</sup>Computer Science & Technology, Department of Technology, Kolhapur

---

**Abstract**—The Internet has evolved significantly over the past few decades. There are many reasons behind this explosive growth in web traffic. Just about every website has some form of navigation. Unfortunately, navigation of every website is not good. Most of the time a websites navigation is put together by web designers who know a lot about making pretty websites, but very little about marketing a website or creating a website built from the users point of view, which results into navigation-related web usability problems for users. In this paper on Enhancing web navigation usability using web usage mining we are discussing in detail about developing solution model to users navigation related web usability problems and evaluating comparison between actual and anticipated usage behavior which will be useful to users to provide better effectiveness that is higher task completion rate and efficiency that is less time for given tasks.

**Keywords**— Web Server log, Web navigation Usability, Usage pattern Extraction, Web Usage mining, Cognitive User Model.

---

### I. INTRODUCTION

The Internet has evolved significantly over the past few decades. There are many reasons behind this explosive growth in web traffic. These reasons include: the ease of use of the Web; the availability of graphical user interfaces for navigating the Web; the availability of editors, support tools for creating and publishing Web documents; an emerging trend among researchers, educational institutions, commercial organizations to use the Web for disseminating information in a timely fashion; the machine- independent nature of the languages and protocols used for constructing, exchanging Web documents, continuing exponential increase in the number of Internet hosts and users [4].

Web navigation refers to the process of navigating a network of information resources in the World Wide Web, which is organized as hypertext or hypermedia. Just about every website has some form of navigation. Unfortunately, navigation of every website is not good. Most of the time, a websites navigation is put together by Web designers who know a lot about making pretty websites, but very little about marketing a website or creating a website built from the users point of view. Therefore, it is necessary to identify navigation- related Web usability problems which will be helpful to users to provide better effectiveness (higher task completion rate) and efficiency (less time for given tasks). One of the greatest advantages of designing web-based user interfaces over traditional user interfaces is the ability to keep track of user interactions with the site. Thanks to the simple (yet extremely useful) concept of server log files, users' interaction with a website is kept in a raw format that can be easily processed by automated tools. This information is stored on most web servers by default. Statistical testing and reliability analysis be used effectively to assure quality for Web applications. Web usage and failure information extracted from existing Web logs. The usage information is used to build models for statistical Web testing. The related failure information is used to measure the reliability of Web applications and the potential effectiveness of statistical Web testing [14].

Web Usage mining applies data mining technique to extract knowledge from these web log files [3]. Additionally, various tools 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, we can compare actual user navigation trails with the designers expected navigation trails and try to improve the interface of the site accordingly.

**This involves:**

Data Preparation [3] and Pre-processing [5] the log files and converting it to sequential data. The data pre-processing phase includes data cleaning, user identification, session identification, site structure, link details formation, path completion, event generation.

Sequential Pattern mining includes a set of sequences and support threshold to find the complete set of frequent sub sequences.

Finding user navigation patterns from the sequential data using different pattern finding algorithms. The raw log files from the web server on which the Sheet Exchange website resides are first simplified and converted into sequential data. Then a number of pattern finding algorithms are applied. Sequential pattern mining algorithms include: Generalized Sequential Pattern (GSP) Mining Algorithm.

Path completion [7] is a critical and difficult task in the preprocessing phase of web usage mining. For that rules are discovered manually of missing references based on site structure, referrer and other heuristic information.

**Transaction Identification:** Client Side cookies are used for identifying unique users and also to identify and define unique transactions.

**Usage Mining:** Using different techniques Such as statistical analysis, association rules, sequential pattern analysis, clustering and so on applied to the Web domain and to the available data for Pattern Discovery and Extraction [8].

Cognitive user model which will discover anticipated usage behavior. Cognitive user model is used to simulate or predict human behavior or performance based on discovered patterns.

Updating web-navigation paths and comparing original and awaited user behavior based on previous results in terms of log records accessed per session. The ultimate goal of this project is to identify web navigation usability problems, and based on cognitive user model showing quantified usability improvement by the higher task success rate, lower time and effort for given tasks after suggested corrections were implemented over website navigation structure.

## II. RELATED WORK

Ruili Geng and Jeff Tian members of IEEE have proposed a transaction paper [1] on Improving Web Navigation Usability by Comparing Actual and Anticipated Usage which is the main base paper. The paper is all about identifying navigation related Web usability problems faced by user based on comparing actual and anticipated usage patterns. Based on these comparison results (performed on small service-oriented website) they have shown it help us to discover usability issues and suggest corrective actions to improve usability.

B. Liu and his team [2] wrote a book Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. The chapter 6-12 of this book describes all basic concepts related to web specific mining.

R. Cooley, B. Mobasher, and J. Srivastava, proposed a paper [3] on knowledge information system, Data preparation for mining World Wide Web browsing patterns, which explains basic concept of Web Usage Mining. Also this paper presents several data preparation techniques in order to identify unique users and user sessions. Also, a method to divide user sessions into semantically meaningful transactions is defined and successfully tested against two other methods. Transactions identified by the proposed methods are used to discover association rules from real world data using the WEB-MINER system.

Om Kumar C. U. and P. Bhargavi published an article [4] on Analysis of web server log by web usage mining for extracting users patterns , In this paper they described the Web Server Log files and use of Web mining techniques to extract usage patterns by using WEKA(A Software tool).

C.P.Sumathy, R.PadmajaValli, T.Santhanam published an article[5]on An overview of preprocessing of web log files for web usage mining, where they discusses problems with data stored in the log files do not present in an accurate picture of the users accesses to the Web site. Hence, preprocessing of the Web log data is an essential and pre-requisite phase before it can be used for knowledge-discovery or mining tasks.

T. Arce, P. E. Roman, J. D. Velasquez, and V. Parada published a paper [6] Identifying web sessions with simulated annealing on Expert System Application where they discusses how web site redesigning stage is compulsory to take into account the behavior of the users.Here they have proposed heuristic approach based on simulated annealing to solve the sessionization problem.

Nirali Honest and Dr. Atul Patel, Dr. Bankim Patel proposed an IEEE conference paper [7] on A study of path completion techniques in web usage mining. They proposed work on path completion by considering different types of path generated in accessing the website designed using cms and gives a novel algorithm to form the path.

Mr. Akshay Upadhyay, Mr. Balram Purswani published a paper [8] for International Journal of Scientific and Research Publications on Web Usage Mining has Pattern Discovery where they proposed knowledge in respect of pattern discovery of web usage mining, also they described how Users behavior of page browsing should be in hand with the website designers and study about the visitors activities through the web analysis and find patterns of the visitors activities.

Melody Y. Ivory and Marti A. Hearst published an article [9] on The State of the Art in Automating Usability Evaluation of User Interfaces where they proposed extensive survey of usability evaluation methods, organized according to a new taxonomy that emphasizes the role of automation. The survey analyzes existing techniques, identifies which aspects of usability evaluation automation are likely to be of use in future research, and suggests new ways to expand existing approaches to better support usability evaluation.

M. Heinath, J. Dzaack, and A. Wiesner published a paper[10]on Simplifying the development and the analysis of cognitive models in cognitive science conference where they proposed drawbacks of cognitive processes and their underlying structures and given solution to these problems by developing HTAmap and SimTrA to simplify the development and analysis of cognitive models.

Tonio Carta, Fabio Paterno, and Vagner Figueredo de Santana published article [11] on Web Usability Probe: A Tool for Supporting Remote Usability Evaluation of Web Sites on springer. They presented a tool that supports remote usability evaluation of Web sites. The tool considers clientside 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 evaluator to perform a comparison between actual user behavior and an optimal sequence of actions.

T. Tullis and B. Albert [12] wrote a book measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics (Interactive Technologies which covers all aspects of someone's interaction with a product, application, or system. Also it says that user experience is measurable or quantifiable. Also it describes usability issues, various user experience metrics.

### **III. CHOICE OF THE TOPIC WITH REASONING**

WWW is a system of interlinked hypertext documents accessed via the Internet. Around 11 Hundred million people access internet daily. Therefore information available on WWW is also growing. With this continued growth of information and proliferation of web services and web based information systems, web sites are also growing to host them.

Every website has some form of navigation. Unfortunately, navigation of every website is not good. Most of the time, a websites navigation is put together by Web designers who know a lot about making pretty websites, but very little about marketing a website or creating a website built from the users point of view.

To overcome the web usability problems then Building and ensuring easy-to-use Web systems is becoming a core competency for business survival. Usability is defined as the effectiveness, efficiency, and satisfaction with which specific users can complete specific tasks in a particular environment. Three basic Web design principles, i.e., structural firmness, functional convenience, and presentation delight, were identified to help improve users' online experience. Structural firmness relates primarily to the characteristics that influence the website security and performance. Functional convenience refers to the availability of convenient characteristics, such as a sites ease of use and ease of navigation, that help users' interaction with the interface. Presentation delight refers to the website characteristics that stimulate users' senses.

#### IV. DATA PREPARATION AND PRE-PROCESSING

The web server logs are synthetically generated over specific website. And these access logs are used for pre-processing to remove unwanted noisy data. This pre-processing phase consists of following stages:

##### Stage 1: Data Cleaning

Data cleaning is usually site-specific, and involves tasks such as, removing irrelevant entries such as those that represent multimedia data and scripts or uninteresting entries such as those that belongs to top/bottom frames.

##### Stage 2: User Identification

Since several users may share a single machine name, certain heuristics are used to identify users. We use the phrase user activity record to refer to the sequence of logged activities belonging to the same user.

##### Stage 3: Session Identification

Sessionization is the process of segmenting the user activity record of each user into sessions, each representing a single visit to the site. The goal of a sessionization heuristic is to reconstruct, from the click stream data, the actual sequence of actions performed by one user during one visit to the site.

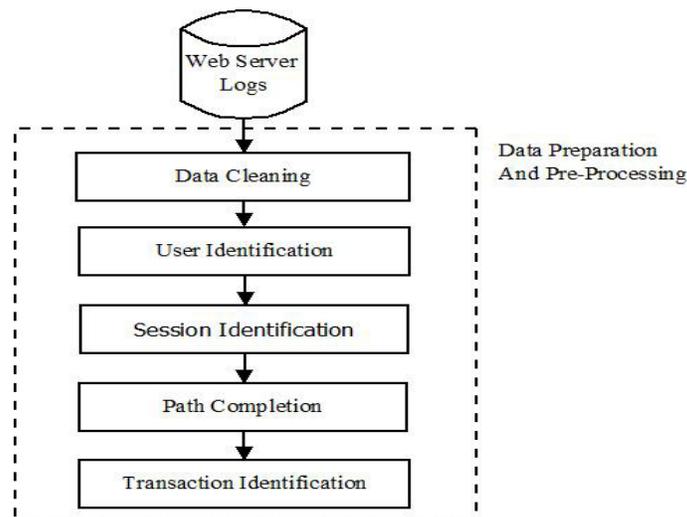


Fig. Data Preparation and pre-processing

#### Stage 4: Path Completion

Here, rules for missing references are discovered manually based on site structure, referrer, and other heuristic information.

#### Stage 5: Transaction Identification

Most probably using client side cookies are used for it.

#### Stage 6: Usage Mining Technique for Pattern Discovery and Extraction

Using different techniques Such as statistical analysis, association rules, sequential pattern analysis, clustering and so on applied to the Web domain and to the available data for Pattern Discovery and Extraction.

### V. COGNITIVE USER MODEL

Cognitive User Model which will specify the anticipated Usage behavior based on the patterns discovered in previous pre-processing phase.

Cognitive modeling methods to understand complex interactive behavior involved in three tasks: (1) icon search, (2) graph reading, and (3) information retrieval on the World Wide Web (WWW).

Here, traditional Cognitive user model(ACT-R(Adaptive Control of Thought Rational)/EPIC(Executive Process/Interactive Control)) is used for building expert system which will predict anticipated usage behavior or performance based on discovered patterns.

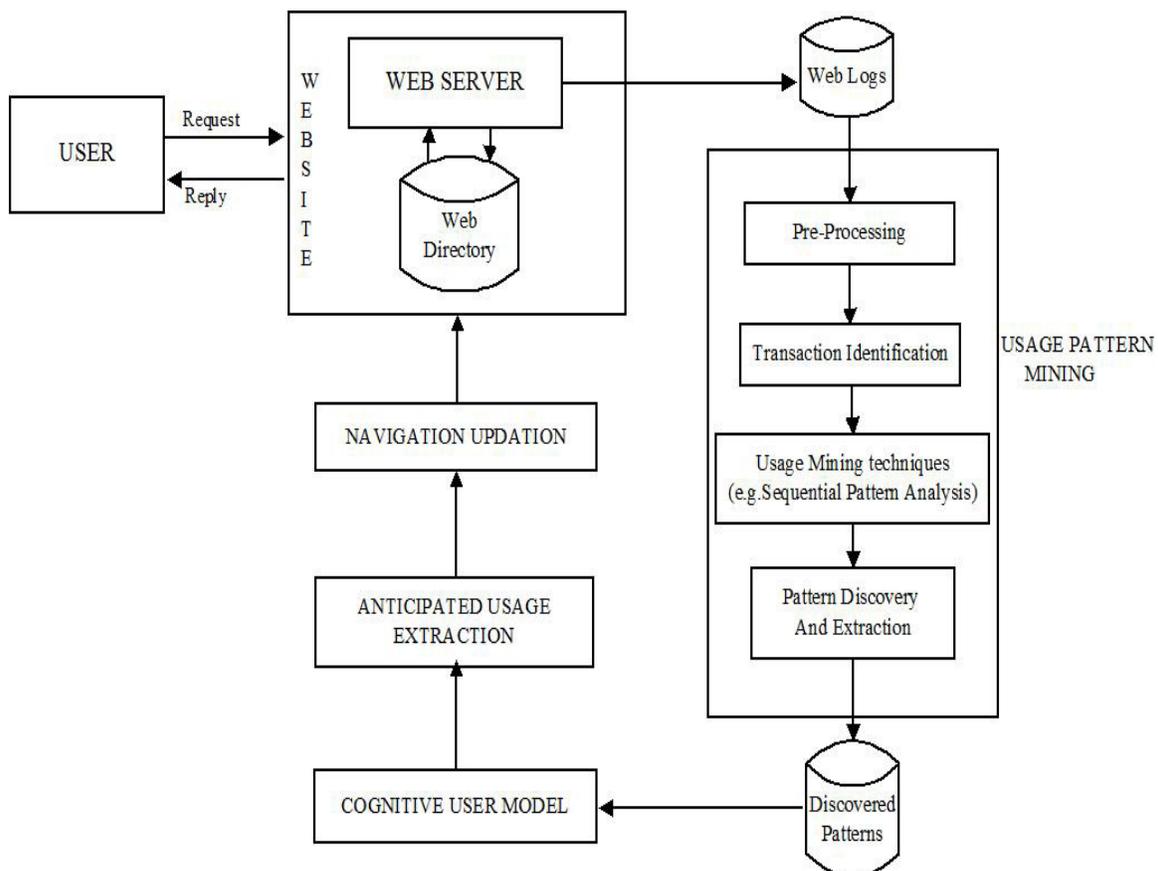


Fig. System Architecture

## VI. WEBSITE LINKS UPDATION IN AUTOMATED MANNER

The final anticipated behavior computed from system may be provided to the developer team for necessary modifications to achieve web-navigation improvement.

But, this module perform necessary modification by updating links on web pages automatically with the help of Document Object Model (DOM) technique, which will saves effort and time of developers team.

## VII. CONCLUSION

This method can contribute significantly to continuous usability improvement over these prolonged maintenance cycles performed over web navigation structure. The usability improvement in successive iterations can be quantified by the progressively better effectiveness (higher task completion rate) and efficiency (less time for given tasks).

## VIII. ACKNOWLEDGMENT

I express my sincere thanks to guide H. P. Khandagale for their valuable guidance. I also express my sincere thanks to the department of technology, Shivaji University, Kolhapur for providing the necessary resources to carry out this work.

## REFERENCES

- [1] Ruili Geng , 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] B. Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. New York, NY, USA: Springer, 2007.
- [3] R.Cooley, B.Mobasher, and J.Srivastava, Data preparation for mining world wide web browsing patterns , Knowl. Inf. Syst., vol. 1, no. 1,pp.532, 1999.
- [4] Om Kumar C. U. and P. Bhargavi, Analysis of web server log by web usage mining for extracting users patterns, Vol. 3, Issue 2, ISSN 2249-6831, 123-136 , (IJCSSEITR), June 2013.
- [5] C.P.Sumathy, R. Padmaja Valli, T. Santhanam, An overview of pre processing of web log files for web usage mining, JATIT, vol. 34No. 1, ISSN: 1992-8645, 15th Dec.2011.
- [6] T. Arce, P. E. Romn, J. D. Velsquez, and V. Parada, Identifying web sessions with simulated annealing, Expert System Application., vol. 41, no. 4, pp. 15931600, 2014.
- [7] Nirali Honest and Dr. Atul Patel, Dr. Bankim Patel, A study of path completion techniques in web usage mining, IEEE International Conference on Computational Intelligence and Communication Technology,2015.
- [8] Mr. Akshay Upadhyay, Mr. Balram Purswani, Web usage mining has pattern discovery, International Journal of Scientific and Research Publications, Volume 3, Issue 2, February 2013.
- [9] M. Y. Ivory and M. A. Hearst, The state of the art in automating usability evaluation of user interfaces, ACM Computation Surveys, vol. 33, no. 4,pp. 470516, 2001.
- [10] M. Heinath, J. Dzaack, and A. Wiesner, Simplifying the development and the analysis of cognitive models, in Proc. Eur. Cognitive Sci. Conf.,Delphi, Greece, 2007, pp. 446451.
- [11] T. Carta, F. Patern, and V. F. D. Santana, Web usability probe: A tool for sup- porting remote usability evaluation of websites, in Human-Computer Interaction - INTERACT 2011. New York, NY, USA: Springer, pp. 349-357, 2011.
- [12] T. Tullis and B. Albert, Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics (Interactive Technologies). San Mateo, CA, USA: Morgan Kaufmann, 2008.
- [13] M. F. Arlitt and C. L. Williamson, Internet Web servers: Workload characterization and performance implications, IEEE/ACM Trans. Netw., vol. 5, no. 5, pp. 631645, Oct. 1997.
- [14] C. Kallepalli and J. Tian, Measuring and modeling usage and reliability for statistical Web testing, IEEE Trans. Softw. Engin., vol. 27, no. 11, pp. 10231036, Nov. 2001.