ISSN: 2321-7782 (Online)
Volume 3, Issue 10, October 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

A Review on Analyzing Actual and Anticipated Usage to Promote Web Navigation Usability

RafatRana S.H. Rizvi[1]
Computer Science and Engineering
H.V.P.M's C.O.E.T
Amravati, India

Ranjit R. Keole[2]
Professor
Information Technology
H.V.P.M's C.O.E.T Amravati, India

Abstract: ***Usability is defined as the effectiveness, efficiency, and satisfaction with which specific users can complete specific tasks in a particular environment. This paper presents a new method to identify navigation related Web usability problems based on comparing actual and anticipated usage patterns. The actual usage patterns can be extracted from client-side logs routinely recorded for operational websites by first processing the log data to identify users, user sessions, and user task-oriented transactions, and then applying an usage mining algorithm to discover patterns among actual usage paths. The anticipated usage, including information about both the path and time required for user-oriented tasks, is captured by our ideal user interactive path models based on their cognition of user behavior. The comparison is performed for checking results and identifying user navigation difficulties.***

Keywords: Usability, Server Logs, IUIP, Web Navigation, Web analytics.

I. INTRODUCTION

As the World Wide Web becomes prevalent today, 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. Usability is a quality attribute that assesses how easy user interfaces are to use.The word "usability" also refers to methods for improving ease-of-use during the design processs. Three basic Web design principles, i.e.,structural firmness, functional convenience, and presentational 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 site's ease of use and ease of navigation, that help users' interaction with the interface. Presentational delight refers to the website characteristics that stimulate users' senses.

Some people distinguish between the terms usability and user experience.Usability is usually considered the ability of the user to use the thing to carry out a task successfully, whereas user experience takes a broader view, looking at the individual's entire interaction with the thing, as well as the thoughts, feelings, and perceptions that result from that interaction.

Usability Engineering is a discipline that provides structured methods for achieving usability in user interface design during product development. User Interface Design describes any kind of interface design of an interactive system and also the information design of non-interactive systems (e.g. subway map). Thus User Interface Design is a subset of Usability Engineering. 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, significant challenges exist, including 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.

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 .

This paper extract actual user behavior from Web server logs, capture anticipated user behavior with the help of cognitive user models, and perform a comparison between the two. This deviation analysis would help us identify some navigation related usability problems. Correcting these problems would lead to better functional convenience as characterized by both better effectiveness (higher task completion rate) and efficiency (less time for given tasks). This new method would complement traditional usability practices and overcome some of the existing challenges.

II. Literature Review And Related Work

In 1997,M.F.Arlitt and C.L.Williamson had analyzed server-side logs by which Web workload was characterized and used to suggest performance enhancements for Internet Web servers. But, they do not provide all of the information that is of interest.For example, the log entries tell only fhe number of bytes transferred for a document, not its actual size; there is no record of the elapsed time required for a document transfer.

In1997, L. Tauscher and S. Greenberg, proposed that users' revisitiation patterns can be discovered by mining server logs to develop guidelines for browser history mechanism. Through history, a user can return quickly to a previously visited page, possibly reducing the cognitive and physical overhead required to navigate to it from scratch.But in this paper how often user revisit their pages has not been analysed.

In 1999,Jeff Tian ,suggested that 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.The drawback of this paper is that it explore only the limitations of log file data for usability analysis but do not provide solutions to overcome limitation.

In 2005, 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. 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.

In 2011, T. Carta, F. Patern`o, and V. F. D. Santana , presents a tool that supports remote usability evaluation of Web sites. This work presents Web Usability Probe (WUP), a tool that follows a proxy based architecture, performs remote evaluation, and considers client-side logs as data source.Many problems was analysed like user is not getting accurate information,there was lack of user guidance,etc.But these problems were not solved.

In 2014, T. Arce, P. E. Rom´an, J. D. Vel ´asquez, and V. Parada , presents a heuristic approach based on simulated annealing for the sessionization problem. 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.

III. Proposed Methodology

The proposed method identify navigation related Web usability problems based on comparing actual and anticipated usage patterns. First, method extract actual navigation paths from server logs and discover patterns for some typical events. In parallel, it construct IUIP models for the same events. IUIP models are based on the cognition of user behavior and can represent anticipated paths for specific user-oriented tasks. The result checking employs the mechanism of test oracle. An oracle is generally used to determine whether a test has passed or failed . Here, IUIP models is used as the oracle to identify the usability

issues related to the users' actual navigation paths by analyzing the deviations between the two. This deviation analysis would help us identify some navigation related usability problems. Correcting these problems would lead to better functional convenience as characterized by both better effectiveness (higher task completion rate) and efficiency (less time for given tasks). This new method would complement traditional usability practices and overcome some of the existing challenges.Proposed method also anticipate user to access web page currently used by some another user of same category.

Fig. 1 shows the architecture of our method. It includes three major modules: Usage Pattern Extraction, IUIP Modeling, and Usability Problem Identification.

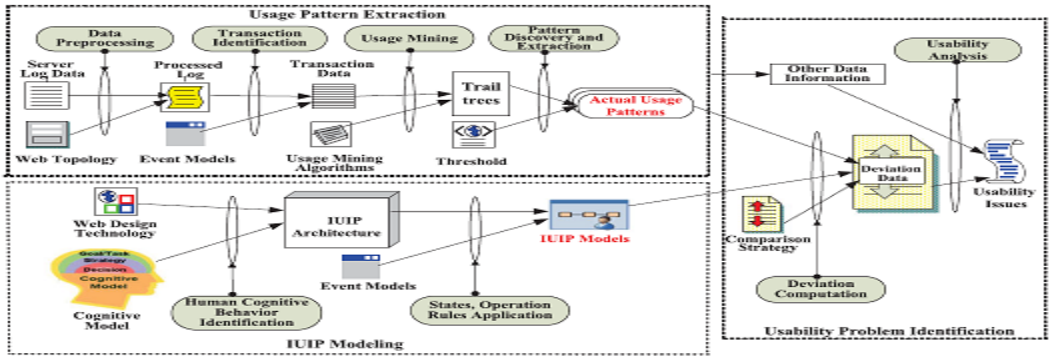


Figure 1. Architecture of a new method for identifying usability problems

3.1.USAGE PATTERN EXTRACTION

Web server logs are our data source. Each entry in a log contains the IP address of the originating host, the timestamp, the requested Web page, the referrer, the user agent and other data. Typically, the raw data need to be preprocessed and converted into user sessions and transactions to extract usage patterns.

A. Data Preparation and Preprocessing :-The data preparation and preprocessing include the following domain-dependent tasks.

1) *Data cleaning:* This task is usually site-specific and involves removing extraneous references to style files, graphics, or sound files that may not be important for the purpose of our analysis.

2) *User identification:- Next Unique users must be identified .For the* log/site based methods ,there are heuristics that can be used to help to identify unique users.

3)*Session identification*:-The goal of session identification is to divide the page access of each user into individual sessions, with each representing a single visit to a site.

4)*Path completion*:-Another problem in identifying unique user session is determining if there are important accesses that are not recorded in the access log .This problem is refer to as Path completion.Methods similar to those used for user identification can be used for path completion.

B. Transaction Identification:- E-commerce data typically include various task-oriented events. Such task-oriented events are order, shipping, and shopping cart changes. In most cases, there is a need to divide individual data into corresponding groups called Web transactions. A transaction usually has a well-defined beginning and end associated with a specific task. For example, when a user places something in his shopping cart is beginning of transaction and when he has completed the purchase on the confirmation screen is end of transaction.

C. Trail Tree Construction:-The transactions identified from each user session form a collection of paths. As multiple visitors may access the same pages in the same order, we use data structure called as Trie to merge the paths along common prefixes. A trie, or a prefix tree, is an ordered tree used to store an associative array where the keys are usually strings. All the

descendants of a node have a common prefix of the string associated with that node. The root is associated with the empty string. To construct a tree structure ,we adapted the trie algorithm that also captures user visit frequencies, which is called a *trail tree* . In a trail tree, a complete path from the root to a leaf node is called a *trail.*

3.2. IDEAL USER INTERACTIVE PATH MODEL CONSTRUCTION

IUIP models are based on the cognitive models , particularly the ACT-R model. ACT-R/PM is an efort to augment a system that has been remarkably successful at describing and predicting human behavior in primarily cognitive domains with a perceptual-motor system. Due to the complexity of ACT-R model development and the low-level rulebased programming language it relies on , we constructed our own cognitive architecture and supporting tool based on the ideas from ACT-R. In general, the user behavior patterns can be traced with a sequence of states and transitions . The sequence of states and transitions is called task solving process.Each task solving process include 3 types of information:-state and transition itself,repetition of state and transition,sequential order of state and transition. For a particular goal, a sequence of related operation rules can be specified for a series of transitions. IUIP model specifies both the path and the benchmark interactive time (no more than a maximum time) for some specific states (pages). The benchmark time can first be specified based on general rules for common types of Web pages. For example, human factors guidelines specify the upper bound for the response time to mitigate the risk that users will lose interest in a website. Humans usually try to complete their tasks in the most efficient manner by attempting to maximize their returns while minimizing the cost.

3.3. USABILITY PROBLEM IDENTIFICATION

The actual users' navigation trails extracted from the aggregated trail tree are compared against corresponding IUIP models automatically. This comparison will yield a set of deviations between the two.We can identify some common problems of actual users' interaction with the Web application by focusing on deviations that occur frequently. Combined with expertise in product internal and contextual information, our results can also help identify the root causes of some usability problems existing in the Web design.

IV. WEB SERVER LOGFILES, LIMITATIONS OF LOGFILES , WEB ANALYTICS

4.1.WEB SERVER LOGFILES

Web server logfiles are a recording of user interactions on a website . They are automatically generated by the web server software hosting the website . Whenever a user clicks on a clickable item on the website (such as a link), a request is sent to the web server, and new entries are recorded into the logfile . A typical entry in combined log format resembles the following: 90.07.06.86 – hciguy [15/Nov/2012:16:00:00 -0700] "GET /index.html?search=routers HTTP.1.0" 200 2326 "http://www.google.com" "Mozilla/5.0 (Macintosh; U; PPC Mac OS X; en) AppleWebKit/412.7 (KHTML, like Gecko) Safari/412.5"

A single log entry like the one above details a user's IP address (allowing them to be uniquely identified, and identifies their geographic location); the date and time of their interaction (Nov 15th at 4pm), describes their interaction (index.html?search=routers is called a request and describes a search performed on the homepage for routers); indicates the prior website that referred them (www.google.com); and provides information on their web browser and computer (Macintosh running Mozilla version 5.0). Although a single log entry details only a user's single click, a series of log entries describes a user's visit to the website. For example, the series of requests can be used to follow the user's path through the website and identify interactions with content. The time difference between timestamps provides the duration between interactions.

4.2.LIMITATIONS OF LOGFILES

Logfiles, however, are incapable of capturing all interaction information, such as a user's thoughts, intentions, gaze, voice, a visual of the user, time spent interacting with another website or program, time spent away from the computer, or network

latency. In addition, logfiles are incapable of capturing cursor movements, clicks on non-clickable items, cursor-selected text, page scrolling, interactions with the Internet browser, and whether a link was opened in the same window or a new one. Logfiles also do not capture interactions with dynamic content such as AJAX or Flash, which do not send requests to the webserver and thus do not prompt logfile entries to be generated. Also, if a website has frequently changing content or structure, then the older log data will be invalid when cross-referenced with the current website; although this can easily be remedied by keeping backups of each website version .

4.3.WEB ANALYTICS

Web analytics is "the measurement, collection, analysis and reporting of Internet data for the purpose of understanding and optimizing Web usage" and makes use of logfile or JavaScript tracking data. The result of web analytics is typically reported through visualizations of website usage rather than raw numerical data. In other words, visualization is the primary way that raw logfile data is transformed into visual information. The likely reason for this is that visualizations present data "in forms allowing humans to exploit their innate visual analysis capabilities to interpret results" . Techniques for analysing and visualizing/reporting web analytics data come in four flavours: frequency analysis, time analysis, exit analysis, and pattern analysis.

- » Frequency analysis looks at the frequency of user interactions and can reveal information about users' preferences and behaviour. For example, the frequency of a webpage's accesses (called page views) can 50 demonstrate its value to users . Frequency is calculated by simply tallying the number of times an interaction occurs.
- » Time analysis looks at the amount of time spent on the website in a single visit and can indicate proficiency, the degree to which a particular webpage forces user cognition , or the interestingness of the content . For example, if a user spends a large amount of time on a page that requires them to complete an action, this could mean that they are not proficient; the page requires a high degree of cognition ; or that the page is poorly designed . Alternatively, this could indicate a large network latency, or that the user ceased interacting with the website and started a new task elsewhere. The amount of time a user spends on a content-abundant page can be used to determine its *interestingness*, defined as the affinity a user has for a particular web page or feature .
- » Regarding exit analysis, Kaushik states that where a user exits a website can indicate their degree of success. For example, a user exiting on a content page relevant to their search criteria suggests that they found what they were looking for, whereas a user exiting on a help page suggests that the user did not find what they were looking for.
- » Pattern analysis aims to discover patterns within analytics data. Some of these patterns provide insights on past usage; others are predictive in nature . *Association rules*, for example, determine the probability that Content B will be accessed with Content A (not time-ordered) and can be used to understand and predict usage patterns

V. CONCLUSION

This paper 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 method can identify areas with usability issues to help improve the usability of Web systems. Server logs in our method represent real users' operations in natural working conditions, and our IUIP models injected with human behavior cognition represent part of cognitive experts' work.

Some of research topics are as following:

- » Presenting some strategies to further expand our usability research to cover more usability aspects to improve Web users' overall satisfaction.;
- » Introducing some methods to explore additional approaches to discover Web usage patterns and related usability problems generalizable to other interesting domains;

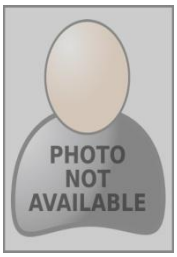
ACKNOWLEDGMENT

My thanks to the Guide, Prof. R.R.Keole and Principal Dr.A.B.Marathe, who provided me constructive and positive feedback during the preparation of this paper.

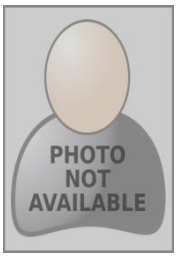
References

1. 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.
2. L. Tauscher and S. Greenberg, "Revisitation patterns in World Wide Web navigation," in Proc. ACM SIGCHI Conf. Human Factors Comput. Syst., New York, NY, USA, 1997, pp. 399–406.
3. Tec-Ed, "Assessing web site usability from server log files," White Paper, Tec-Ed, 1999.
4. 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.
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. T. Arce, P. E. Rom´an, J. D. Vel´asquez, and V. Parada, "Identifying web sessions with simulated annealing," Expert Syst. Appl., vol. 41, no. 4,pp. 1593–1600, 2014.
7. 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.
8. 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.
9. J. Nielsen, Usability Engineering. San Mateo, CA, USA: Morgan Kaufmann,1993.
10. T. Tullis and B. Albert, Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics (Interactive Technologies). San Mateo, CA, USA: Morgan Kaufmann, 2008.

AUTHOR(S) PROFILE



RafatRana S.H. Rizvi, received the B.E.degree in Computer Science and Engineering from H.V.P.M's College Of Engineering And Technology, Amravati in 2014. She is currently persuing Master's Degree in Computer Science and Engineering from H.V.P.M's College of Engineering And Technology, Amravati.



Prof. Ranjit R.Keole, received the B.E.and M.E degree in Computer Science from Prof. Ram Megha Institute of Technology, Badnera in 1992 and 2008, respectively. His field of specialisation is web Mining. He is currently working as Associate Professor at H.V.P.M's college of Engineering and Technology, Amravati.

