

# A Review on Enhancing Web Navigation Usability by Analyzing and Comparing Actual and Anticipated Usage

Payal D. Chintawar<sup>1</sup>, Prof. K. P.Wagh<sup>2</sup> and Dr. P.N.Chatur<sup>3</sup>

<sup>1</sup> PG Scholar, <sup>2</sup> Assistant Professor, <sup>3</sup> Associate Professor

<sup>1,3</sup> Department of Computer Science and Engineering,

<sup>2</sup> Department of Information Technology,

<sup>1,2,3</sup> Government College of Engineering, Amravati. INDIA.

## ABSTRACT

*The Internet has evolved significantly over the past few decades. Many users are switching to the internet. For survival of business organizations, it has become necessary to increase their web based activities regarding business work. Usability of web system is the important concept that deals with effectiveness, efficiency and satisfaction with which specific tasks are completed in a particular environment by users. Actual usage patterns are extracted from web server logs. Web usability is improved by first preprocessing and analyzing the web data logs located at the web server. Pattern mining and extraction methods are applied on the preprocessed data to get the meaningful data out of the log file. Using this data user's actual navigation behavior is found out. IUIP model is a cognitive user model constructed for discovering anticipated usage behavior based on the cognition of user behavior. By comparing extracted patterns and cognitive user model, system finds the usability issues in the web system.*

**Keywords—** *Cognitive User Model, Usage Pattern, Usability, IUIP, Web Server Logs.*

## I INTRODUCTION

Internet has become prevalent in everyday life. As digitalization has become part of all the activities today, for business survival it is necessary to build web systems that are easy-to-use. An important input for efficient web system design is analysis of how a web system is being used by users. Thus it is important for a business organization to analyze users using their web systems and interests of these users. This is called as usage analysis. Web usage mining can be done by applying mining techniques on huge web data. Some of these techniques are association rule generation, sequential pattern generation and clustering [2]. Using usage information the needs of users of the web site can be better understood. Understanding users' needs can be used to restructure the website in order to increase user satisfaction.

Log data is routinely captured at web servers. This data is called as actual usage data. Some of the applications where log data is used are usage based testing, guiding user interface design and understanding user behavior. With this data analysis it is possible to understand user behavior. This is also used to guide business intelligence

recommendation which improves the growth rate of the organization. To understand the human factor of the user interface design cognitive science is used. Basically cognitive science works on the mechanisms that work on the strengths and limitations of the human perception. Usability engineering provides methods for addressing usability issues and for measuring usability. Usability testing is a method used to understand the usability of a website. This involves activities like observing users who are interacting with the website.

Navigation related web usability problems in the site are identified by understanding most likely accessed patterns of users of website. Web navigation usability is improved by identifying and solving the web navigation usability problems. Thus web developers can customize and adapt the site's interface for the individual user to improve the site's structure within the underlying hypertext system.

## **II RELATED WORK**

Ruili Geng and Jeff Tian[1] proposed a method for identifying web usability problems by extracting usage patterns using cognitive user model and comparing it with actual usage pattern. Actual user behavior is captured from web server logs. Anticipated user behavior is captured with the help of cognitive user model. IUIP models is proposed using human behavior cognition which represents part of cognitive expert's work. The method corrects usability problems leading to better functional convenience as characterized by better effectiveness and efficiency. This method is used for continuous usability improvement of the web system. This heuristic approach reduces the processing times up to 166 times.

R. Cooley, B. Mobasher, and J. Srivastava [2] stated the method for preprocessing of server logs and transaction identification. When data mining techniques are applied and used for large web data stored at repositories then this field of work is called as web usage mining. Sequential pattern generation, association rule generation and clustering are the commonly used for web usage mining. Browser behavior models are derived to predict user behavior from the server logs and the site files. This work states the limitations of maximum forward reference approach and auxiliary content transactions. A method to create semantically meaningful transactions from user sessions is tested successfully against other two methods. WEB-MINER system is developed for discovering association rules from the real world data.

J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin [3] proposed the work on ACT-R model. ACT-R provides a process model of human performance in interactive tasks that works well with complex interfaces. Using this model, cognitive factors are specified like domain knowledge and problem solving strategies. This is done by developing cognitive models having interactive behavior. It consists of some of the modules that takes information as a input from the environment, processes this information and executes actions for achieving specific goals. After this process pattern matching process is done. In this process productions are found out and matched with current content. This model is used to better understand the decisions made by web users by following various links to satisfy the user goals. ACT-R model has its own limitations because of the complexity of deployment and the low level programming languages used.

Guosheng Kang, Mingdong Tang, Jianxun Liu Xiaoqing and Buqing Cao [4] proposes a novel diversity aware service ranking algorithm for diversified top k services. QoS is the advanced feature that prioritizes internet traffic to neutralize the effect of busy bandwidth. The QoS of web users are getting may be different. There is need to check user service usage history. This is done by processing Query log and profile of user called historical approach. This can also be done by collaborative approach that identifies potential user interest. Hybrid method finds similar users and then finds their interests. User's satisfaction decreases due to similar recommendations. The functional relevance of web services is computed. Better web services are provided for sharing data, computing resources and programs on the internet.

F. E. Ritter, A. R. Freed, and O. L. Haskett [5] proposed a method for effective task analysis of the website. This method helps to improve design of website and user satisfaction by analyzing several university websites, their departmental hardcopies, search engine queries and interviews of some of the website users. The tasks that website supports are discovered by design team. Most viewed pages are identified and then after further analysis some of the tasks are identifies that are to be made easier. For analysis geographically dispersed users are taken into consideration. The websites selected for analysis are compared based on task completion rate. At last some of the implications for website design are given. This includes getting website listed on search engine and direct banner adds.

C. M. Nadeem Faisal, Martin Gonzalez-Rodriguez, Daniel Fernandez Lanvin, and Javier de Andres Suarez [6] proposed a model and a hypothesis to avoid annoyance towards websites. For web design attribute selection, user preferences are evaluated. Content quality and navigation deals with user trust whereas interactivity, color selection and typography deals with user satisfaction. A questionnaire is designed for university students. Partial least square method is used for analyzing the data collected from university students. It is found that trust and loyalty factors works better than satisfaction and loyalty factors for web design.

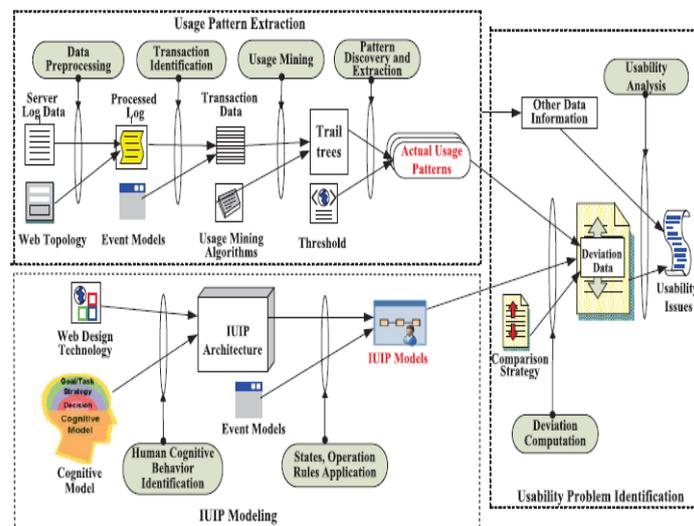
Sneha V. Dehankar, K.P.Wagh and P. N. Chatur [7] proposed a system for classifying various links into predefined classes based on highest probability of word of every document. Only several and relevant links are being classified using Apriori algorithm. This classification helps to reduce the time complexity for scanning all the links. In this method frequent item sets are found by applying association rule. The web page classification system gives the efficient and accurate classifier.

Hetal C. Chaudhari, K. P. Wagh and P. N. Chatur [8] proposed TF-IDF based Apriori scheme. Clustering is applied on input that is results of the query fired. On this result modified TF-IDF Apriori method is applied followed by ranking of the documents. Formulation of equation for computing threshold is done to discard the rows and columns of IF-IDF table created. Cosine similarity is computed between every pair of documents. This method is used along with TF-IDF Apriori method to further provide ranked results for the fired query. The F-measure value obtained using this method is 81%.

Sanjay D. Sawaitul, Prof. K. P. Wagh, Dr. P. N. Chatur [9] proposed a method for storing and analyzing weather parameters. This method is used further for predicting future weather. A wireless kit for weather forecast is introduced. Based on the changes in one weather parameter, changes in other weather parameters can be predicted. Old weather forecast models uses methods like soft computing and image retrieval and algorithms like k nearest neighbor. Back propagation is more reliable, accurate and consistent method. Neural network is a signal processing approach for modeling weather forecast systems. This method is alternative to the traditional weather forecast methods.

P. H. Govardhan, K. P. Wagh, P. N. Chatur [10] proposed a method for preprocessing results of web queries that is web pages followed by measuring similarity between the documents which is motivated by the potential effectiveness gains postulated by the cluster hypothesis. Web documents are the rich source of information. These web resources are processed considering all possibilities that is present and absent features thus simplifying the tedious task of similarity finding.

### III ARCHITECTURE OF METHOD



**Fig. 1: Architecture of the new method for identifying usability problems of web system.** Architecture of the method is shown in Fig. 1. It includes three major modules:

- Usage Pattern Extraction
- IUIP Modelling
- Usability Problem Identification.

First of all actual navigation paths are extracted from web server logs. Preprocessing is done on these web server logs. Then the patterns discovery for some identified events is done. Construction of IUIP models for these events is done. IUIP models are based on the cognition of user behavior and can represent anticipated paths for user oriented tasks. Test oracle is a tool used for result checking.. IUIP models are used as the oracle to identify the usability issues by finding deviations between extracted actual pattern and cognitive model.

### **3.1 Usage Pattern Extraction**

Proposed system improves web usability based on web data logs located at the web server. Log files are files that list the actions of users on web system. Web servers are the computers that deliver the web pages. Each entry in a log file contains the IP address of the timestamp, the originating host, the requested web page, the user agent, the referrer, request type and other relevant data.

#### **3.1.1 Preprocessing**

The raw data in web log file is preprocessed to extract usage patterns. This includes following steps :

- Data Cleaning - This step removes unnecessary and redundant log entries in the log file.
- User Identification - This step differentiates log entries and thus the different users.
- Session Identification - In this step the activities done by users from the moment they enters the website to the moment they exits it is recorded.
- Path completion - This step identifies the hyperlinks between the previous page and next page

#### **3.1.2 Transaction Identification**

After preprocessing transaction identification is done. Sequences of page references are grouped into logical units. These logical units represent web transactions. After this mining and pattern extraction techniques are applied on web usage data. Transaction is a group of page references made by a user during single visit to the web system. A set of tasks is identified which represents the event model or task model. Using task model, transaction identification is done.

#### **3.1.3 Trail Tree Construction**

Transaction is nothing but the collection of paths. Users visiting paths and its frequency are stored by construction a trail tree. Path is also called as trail. Mining and pattern discovery is done on this trail tree. For trail tree construction tri algorithm is used. This stores the unique paths and the frequency of users using these paths. Single node in a trail tree represents a single page. Annotation given to every node gives the number of users reached to the page following that particular path in which it is represented.

### **3.2 Construction of Ideal User Interactive Path Model**

IUIP model is the abbreviation of Ideal User Interactive Path model. This model is based on the ACT-R model. Using the sequence of tasks or transaction the pattern of user behaviour is traced using this model. This model consists of states and transitions. States represents web pages whereas transition is the hyperlink between one page to another. IUIP model specifies benchmark interactive time and the path. Benchmark time is a maximum time after which it is considered that user will lose the interest in the page. Expert and novice both should construct the IUIP model separately. Best part about IUIP model is that it can be reused. It is constructed using the programming languages C++, visual diagram software DIA, XML, etc. In this model, we can create our own symbols using open source visual diagram software DIA and XML.

### **3.3 Usability Problem Identification**

Comparison is done between actual usage pattern of users and IUIP model. Thus deviations are found out. This identifies 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, these results help identify usability problems and root cause of these problems existing in the web system design. The calculation of deviations between actual users' usage patterns and IUIP can be divided into two parts based on time spent and logical choices made by users at each page:

- Logical deviation is calculated as follows:  
When the path choice anticipated by the IUIP model is available but not selected, a single deviation is counted. Add all the deviations found for each page over all the selected user transactions.
- Temporal deviation is calculated as follows:  
When a user spends more time at a specific page than the benchmark specified for the corresponding state in the IUIP model, a single deviation is counted. Add all the deviations found transactions for each page over all the selected user.

## **IV PERFORMANCE ANALYSIS**

By checking the extracted usage patterns against the four IUIP models, temporal and logical deviations were obtained. Some of the logical and temporal deviation methods the issues are found out. 17 usability issues are found. After finding the issues restructuring of the website is done. Again web logs are collected and compared the usability of improved web system to the old one in terms of task success rate, average efforts required and the average time required for the task completion.

## **V CONCLUSIONS**

Internet has become prevalent in everyday life. As This approach identifies the interest of users and helps to make proper changes to the business strategies. This leads to increase the users trust and profit margin in business. This improves the GUI and usability of the web system. This method identifies different users, their session time, etc. to provide web recommendation for building effective business analytic system. In the future, large scale web applications can be studied for the validation study. Additional approaches can be found to discover user web usage patterns. The additional approaches related to usability problems generalizable to other interesting domains can be found out. Recommendation analysis can be done on new factors.

## **REFERENCES**

- [1] RuiliGeng, and Jeff Tian, Member, IEEE, "Improving Web Navigation Usability by Comparing Actual and Anticipated Usage", IEEE Transactions on Human–Machine Systems, Volume No. 45, 2015.
- [2] 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.

- [3] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, "An integrated theory of the mind," *Psychol. Rev.*, vol. 111, pp. 1036–1060, 2004.
- [4] Guosheng Kang, Mingdong Tang, Jianxun Liu Xiaoqing and Buqing Cao, "Diversifying Web Service Recommendation Results via Exploring Service Usage Histor" *IEEE Transactions on Services Computing* Vol. 9, Issue 4, July-Aug. 2016.
- [5] 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.
- [6] C. M. Nadeem Faisal, Martin Gonzalez-Rodriguez, Daniel Fernandez-Lanvin, and Javier de Andres-Suarez, "Web Design Attributes in Building User Trust, Satisfaction, and Loyalty for a High Uncertainty Avoidance Culture", *IEEE Transactions on Human–Machine Systems*, Volume No. 47, Issue 6, Dec 2017.
- [7] Sneha V. Dehankar, K. P. Wagh and P. N. Chatur, "Web Page Classification Using Apriori Algorithm and Naive Bayes Classifier" *IJARCSMS* volume 3, issue 4, April 2015, pg-527- 533.
- [8] Hetal C. Chaudhari, K. P. Wagh and P. N. Chatur, "Search Engine Results Clustering Using TF-IDF Based Apriori Approach" *IJECS* vol. 4, Issue 5, May 2015, Pg-11956- 1196.
- [9] Sanjay D. Sawaitul, Prof. K. P. Wagh, Dr. P. N. Chatur, "Classification and Prediction of Future Weather by using Back Propagation Algorithm-An Approach", *IJETAE*, ISSN 2250-2459, Volume 2, Issue 1, January 2012.
- [10] P. H. Govardhan, K. P. Wagh, P. N. Chatur, "Web Document Clustering using Proposed Similarity Measure", *International Journal of Computer Applications (0975 – 8887) National Conference on Emerging Trends in Computer Technology (NCETCT-2014)*.
- [11] M. F. Arlitt and C. L. Williamson, "Internet Web servers: Workload characterization and performance implications," *IEEE/ACM Trans. Netw.*, vol. 5, no. 5, pp. 631–645, Oct. 1997.
- [12] C. Kallepalli and J. Tian, "Measuring and modeling usage and reliability for statistical Web testing," *IEEE Trans. Softw. Engin.*, vol. 27, no. 11, pp. 1023–1036, Nov. 2001.
- [13] Tec-Ed, "Assessing web site usability from server log files," White Paper, Tec-Ed, 1999.