

# User Specific Web Usage Analysis on Web Navigation

Vrushali P. Bonde<sup>1</sup>, Prof. Krutika K. Chhajed<sup>2</sup>

*Student of M.E. (CSE), Assistant Professor, Comp. Sci. Dept., P. R. Pote College, Amravati  
vrushalipbonde@gmail.com, krutika.chhajed@gmail.com*

**Abstract** - As the World Wide Web becomes prevalent today, building and ensuring easy-to-use Web systems is becoming a core competency for all the online business activities. So to help these online business activities, a business model is proposed and develop in this paper, which keeps track of user's navigation on the web pages. Web navigation is the process of navigating network information in the World Wide Web. The paper provides a new method to identify navigation related Web usability problems. There are many reasons behind this explosive growth in web traffic. Unfortunately, not every website navigation is good. Most of the time, website navigation is put together by web designers who know a lot about making decorated websites, but very little about marketing a website or creating a website built from the users point of view, which results into web navigation usability problems for users. But as the web navigation to identify user's interest is very much essential, this paper tries to develop the interface for proper web navigation usability improvement. The dataset taken for this functionality, is generated from online shopping interface, which helps to identify the user's interest according to their products searching behavior. Here, the system uses usage mining with apriori algorithm for understanding user's interest and generating recommendations for them. Based on the users visiting behaviors the owners get the idea regarding user's interest, leading to the growth and development of online business.

**Index Terms** - Navigation Efficiency, Usability, Web Server Logs, Web Navigation, Web usage mining.

## I. INTRODUCTION

The World Wide Web today has expanded to serve millions of different users for a multitude of purposes in all parts of the world. Naturally, Web content nowadays needs to be filtered and personalized based on the particular needs of individual users. The user's interests, expectations and expertise, cognitive style and perception are some of the factors that need to be considered when creating personalized interactive systems. Web systems that are easy-to-use by all users. Usability of this system is defined by its effectiveness, efficiency, and satisfaction with which all the users and providers can complete specific tasks in a particular environment. The use of internet has become vital in the part of society, more and more users are switching to the internet for their all basic work such as shopping etc, so it is important for a business organization to analyses which users are using their systems. As many of the user are now a days preferring their work should get done by seating at their place, So, for all the business organizations and people it becomes necessary to increase their web based activities, regarding business work [1]. This should include Structural firmness in your website that influence the website security and performance. Your website should be functionally

convenient, that is available for all convenient characteristics, such as use and navigation.

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 a 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.

Here, the system propose a new method to identify navigation related usability problems by comparing Web usage patterns extracted from server logs against anticipated usage represented in some cognitive user models. The routing algorithm proposed here helps to identify uses navigation

pattern and the pattern matching algorithm helps to keep the log of user's session. By using user's session it is easy to identify user interest. As he develop interface is of online shopping, system user's apriori algorithm that helps to suggest related product to the customer, by extracting the users log pattern for his past history products. Along with adding all the functionality to our web based system, measuring usability from all functions leads to find out our strengths and issues occurring on daily basis is also necessary. 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 [3].

After that, by applying some computation technique by usage mining using apriori algorithm we take user sessions and transactions to extract usage patterns. This will lead to generate usability report helps to identify some issues and apply corrective measures on them leading to the development of our business. This system perform the navigation of user and retrieve the best solution from that, so it will set implemented for decision making in developing business plans and strategies. With the increasing use of Interest, 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. 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 people visiting different organizations portal. The remaining paper is organized as, Section II gives the problem definition and Section III gives the architecture of our proposed model. Section IV describe the algorithm used for generating recommendation from users log date. Section V gives the diagrammatic implementation of the proposed system followed by the steps of proposed working. Finally section VI, concludes the paper.

## II. PROBLEM DEFINITION

Different web mining technique used by different engineers mostly uses server logs for analyzing and understanding users behavior. But this server logs helps to carry out the tasks for improving their

experience of identification only. This server logs do not contain the user's goals. Along with this there are significant challenges exist like accuracy of problem identification due to false alarms common in expert evaluation, unrealistic evaluation of usability due to differences between the testing environment and the actual usage environment, and increased cost due to the prolonged evolution and maintenance cycles typical for many Web applications. So for extracting the useful and important users visiting behaviors and navigation patterns that keeps track of user's interest for development of business, some technique needs to be develop.

## III. SYSTEM ARCHITECTURE

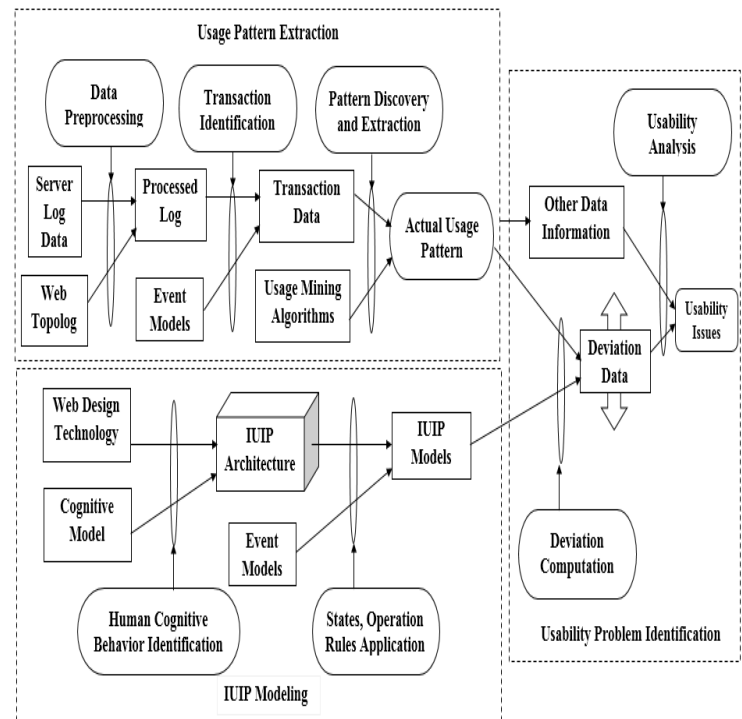


Figure 1: Architecture for identifying usability problems

System propose a new method to identify navigation related usability problems by comparing web usage patterns extracted from server logs against anticipated usage represented in some cognitive user models. Figure.1 shows the architecture of proposed method. It includes three major modules: Usage Pattern Extraction, IUIP Modeling, and Usability Problem Identification. First, we extract actual navigation paths from server logs and discover patterns for some typical events. In parallel, System

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 functional convenience aspect of usability for this website is quantified by its task completion rate and time to complete given tasks. The ability to implement recommended changes and to track quantifiable usability improvement over iterations is an important reason for us to use this website to evaluate the applicability and effectiveness of this method.

#### IV. ALGORITHMS USED

The algorithms used here is required to generate recommendation by extracting users log of data from the web server log generated by users visiting behavior on web page. The Apriori algorithm suits best for this task and generate proper output.

##### A. Apriori Algorithm

During the literature survey we found various frequent set mining algorithms are implemented for web personalization using proxy web access log mining. Additionally various other sequential and frequent pattern mining approaches are developed. But they not provide the sufficient performance for large data set. And as the usage of online activities for all the task is growing on increasing it becomes necessary to work with large volume of data. The Apriori Algorithm is an influential algorithm for mining frequent item-sets for Boolean association rules.

The traditional simple Apriori algorithm is most frequently used by different researchers and groups to mine log data. This algorithm has some problem with their performance we observe that when the item set are increased then the time and memory required is increased exponent manner. To overcome this problem we propose new Modified Apriori algorithm

Useful Terms:

- *Frequent Item-sets*: The sets of item which has minimum support (denoted by  $L_i$  for  $i$ th-Item-set).
- *Apriori Property*: Any subset of frequent item-set must be frequent.

- *Join Operation*: To find  $L_k$ , a set of candidate  $k$ -item-sets is generated by joining  $L_{k-1}$  with itself.
- Use the frequent item-sets to generate association rules.
- *Join Step*:  $C_k$  is generated by joining  $L_{k-1}$  with itself.
- *Prune Step*: Any  $(k-1)$ -item-set that is not frequent cannot be a subset of a frequent  $k$ -item-set

Modified Apriori Algorithm Pseudo-Code:

##### Variables:

$C_k$ : Candidate item-set of size  $k$

$L_k$ : frequent item-set of size  $k$

$L_1 = \{\text{frequent items}\}$ ;

##### Process:

For ( $k = 1$ ;  $L_k \neq \emptyset$ ;  $k++$ ) do begin

$C_{k+1} =$  candidates generated from  $L_k$ ;

For each log  $t$  in database do

If ( $t ==$  input set) then

{

Increment the count of all candidates in  $C_{k+1}$

}

Those are contained in  $t$

$L_{k+1} =$  candidates in  $C_{k+1}$  with  $\text{min\_support}$

End

Return  $\cup_k L_k$ ;

**V. SYSTEM IMPLEMENTATION**

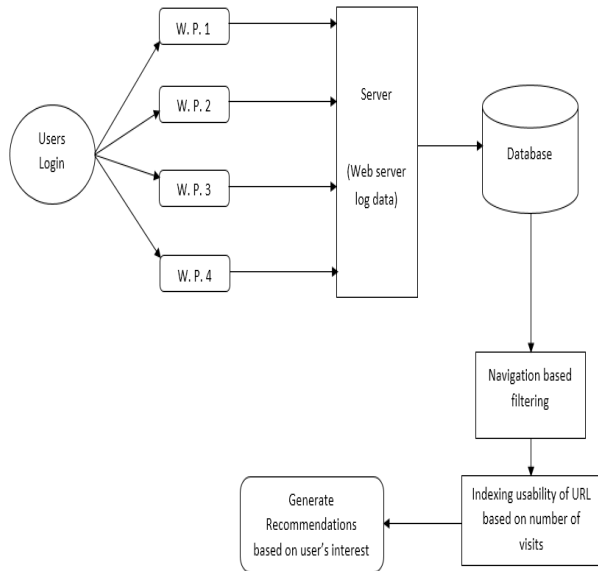


Figure 2: Dataflow diagram of proposed system

The above is the dataflow diagram of our proposed system.

Here, W.P represents the Web Page that the users are visiting after login the system. The server is used to store the log of users visited web page. For further calculations and recommendations database is used. Finally, user’s recommendations for web page is generated and displayed.

**Stepwise Workflow:**

The proposed system will work in following steps:

Step 1: For performing user’s specific web usage analysis based on web navigation, we have create the interface online shopping portal as business model.

Step 2: For making it user’s specific, firstly any new users has to register with the system. After successful registration only user can login the system for further use.

Step 3: After login the system user log gets maintain and now as an online shopping portal, users visited the different web pages according to their interest to see or purchase the products of different categories.

Step 4: As the users log is continuously updating till users logout the portal, his visiting behavior on different webpages gets stored on the server in the form of URL’s of the visited pages

Step 5: After that this server logs are stored in the database. Now, administrator of the system can see users visiting on different web pages along with page category, date and time of visiting.

Step 6: After getting these users visiting behavior, administrator of the system, can generate recommendation that can be used by business organizations to understand users interest and their visiting behavior.

Step 7: Finally, with this information any business organization helps to make changes if their customer demands for any changes in their current system. Along with this they can generate recommendation for their customers in future leading to the growth and development of their business.

**VI. RESULT ANALYSIS**

After successful implementation of our develop business model. Table 1 below gives the result after performing calculation of all user’s logs. The result obtained is of users visited web pages (may be in number of times) i.e. frequency of occurrences, their session time. The final column represents match ration which is the most useful feature to calculate and generate recommendations.

PageVisited	Occorance Frequency	Session time	Match Ratio
Electronics	24	38733	16.13875
HomePage	41	82402	20.09805
Cloths	25	56665	22.666
Baby	0	63190	0
Furniture	12	27560	0
Watches	21	97626	0

Table 1: Frequency and number of occurrences calculated

The method used for result analysis in the proposed system is with the help of match ratio, which always give more accurate result than session time that is used by existing system for identifying user’s behavior

The following formula used by proposed system to calculate match ratio,

$$\text{Match ratio} = \text{time} / \text{occurrence frequency} * 100$$

Here,

time - is the overall time spend by the user on the portal

Occurance freuency - is the the frequency for number of times user visited that page

This match ration is useful feature for the calculation of users visiting behavioire on any online protal.

- The existing system only uses session time as a major for calculating users behavioire but it is not perfect major. Because it is hard to get which page is visited perticularly in that session time.
- But the proposed system uses match ration for the calculatin of users visiting behavour, that takes both session time and number of occurrence as input for calculation, leads to more perfect generation of users visiting behavior.

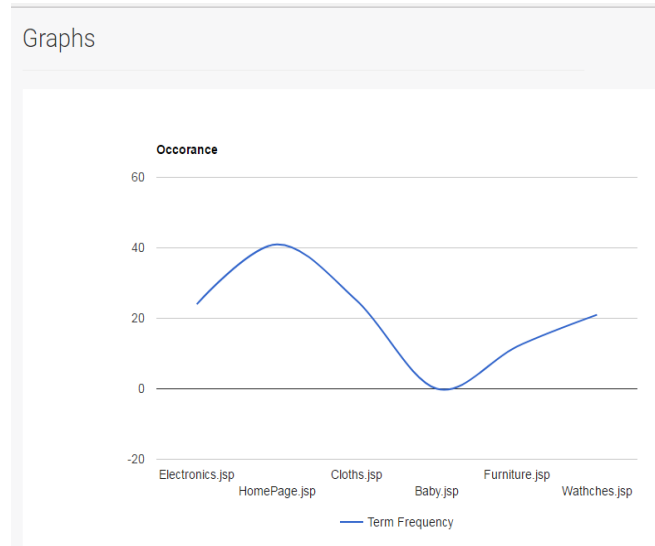


Figure 3: Category Occurrences graph

The Figure 3 is the resulting graph shows the number of occurrences of user's visits to different category of web pages. This results in clear understanding of user's behavior of visiting to the administrator of any online business organizations.

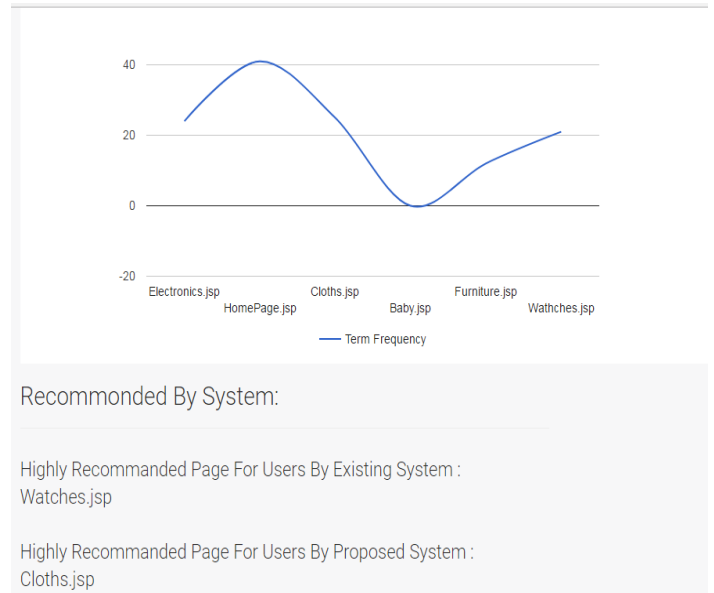


Figure 4: Recommendation generated based on frequency and occurrences

- The Figure 4 above gives the clear understanding of comparative analysis between existing and the proposed system develop here. As number of occurrence of clothes are more than watches, then the recommended system should be for cloths and not for watches.
- But as the existing system uses session time as parameter for calculation it will give results recommendation for watches category pages as more visiting pages.
- But, as there is slight difference in the number of occurrences of cloths category and as it is more in number than watches, but not recommended by existing system only because it required more perfect calculations.
- This drawback is overcome by our proposed model here. The proposed system develop here uses the match ratio as the parameter for calculation of recommendation generation. And it can clearly see that the proposed Concept producing proper results that satisfies our objectives.

## VII. CONCLUSION

The activities and transaction on web is going on increasing continuously with the advent of internet. These users log is much more useful for large number of business organizations that helps to understand user's interest which helps to increase their business. But there are different kinds of issues associated web usability problems that needs to be overcome in order to satisfy the growth and needs online business. This project identifies web usage mining as the part of data mining as an effective technique that helps to identify important web usage data. System proposed a concept that efficient business logic has been developed by implementing temporal deviation on anticipated data by mining of server logs. Which are applied with the usage navigation in association with apriori algorithm that helps to understand user's interest analysis. This will helps to achieve a best recommendation of web pages and products in e-commerce and other searches, leading to the growth of online business.

## 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](https://en.wikipedia.org/wiki/Server_log).
- [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] 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.
- [7] 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.

- [8] A. McDonald and R. Welland, "Web engineering in practice," in *Proc. 10th Int. World Wide Web Conf.*, May 2003, pp. 25–30.
- [9] Jaideep Srivastava, Prasanna Desikan, Vipin Kumar, "Web Mining — Concepts, Applications, and Research Directions", [Online]available:<http://dmr.cs.umn.edu/Papers/> October 2001, pp. 21–30.