

# Prediction of user Behavior through Web usage Mining

Naresh Kumar Kar, Subhash Chandra Shrivastava, Megha Mishra, Vishnu Kumar Mishra

**Abstract:** In real life, many users draw in towards online ticket reservation, so numerous transactions are in websites. A weblog includes sequences of entrances updating often by individual while opening the web site. Based upon the customer interest, it might be categorized as relevant & unassociated information. The relevant information might be thought about as success feedback; however, unassociated data could be deliberated as failing reaction. It analyzes the pattern of individual navigating while searching, for that internet use mining have to be examined.

The stages consisting for the procedure of internet usage mining are preprocessing, discovery of pattern, information collection, and pattern analysis. Along these steps this paper gave the user behavior analysis.

**Keywords—** Navigation pattern, Web usage mining (WUM), Web system mining (WSM), web page mining (WCM), Automatic Log Mining using Genetic (ALMG), hypertext transfer protocol (HTTP), Internet Use for Info Provisioning (IUIP) Modeling, hypertext mark-up language (HTML) code.

## I. INTRODUCTION

World Wide Web (WWW) is enormously in vogue and interactive. It's become a vital supply of understanding as well as services. The web is big, numerous as well as dynamic. Extraction of eye-catching details from internet expertise has ended up being in style and also as consequences of that web mining has actually drawn in great deal of interest in latest time [1] internet mining will be described approximately as information processing victimization expertise produced by web [2]. It is split in to 3 courses mainly they are WSM, WCM and WUM [3]. WSM attempts to determineweb link framework of links at the inter-document level as well as produces an architectural synopsis to look at knowledge connected with arrangement of a nominated computer gadget.

WCM mainly concentrates on inner-document structure to examine out useful information within content of sites like totally free message within an on the internet page, semi-structured knowledge such as pictures, HTML code, as well as downloadable documents. WUM makes an effort to discover valuable information from second knowledge, especially those had in blog documents. Various resources are internet browser logs, individual profiles, individual sessions,

**Revised Manuscript Received on October 22, 2019.**

**Naresh Kumar Kar**, Research Scholar, 1Department of CSE Rungta College of Engineering & Technology, Bhilai, Chhattisgarh, India, Email: nareshkar@gmail.com

**Subhash Chandra Shrivastava**, Department of CSE, Associate Professor, Rungta College of Engineering & Technology, Bhilai, Chhattisgarh, India, Email: subhash\_2911@rediffmail.com

**Megha Mishra**, Department of CSE, Associate Professor, SSGI, Bhilai, Chhattisgarh, India, Email: meghal6shukla@gmail.com,

**Vishnu Kumar Mishra**, Department of CSE, Associate Professor, SSGI, Bhilai, Chhattisgarh, India, Email: vshn123mshr@gmail.com.

book markings, folders as well as scrolls. These understanding are gotten from interactions of the customers with net. Efficient computer gadget monitoring, making adjustive internet websites, business and assistance services, customization, as well as network traffic circulation study expeditiously utilize WUM for higher proficiency. WUM concentrates on methods, which certainly anticipate individual's behavior. In [3], there are 3 important challenges for playacting WUM: Preprocessing, Pattern Exploration, & Pattern Evaluation. Due to the value of pattern discovery, this manuscript concentrates on defining this segment & also exists a recap of WUM and also conjointly offers a study of numerous methods of pattern extraction utilized for WUM.

Online ticket booking is a type of digital commerce that permits customers to reserve the ticket from a bus owner online, making use of an internet browser. Alternate names are: train ticket reservation, flight ticket booking. Mobile business defines regarding on the internet stores in application. Online clients frequently utilize a bank card to create transactions. Several online stores will certainly not accept credit cards. Some of them needcustomer's payment as well as shipping address to be in the very same nation. Various others on the internet buying allowing clients to any type of nation to send out gifts anyplace.

Consumers search the ticket by visiting the web site of bus proprietor straight or by searching in different suppliers by a ticket booking online search engine. As soon as a certain seat was found on internet site, a lot of on the internet ticket bookers will picked ticket reservation software application to permit the consumer to collect several seats as well as readjust seat area like loading a physical booking cart in a ticket reserving workplace. Some reserving office enables customers to enroll in a permanent account use. Less sophisticated stores might respond on customers to email or phone their orders.

Customer demands and also assumptions are not equivalent for all consumers. Age, sex, experience as well as society are all important variables for ticket booking. Individuals with on-line experience focus extra on variables, which straightly affect the task, though other individuals focus on comprehending the details. To enhance on-line ticket reservation, services will certainly be using significant money and time to specify, layout, establish, examination, apply and preserve the booking office. It is to get rid of mistakes and even more attracting online ticket booking, lots of web ticket scheduling developers study research study on consumer assumptions.

The remaining sections of this manuscript structured as follows. In the following section 2 connected deal with WUM & also pattern discovery have evaluated. In section 3 reviewed system design and also section 4 defines proposed work and outcomes. Lastly a verdict of this work is depicted in section 5.

### II. RELATED WORK

The work [4] predicted a brand new fertilization in 1996 that was described as "Web mining". He utilized information methods to automatic detection and also essence info from thick knowledge on earth Wide internet. WUM was early predicted by [5], [6], [7], [8], WUM system described as advice that maximizes the internet web server efficiency by giving practical information technique provides an objective behavior for individual navigation. The work [9] utilized the Andre Markoff chains to design customer's instructions practices. They forecasted a method for developing an Andre Markoff version of internet website supported earlier individuals' behavior. Then Andre Markoff design is working to form link forecasts that help with novel customers to browse the net web site.

The work [10] proposed a framework for removing consumer's navigational behavior utilizing a graph partitioning method. A plan less graph sustained property in among every shot of web sites was thought about and also but also predicted a novel formula for allocating weights to graph edges. In [11] the forecast of individual's navigation patterns will be forecasted victimization jumble & classify from blog site knowledge. Preliminary segment of this approach concentrates on dividing individuals in blog site knowledge, as well as within 2nd area cluster technique is working to gather the customers with related preferences and also within 3rd segment the outcomes of classification and jumble are wont to forecast the individuals' following requests. The work [12] discovered serial accessibilities from blog documents, (Algorithmic Guide Line) GA guideline that mentioned to as ALMG. In their work, GA sustained natural procedure technique for pattern extraction, was wont to located finest explanations for time extreme disadvantage to discover serial gain access to from blog site understanding.

Kim and Zhang make utilization of GA guideline to be told the important features of HTML tags that are wont to re-rank the documents recovered by popular weight systems for web structure mining. The work [13] gift a genetic search method for a mission engine. The work [14] proposed a hymenopterous insect agglomeration algorithmic guideline to locate web usage patterns (information collections) and also a straight genetic programming strategy to examine the traveler trends [15]. Several systems have established internet mining for automatic personalization [16], [17] as well as [18] they usually brings with it 2 significant procedures: off-line mining & on-line recommendation within off-line mining method, all accessibility tasks of customers in website are taped into log documents by web server. After that, few net mining procedures have applied to web server logs to extract hidden navigation versions of customers within online reference method; consumer's demands from his existing energetic session have recorded. The work [19] produced an endeavor to incorporate each use and material features of a website into web mining structure for internet personalization. A "post-mining" method was applied to attain the uniform

image for every website use as well as website material accounts to facilitate the amount of time customization. Nonetheless, the strategies forecasted in [14] were limited to employment of pile to independently build web site use and web content profiles. Authors in [20] have stated concerning web link forecast and also course examination for improved consumer navigations. He suggests a Markov chain method to calculate the individual access pattern sustained individual gain access to logs formerly collected. The work [21] introduces the fertilization of victimization the outside forward recommendations so regarding disturb down customer sessions into transactions for mining of traversal patterns. An outside ahead reference will be that last page asked for by customer before backtracking occurs, where as the individual demands a web page antecedently saw throughout that precise customer session.

The work [23] found that-- Weblog are evaluations for the way of IT managers to make sure appropriate data transfer as well as web server ability on business website. In the past 5 years, Log documents evaluation has actually progressed with firms; currently mining files have fine-grained detail regarding site visitor profiles as well as purchasing activity. The organizations are now looks for to utilize log files to discover the functionality of internet site. Log documents information might provide valuable understanding right into internet site use. It replicating actual use in all-natural functioning condition associated to synthetic setups of a functionality lab. It stands for task of several customers, over potentially long period of time checked out to restricted variety of individuals for a hr or more each.

The work [24] exposed that-- We report on individuals' revisited patterns to Net pages, and experiential structure for layout of background devices in internet browsers. Through background, it returns to visited web page by decreasing cognitive as well as physical overhead called for to navigate for scrape. Individuals has a tendency to revisit pages for opening websites regularly, browse in extremely small collections of pages, and produces short series of LINK paths. We contrasted various history mechanisms, as well as discovered the stack-based prediction technique widespread in industrial web browsers is substandard to easy method of just newly checked out URLs with matches gotten rid of.

The work [25] discovered that-- Functionality assessment of Web sites is time consuming task done physically. It offers a device that sustains remote functionality analysis of Internet site. It sustains client-side information's on individual communications & JavaScript events. Furthermore, the meaning of customized occasions offering examines the versatility to include particular events to spotted and also thought about for analysis. The tool sustains the Internet site by manipulating a proxy-based style as well as allows the critic to execute actual individual actions & optimal sequence of actions The work [26] proposed that-- Functionality of intricate vibrant human computer system interfaces might be reviewed by cognitive modeling for their underlying frameworks. Although the prediction of human attitude might examine the investigative mistakes in communication layout as well as cognitive needs of the future. This manuscript describes both principles as well as first applications. Both tools were revealed utilizing an instance of procedure control.

### III. ARCHITECTURE OF SYSTEM

In architecture of system, the customer will certainly search the internet site for online ticket booking. After opening a web site, they send request to Internet server & also internet server sends a reaction to internet browser. The HTTP is an application protocol for dispersed, joint, hypermedia details systems. The hypertext is a structured text, which utilizes sensible links in among nodes consisting of message. The collection of entrances among internet browser as well as internet server is saved in Web logs. Weblog is a raw log utilizing this pre-processing strategy will certainly be turned on. Figure 1. Shows the net usage mining.

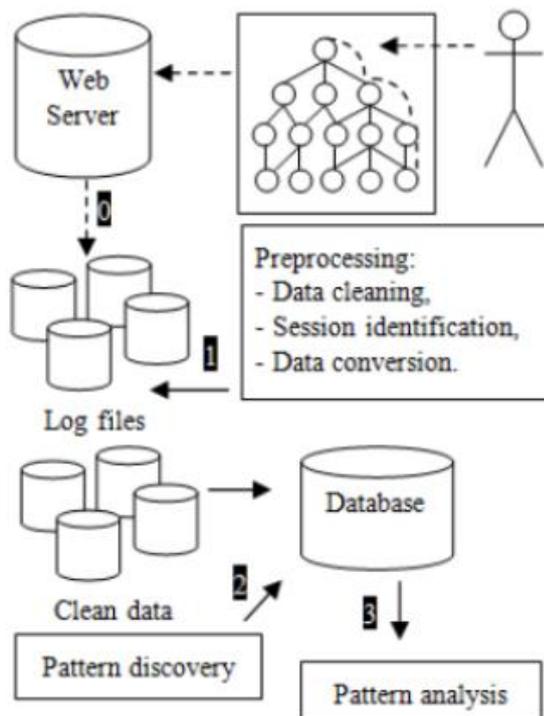


Figure 1. Net Usage Mining

On the internet ticket booking is a kind of electronic commerce it permits clients to get items or solutions from vendor over internet utilizing web browser. Few on the internet shops will certainly decline international credit cards. Some of them need customer billing & address of customer to be in very same country. Various on the internet ticket booking permit consumer's from any kind of country to send gifts wherever. Online clients have to open the internet as well as a legitimate technique of settlement to finish the purchases. Typically higher degrees of education and individual earnings correspond to more positive assumption of on-line purchasing. Raised disclosure to technology likewise raises the probability of creating advantageous attitudes towards the online reservation.

### IV. PROPOSED WORK AND DISCUSSION

The brand-new method to determine navigating relevant functionality difficulties by comparing Web use patterns extracted from server logs versus anticipated usage signified in several cognitive individual versions. It consists of 3 significant components: They are: Usage Pattern extraction, IUIP Modeling, and also Usability problem identification. Initially, it extracts actual navigating paths from server logs &

also determines patterns for distinctive occasions. In parallel, it will build IUIP designs for similar events. IUIP versions are cognition of client attitude as well as it might stand for anticipated paths for details client-oriented jobs. The outcome examining workers the device of examination oracle. The oracle will be used to evaluate that passed or stopped working. Here, we utilize IUIP designs as oracle to recognize the use concerns associated to individuals' actual navigation courses by assessing inconsistencies in among the two.

Weblog documents: Weblog data is log data immediately developed as well as maintained by an internet server. Every "hit" to Internet site, consisting of every vision of a HTML document, photo is logged. The raw web log documents format is fundamentally one line of text for every struck to internet site. A minimum of 2 log documents styles exists: Typical Log File style (CLF) as well as Expanded Log Documents layout,

- 1 The log file will be text file: Its records are same in format.
- 2 Each record in log file signifies a single HTTP request.

**Pre-processing:** The primary step of PUCC is pre-processing of internet log data, where as unformatted log information is exchanged a form, which might be straight related to mining procedure. The pre-processing stages contain cleaning, individual and session identification. The cleaning will be the procedure that eliminates all entries will certainly have no usage throughout evaluation or mining. Figure 2. Shows the preprocessing.

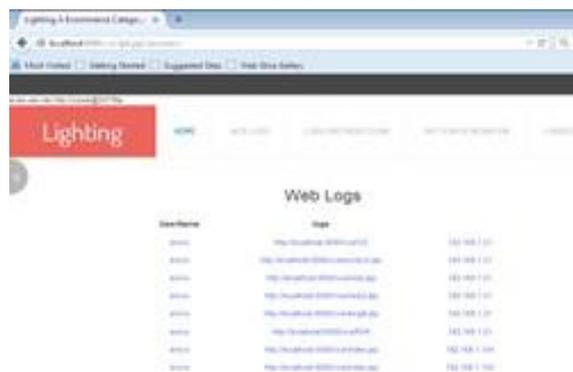
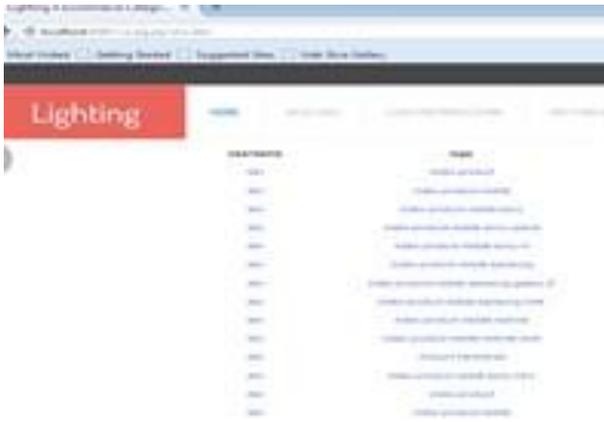


Figure 2. Preprocessing

**Identification of Potential Individuals:** This action of PUCC concentrates on dividing the potential customers from others. It makes use of decision tree category using C4.5 technique to determine interested customers. They make utilization of a collection of decision policies for this persistence. The procedure worked successfully in determining potential customers, but had disadvantages that it entirely ignored the entries made by the network robotics. Online search engine usually utilize network robotics to creep with the website to gather info. This issue is resolved in this manuscript by recognizing robotic entries first prior to segmenting the customer clusters into potential & also not-potential customers. Figure 3. Shows the identified individuals.

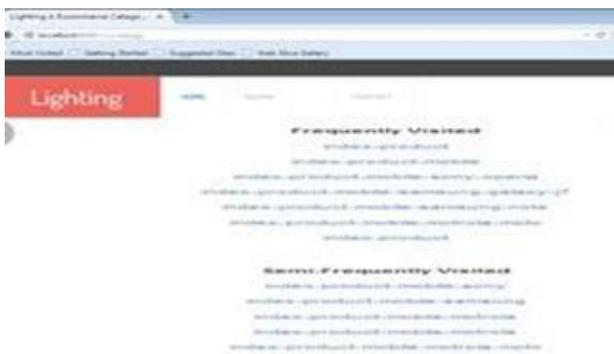


**Figure 3. Identified Individuals**

**Prediction Engine:** The primary objective of prediction engine is to classify user navigating patterns as well as forecasts individual's future requests. The major objective of prediction engine in this fragment of design is to identify customer navigating patterns and also forecasts customer's future requests. For this objective, we utilize LCS algorithm. The major objective of LCS is to locate the longest subsequence usual to all series in a collection of sequences. The algorithm collaborates with 2 functions. The initial residential property state that if 2 sequences X and Y with same aspect, then their LCS will certainly be establish by eliminating the last aspect and then locating LCS of reduced sequence. The 2nd property will be utilized when 2 sequences X as well as Y does not finish with similar symbol.

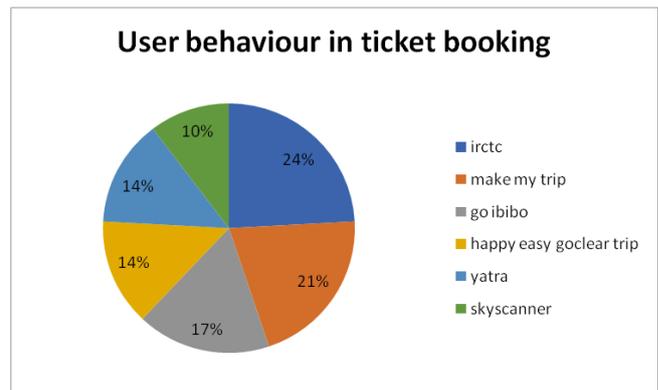
Algorithm Steps: (LCS Algorithm)

1. To determine a longest subsequence common to  $X_i$  and  $Y_j$ , the features  $X_i$  and  $Y_j$  are compared.
2. If equal, then sequence LCS ( $X_{i-1}$ ,  $Y_{j-1}$ ) is protracted by that element,  $X_i$ .
3. If not equal, then longer of 2 sequences, LCS ( $X_i$ ,  $Y_{j-1}$ ), and LCS ( $X_{i-1}$ ,  $Y_j$ ), is recollected.
4. If both of the similar length, but are not same, then both have recollected



**Figure 4. shows the various links after applying algorithm.**

In graph modeling, the statistical analysis & web usage mining are 2 paths to estimate utilization of Web site. The various online websites for ticket booking are shown in pie chart figure 5.



**Figure 5. User behavior analysis**

## V.CONCLUSION

The web log data contains raw information. Therefore, preprocessing the web log data is a considerable as well as one of the most standard action in web mining. It eliminates the undesirable items and acknowledges individuals with browsing information. The various patterns can be found by applying the mining strategies. The removal of relevant data from web browser will be recognized the regular & also semi-frequent consumers for online-ticket booking. Utilizing this kind of info, customers will certainly conserve time. In future, Internet utilization mining will be the base for navigation pattern mining as well as method of clustering will be utilized to usage mining compacts with discovery & evaluation of usage patterns from Internet data, particularly web logs, to develop web based applications as well as additionally to enhance the customer relationship administration. I want to carry out even more detailed experiments to further validate my technique and also surpass it.

## REFERENCES

1. B. Berendt. Web usage mining, site semantics, and the support of navigation
2. J. Borges and M. Levene. Data mining of user navigation patterns. In Proceedings of the WEBKDD'99 Workshop on Web Usage Analysis and User Profiling, August 15, 1999, San Diego, CA, USA, pages 31-39, 1999
3. R. Cooley, B. Mobasher, and J. Srivastava. Web mining: Information and pattern discovery on the world wide Web. In Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), 1997
4. R. Cooley, B. Mobasher, and J. Srivastava. Data preparation for mining world wide Web browsing patterns. Knowledge and Information Systems, 1(1), 1999
5. R. Cooley. Web Usage Mining: Discovery and Application of Interesting Patterns from Web data. PhD thesis, Dept. of Computer Science, University of Minnesota, May 2000
6. R. Cooley. WebSIFT: The Web Site Information Filter System.
7. Oren Etzioni. The world wide Web: Quagmire or gold mine. Communications of the ACM, 39(11):65-68, 1996
8. R. Kosala, H. Blockeel. Web mining Research: A Survey
9. B. Mobasher, R. Cooley, J. Srivastava. Automatic Personalization Based on Web Usage Mining. Communications of the ACM, Volume 43, Number 8 (2000)
10. S.K.Madria, S.S.Bhowmick, W.K.Ng, and E.P.Lim. Research issues in Web data mining. In Proceedings of Data Warehousing and Knowledge Discovery, First International Conference, DaWaK '99, pages 303-312, 1999

11. M.D.Mulvenna, S.S.Anand, A.G.Buchner. Personalization on the Net using Web Mining Introduction.Communicaitons of the ACM, Volume 43, Number 8 (2000)
12. M. Spiliopoulou, L.C.Faulstich, K. Winkler. A Data Miner analyzing the Navigational Behaviour of Web Users
13. M. Spiliopoulou. Web Usage Mining for Web site evaluation
14. M. Spiliopoulou. Data mining for the Web. In Proceedings of Principles of Data Mining and Knowledge Discovery, Third European conference, PKDD'99, P588-589
15. Abraham, A. and Ramos, V. "Web usage mining using artificial ant colony clustering and genetic programming," inHYPERLINK "http://libra.msra.cn/Conference/1848/cec-ieee-congress-on-evolution-ary-computation" IEEE Congress on Evolutionary Computation - CEC, 2003, 1384-1391.
16. Zhou, B., Hui, S.C. and Chang, K. "An Intelligent recommender system using sequential web access patterns," in Proceedings of the 2004 IEEE Conference on Cybernetics and Intelligent Systems, 2004, 1-3.
17. Burke, R. "Hybrid recommender systems: survey and experiments," in User Modeling and User-Adapted Interaction, 2002, 331-370.
18. Ishikawa, H., et al. "An intelligent web recommendation system: A web usage mining approach," in ISMIS, 2002, 342-350.
19. Mobasher, B., et al., "Integrating web usage and content mining for more effective personalization, " in First International Conference on Electronic Commerce and Web Technologies, 2000, 165-176.
20. Sarukkai, R.R. "Link prediction and path analysis using Markov chains," in 9th world wide web conference, 1999.
21. Chen, M.S., Park, J.S. and Yu, P.S. "Data mining for path traversal patterns in a web enviroment," in 16th International Conference on Distributed Computing Systems, 1996, 385-392.
22. M. Rauterberg, —AMME: An automatic mental model evaluation to analyse user behaviour traced in a finite, discrete state space, Ergonomics, vol. 36, no. 11, pp. 1369–1380, 1993.
23. P. M. Sanderson and C. Fisher, —Exploratory sequential data analysis: Foundations, Human–Comput. Interaction, vol.9, nos. 3/4, pp. 251–317, 1994.
24. 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.
25. T. Arce, P. E. Roman, J. D. Velasquez, and V. Parada, —Identifying web sessions with simulated annealing, Expert Syst. Appl., vol. 41, no. 4 , pp. 1593–1600, 2014.
26. 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. 446–451.