iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35037

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Tracking User Interaction with Web and Assisting in Targeted Communication

Dheeraj Ahuja[1], Madheshi Chinmay[2], Krishna Chandra Tripathi[3], M. L. Sharma[4]
[1, 2, 3, 4]*Information Technology, Maharaja Agrasen Institute of Technology Delhi, India*

Abstract: ***Today, we spend most of our time online using some form of digital technology (such as search engines, news portals, or social media sites). Our online presence keeps us involved most of the time and provides a lot of information to Internet customers.***
The development of the web is excellent because every day about a million pages are added. Due to the massive use of the network, the log files of the network increase at a faster rate and the scope becomes enormous. Web Usage Mining uses mining technology on log data to extract user performance, which is used in different applications such as support design, e-commerce, service modification, prefetch, etc.
In this paper, we propose a tool that users can use to collect data on their website, and then use this web log data to track user interactions on your website, which helps in targeted communication.

I. INTRODUCTION

The popularity of the Internet is getting higher and higher, and it provides a large number of services in various fields. One of the most profound services on the Internet is the World Wide Web, which has become the main tool for almost all activities that require a knowledge base.

These portals generate large amounts of data. These data include various records and details of transactions. With this information, market analysis and forecasting can be carried out very accurately. Many organizations and companies also use web portals as a key tool for sharing and accessing information and creating networks/collaboration.

Because the web process is easier to store and retrieve information, the opportunities for data mining are also enormous. If the records are also accessible, it is even more beneficial.

The information extraction method of the web usage mining application extracts the learning from the web logs. Web mining can be divided into web content mining, web structure mining, and web usage mining. Web content mining is a method of separating data from a wide range of collected information metrics. This strategy extracts data from the essence of the web page. Web content mining is again organized as interactive media mining and network print mining. Web structure mining is a method of analyzing the connections between web pages through the web structure. Web frame mining is divided into super interface mining and internal frame mining. Use web mining to verify network logs and to find sample client programs. Web usage mining is also called web log mining.

Understanding customer acquisition examples can help webmasters fine-tune website content, thereby expanding customer reading knowledge. The web routing example can be described by breaking down the web server log that contains the basic data for the client reading exercise. The blog occasion consists of the requesting client's IP address, the date and time of the request, the technology used in the request, the status code, and the request record. Quick exercise arrangements can be categorized into conversations and then used for further investigation to elicit the client's intended behavior. Decomposing online viewing behaviors of customers is one of the basic research and testing topics of web usage mining. If customer advantages can be identified naturally from the customer's online registration information, they can be used to suggest and present data that is valuable to customers and website engineers.

Many experts in the field of network usage mining have completed the analysis of web log information; they are constantly looking for examples of registration information. The main motivation behind studying customer visit logs is to decompose the results of the mining results that customers care about most. By breaking the customer's access to assets over time, repetition, etc., the site structure and plan are changed to predict that more customers need to stay and better serve customers. Customer behavior research has become another entry point for the exam.

This research work mainly considers web mining technology in order to track user interaction with the web.

II. LITERATURE REVIEW

A. Related Work

Web usage mining [2] is the use of data mining strategies to discover the usage patterns of Web information in order to better understand and meet the needs of Web-based applications. Web usage mining comprises of three stages [3], specifically preprocessing, pattern discovery, and pattern analysis**.** This paper depicts every one of these stages in detail. Given its application potential, Web use mining has seen a fast expansion in interest, from both the research and practice communities. This paper provides an in depth taxonomy of the work in this space, including research efforts as well as commercial offerings.

Web Mining can be isolated in to three classes specifically Web Structure Mining (WSM), Web Content Mining (WCM) and Web Usage Mining (WUM).

WSM tries to get the link structure of the hyperlinks at the inter-document level and generates a structural outline to look at information associated with the structure of a specific website.

WCM principally focuses on the structure of inner-document to seek out helpful information within the content of sites like free text within an online page, semi-structured knowledge like hypertext mark-up language code, pictures, and downloadable files.

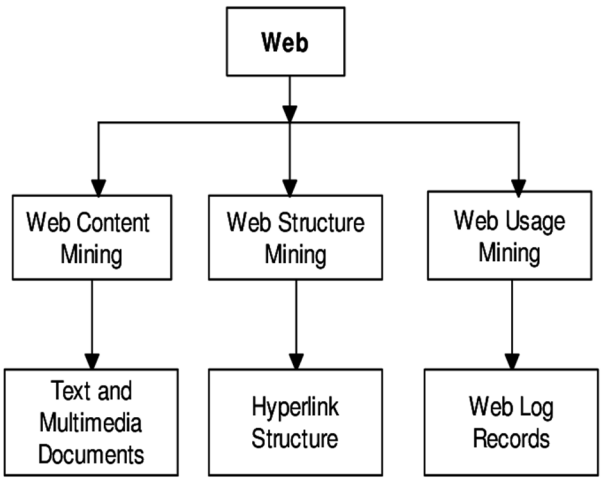
WUM makes an attempt to find helpful information from the secondary information, particularly those contained in Web log files. Alternative sources may be browser logs, user profiles, user sessions, bookmarks, folders and scrolls. This information is obtained from the interactions of the users with the web Effective website management enables responsive websites, business and support services, personalization and network traffic analysis to quickly use WUM to improve performance. WUM focuses on the techniques that may predict user's navigation behavior.

Different web usage mining techniques are mentioned in [4] that may be used to extract patterns from web log files. The Discovered patterns are used for pattern analysis that helps in understanding the user behaviors. As per [5], density-based clustering algorithm has been accustomed to discover navigation patterns. K-Nearest Neighbor algorithm with inverted index has been advised for economical prediction. Thus, many ways from data mining are utilized in the world of accustomed to usage analysis.

DOBBS [6] uses a browser add-on that permits researchers to log browsing behavior of on-line users, capture relevant and different window, session and browser events in anonymous and privacy-preserving manner and send those events to the server. In Dobbs, event is a unit of information. This paper describes all the logged events as well as window events, session events and browsing events. Window events includes events e.g. the opening and shutting of a browser window or tabs and changing within the state of browser window.

Passive browsing is that the time of idleness or inactivity throughout a user's browsing sessions. Parallel browsing is opening of multiple tabs inside one browser window and changing among them. Authors in [7] have analyzed in their study the impact of parallel and passive browsing on the calculation of user's time spent on web page.

Kosinski et al. [8] show that there's a mentally vital association between character, website and site categories. As indicated by this paper, outgoing clients' regularly visit sites known with Music and Social, whereas loners incline toward sites known with funnies, writing, and films. In addition, innovative and liberal are pulled in to journal, media, culture, crystal gazing, eBooks and expressive arts.



B. B-Related Applications

1) *Personalization:* Customizing the Web experience for a client is the sacred goal of many Web-based applications, for example individualized showcasing for web based business (Making dynamic proposals to a Web client, in light of her/his parole notwithstanding utilization conduct is alluring to numerous applications). Web utilization mining is a great methodology for accomplishing this objective. A few tasks learns a client's inclinations by watching the page accesses for every user. A rundown of watchwords from pages that a client has invested a lot of energy seeing is gathered and introduced to the client. In view of input about the watchword list, suggestions for different pages inside the site are made.
2) *Updating in System Performance :* Performance of internet services may be an important drawback for user satisfaction. Web usage mining is a vital analysis space for locating internet traffic procedures, which may be accustomed amplify new policies for raising the web server concert [9, 10]. Web caching technique transmission or account distributions, Load balancing, are the various application areas of web mining for performance upgrading.
3) *Site Restructuring:* The link association and content composition of any web site are two major factors in any internet site. the present trend in web mining tools go towards shorter navigation sequences, in support of that motive the accessibility of objective page in many internet domain must be improved [11]. Restructuring website topology of any internet field will achieve this. The restructuring task is performed with relevance the frequent patterns unconcerned at the end of web usage mining. Web usage data offers information in conjunction with the design of any web site with interests to user's activities [12]. To show page-stay moment offers the pages, that isn't attractive. Vendor of web site will re-establish these pages and observe the behavior of users on these pages. These two organizations techniques content and structure results in adaptive internet sites. The model given in changes web site organization with relevance usage patterns uncovered [13, 14].
4) *Supporting to the Design:* The Usability is most significant problems within the style and accomplishment of websites. The implications formed by web Usage Mining strategies will present ways for up the look of web functions. The utilize stereogram estimate the association and therefore the efficiency of websites from the clients view-point [15]. Web Usage mining strategies to propose acceptable variation for websites.
5) *Internet Business:* Mining E-commerce intelligence from web usage information is considerably vital for web-based businesses. A good advantage from the utilization of web Usage Mining strategies will have client Relationship Management (CRM). Within this case, the aim is on business explicit issues such as: customer removal, customer attraction, cross sales, and customer retention [16] [17] [18].

III. METHODOLOGY

In this paper we present a way that will help you automated the mapping of user behavior on your webpages. It will use the concept of web usage mining for collection of web log data, socket programming for the communication between server and client and selenium web driver for the realtime automation of the web logs already generated at the server side.

A. Socket Pipeline Connection

Socket programming is an approach of connecting 2 nodes on a network to speak with one another. One socket (node) listens on a specific port at an ip, whereas different socket reaches out to the opposite to make a connection. Server forms the listener socket whereas client reaches out to the server.

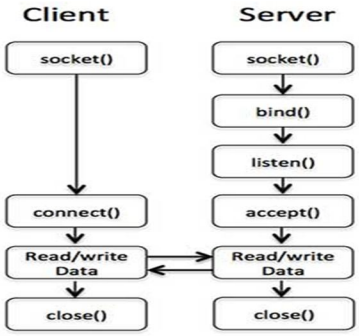


Figure2: State diagram for server and client model

As soon as the client connects from our server, we establish a socket pipeline connection between the server and the client and we capture few details about the user such as connection start time, user fingerprint containing various details such as browser, device, operating system, geo-ip, device-window-size, etc.

B. Collection of Web Log Data

After collection of user's basic details we start storing user action that one performs on the web page. These include clicks, key press, and scrolling.

These actions are stored separately as one json-object. Each individual json action object consists of three fields – time, action and value.

Time field stores the value of the time difference between the previous action and the current action. Action field consist of one of the three action types i.e. click, key press and scroll. Value field consist of the value corresponding to these action types, click action stores the coordinates where the click action has been performed, key press event stores the key that has been pressed and scroll action stores the amount in pixels that the page has been scrolled.

Table I: Attributes of Web log data that helps in Automating User Behavior

Attributes	Behavior
Start Time	Stores the start time of the user session
Browser	Tells which browser has the user used
Os	Tells which Operating System has the user used
Window Size	Signifies the size of user browser
End Time	Stores the start time of the user session
User Actions	Records the time gap, user action type and values of user action

C. Automating the process using Selenium Web Driver

Web Driver drives a browser natively, as a user would, either locally or on a remote machine using the Selenium server, marks a leap forward in terms of browser automation.

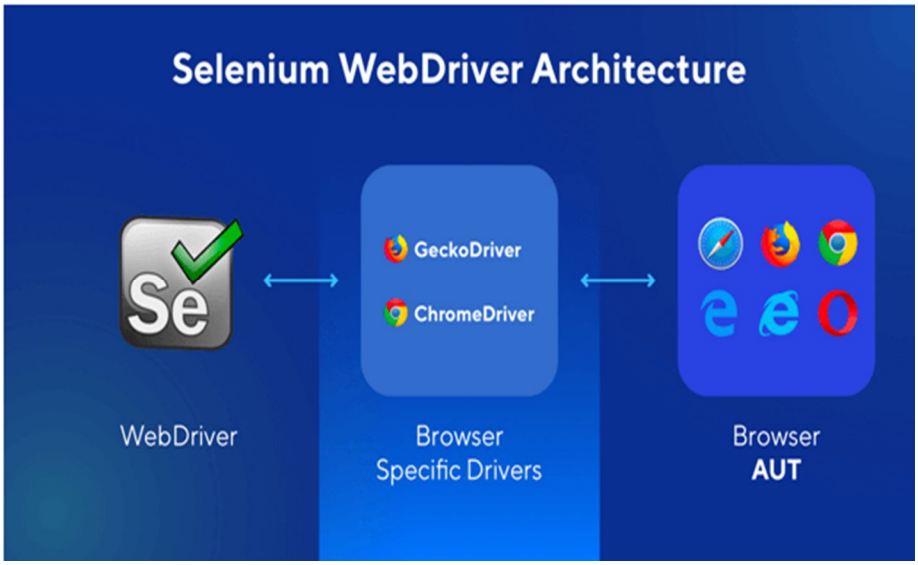


Figure 3: Components of Web Driver

The web log data file that has been created on the server and stores the user's details is now passed on to the selenium web driver script which starts reading the file line by line and mimics the action performed by them onto the web page on that web browser. Each action json object has three fields time, action and value so corresponding actions are executed after taking the pause (in millisecond) as stored in the time field.

D. Socket Pipeline Connection and Collection of web log Data

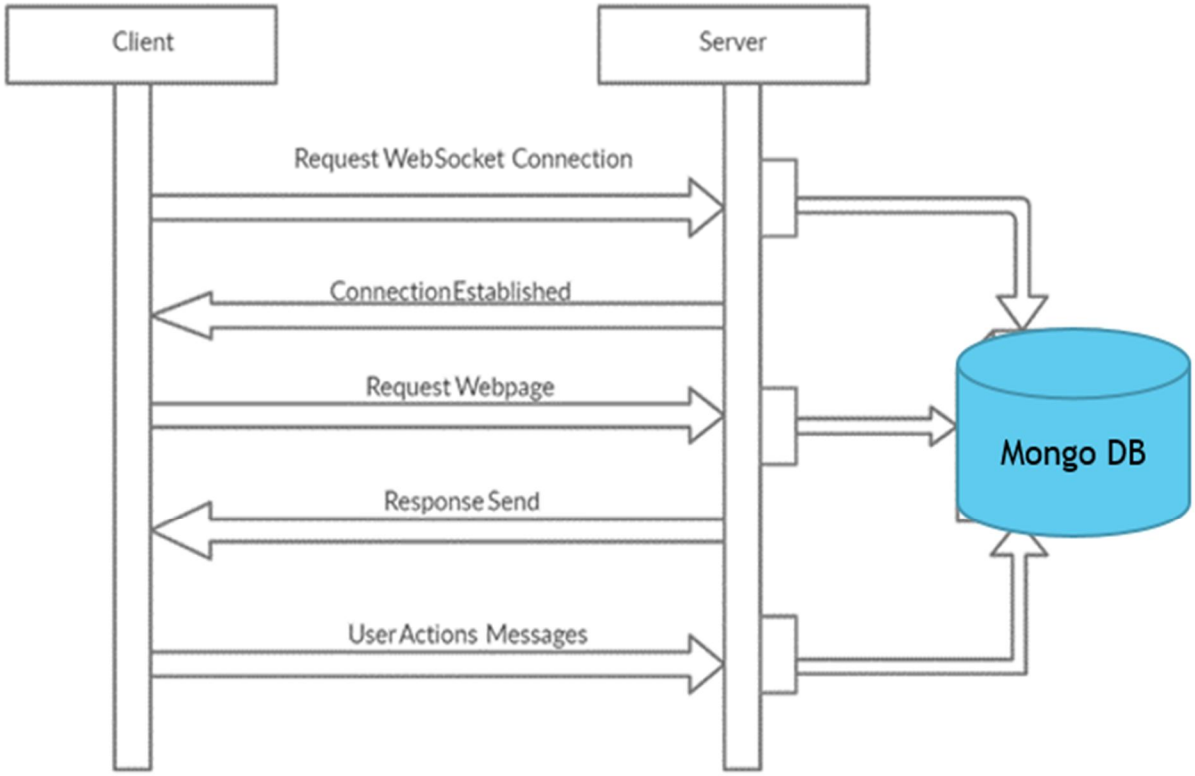


Fig4. The socket connection of the client and server and for the development of json object file has been shown here.

E. Automating the User Behavior Using web log Data

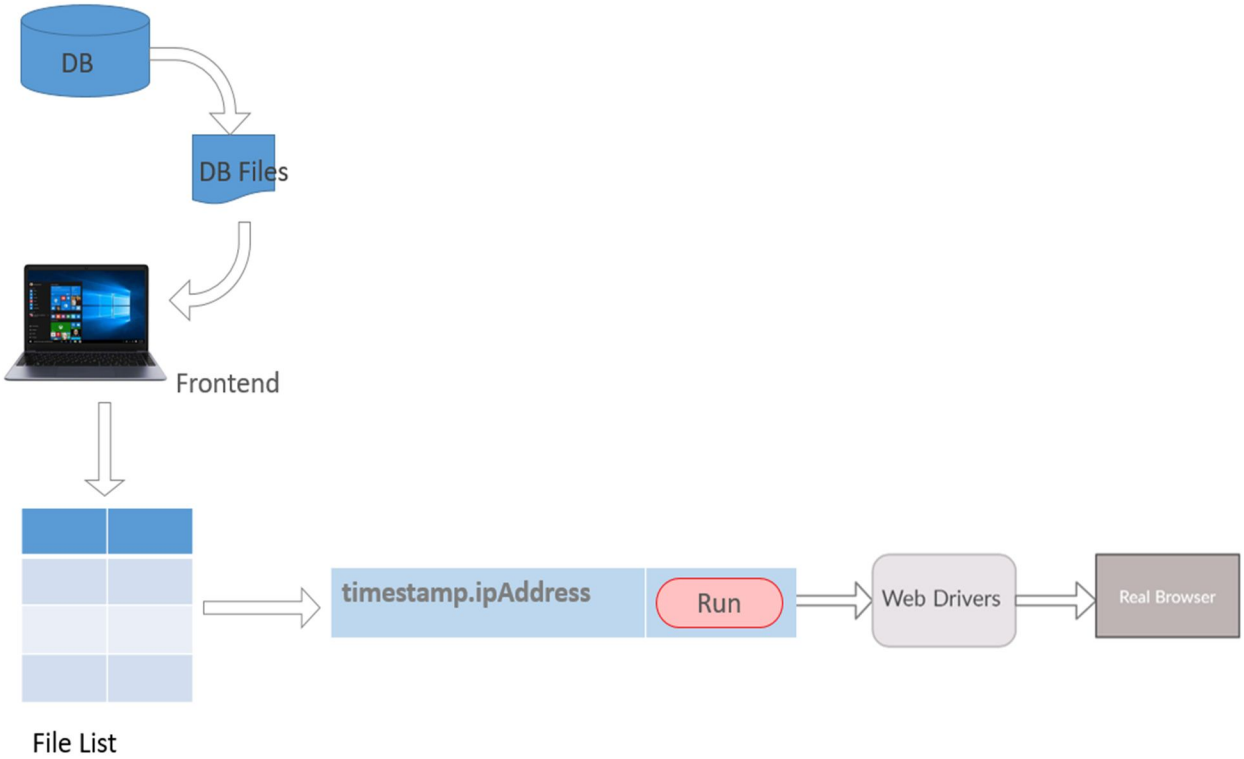
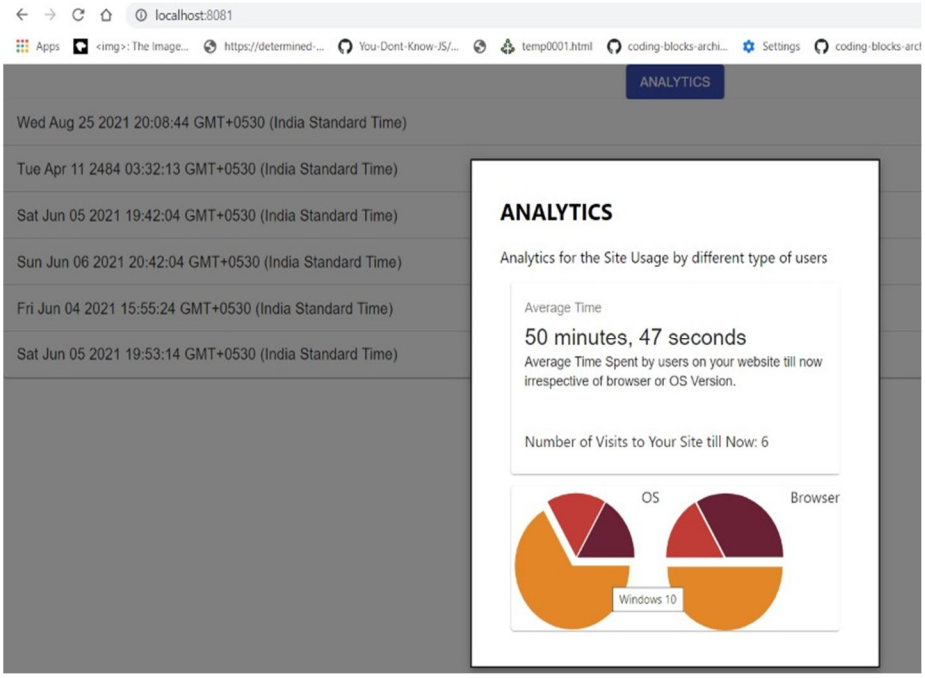
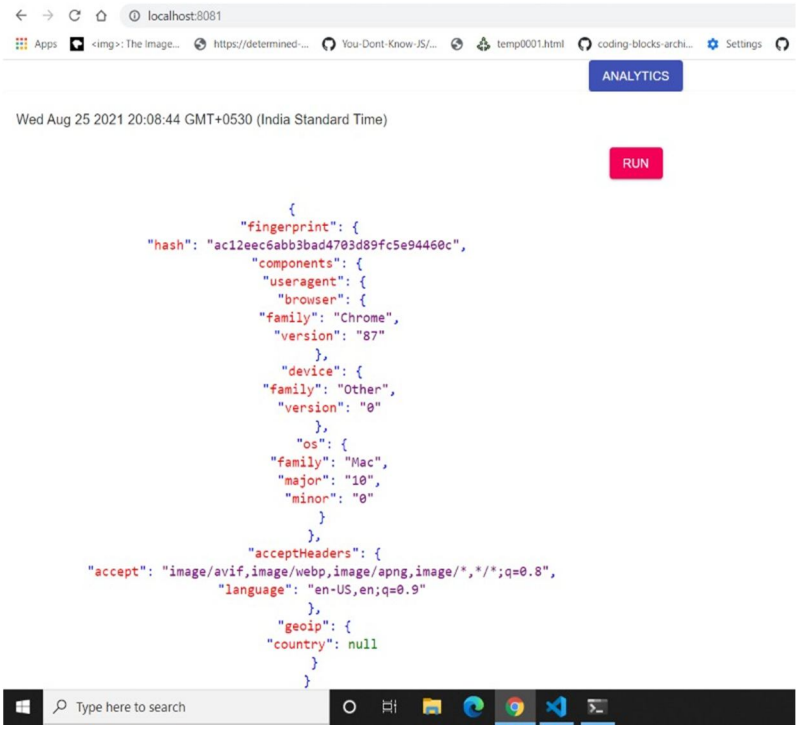


Fig5. The architecture for the automation process has been shown in the below diagram.

IV. RESULTS

In this paper we have implemented a tool using a blend of different concepts like web usage mining and technologies like socket programming and web drivers to automate the user behavior and there by getting some insights for targeted communication with the customers.





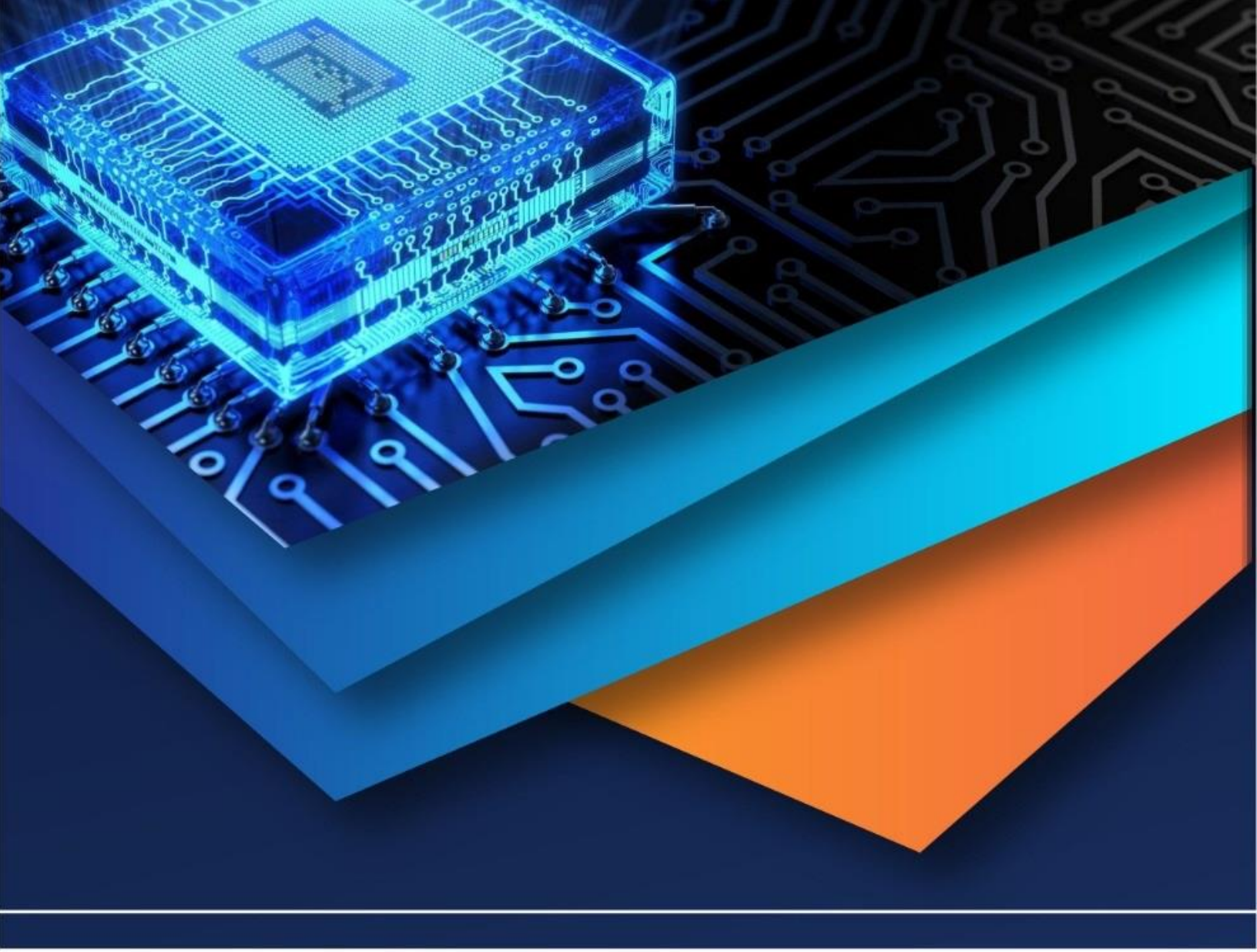
This paper has provided another application of web usage mining. With a large number of web applications being deployed daily, there is significant interest in analyzing Web usage data to better understand Web usage, and apply the knowledge to better serve users .We have used the concept of web usage mining for understanding the user behavior and automate it for getting insights related to user behavior.

REFERENCES

[1] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, "Web usage mining: Discovery and applications of usage patterns from web data," ACM SIGKDD Explorations Newsletter, vol. 1, no. 2, pp. 12–23, 2000.
[2] Bhupendra Kumar Malviya, Jitendra Agrawal, "A Study on Web Usage Mining Theory and Applications" 2015 Fifth International Conference on Communication Systems and Network Technologies, pp. 935-939
[3] Farhana Seemi1, Hania Aslam2, Hamid Mukhtar3, Sana Khattak," Browsing Behaviour Analysis using Data Mining "International Journal of Advanced Computer Science and Applications, Vol. 10, No. 2, 2019,pp.490-498
[4] M. Jafari, F. S. Sabzchi, and A. J. Irani, "Applying web usage mining techniques to design effective web recommendation systems: A case study," Advances in Computer Science: an International Journal, vol. 3, no. 2, pp. 78–90, 2014.
[5] P. Mehta, S. B. Jadhav, and R. Joshi, "Web usage mining for discovery and evaluation of online navigation pattern prediction," International Journal of Computer Applications, vol. 91, no. 4, 2014.
[6] C. von der Weth and M. Hauswirth, "Dobbs: Towards a comprehensive dataset to study the browsing behavior of online users," in Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on, vol. 1. IEEE, 2013, pp. 51–56.
[7] "Analysing parallel and passive web browsing behavior and its effects on website metrics," arXiv preprint arXiv:1402.5255, 2014.
[8] M. Kosinski, D. Stillwell, P. Kohli, Y. Bachrach, and T. Graepel,"Personality and website choice," 2012
[9] Etminani, K.; Delui, A.R.; Yanehsari, N.R.; Rouhani, M.; "Web usage mining: Discovery of the users' navigational patterns using SOM", First International Conference on Networked Digital Technologies (NDT '09), Pp. 224 – 249, 2009.
[10] Pierrakos, D., et al, "Web Community Directories: A New Approach to Web Personalization", 1st European Web Mining Forum, pp. 113-129, 2003.
[11] Jing Wang, Ying Liu, Yong Shi, and Xingquan Zhu, Pushing Frequency Constraint to Utility Mining Model, ICCS 2007 Springer Verlag Berlin Heidelberg, Part III, LNCS 4489, 2007, 685-692.
[12] Kudelka, M.; Snasel, V.; Lehecka, O.; El-Qawasmeh, E.; "Semantic Analysis of Web Pages Using Web Patterns", IEEE/WIC/ACM International Conference on Web Intelligence, Pp. 329 – 333, 2006.
[13] Eirinaki,Magdalini,and Michalis Vazirgiannis, "Web mining for web personalization", ACM, vol. no. 1,pp 1-27,2003.ms and Knowledge Discovery (FSKD '08), Vol. 1, Pp. 52 – 56, 2008.
[14] Jain Pei, Jiawei Han, Behzad Mortazavi_asl and Hua Zhu, Mining Access Patterns Efficiently from Web Logs, PacificAsiaConference on Knowledge Discovery and Data Mining(PAKDD'00), Kyoto, Japan, 2000, 396-407.
[15] Chu-Hui Lee, Yu-lung Lo, Yu-Hsiang Fu, "A novel prediction model based on hierarchical characteristic of web site", Expert Systems with Applications 38, 2011.
[16] A. Anitha, "A New Web Usage Mining Approach for Next Page Access Prediction", International Journal of Computer Applications, Volume 8– No.11, October 2010
[17] Mehrdad Jalali, Norwati Mustapha, Md. Nasir Sulaiman, Ali Mamat, "WebPUM: A Web-based recommendation system to predict user future movements" Expert Systems with Applications 37, 2010.
[18] TrilokNathPandey, Ranjita Kumari Dash, Alaka Nanda Tripathy ,Barnali Sahu, "Merging Data Mining Techniques for Web Page Access Prediction: Integrating Markov Model with Clustering", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 6, No 1, November 2012.

WEB RESOURCES

[1] Socket IO https://socket.io/
[2] Selenium https://www.selenium.dev/



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY