



UNDERSTANDING USER BEHAVIOR THROUGH WEB UTILIZATION MINING

Satyam Ramawat

Department of Computer Science, Christ University, Bangalore, India

Prof. Deepthi Das

Department of Computer Science, Christ University, Bangalore, India

ABSTRACT

This paper is focused on the efficiency of applications in Web Mining Algorithm for web log analysis which can be applied to identify user behavior and the context related to the pattern of an e-commerce website. The main objective is to find user behavior over the internet which can be done through the branch of web mining i.e., web utilization mining. So, it is mainly focused on learning the behavior of web user with their interaction with websites and application of web to extract hidden knowledge and information from World Wide Web. This motive can be accessed from the huge web log data and this data will be available in a log file of the server. Thus, the web server log file includes hidden knowledge and information of a visited user on web. This article provides a survey and analysis of current web mining system and technologies. Therefore, we implement how web mining techniques can apply to the web personalization by providing an efficient algorithm.

Key words: Web utilization mining, Customer behavior, web server log-file, Apriori all.

Cite this Article: Satyam Ramawat and Prof. Deepthi Das, Understanding User Behavior Through Web Utilization Mining. *International Journal of Civil Engineering and Technology*, 9(3), 2018, pp. 522-532.

<http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=9&IType=3>

1. INTRODUCTION

The Web Mining field permit data scientist to extract the hidden expertise and records of each and every user who visited the website. Web content mining, Structure mining and Utilization mining are various recorded branches of WEB MINING.

"Web utilization mining is the method of data mining used to discover the hidden data of a client from web information, to comprehend and better serve the necessities of a web-based application."

The web utilization mining plays a vital role for e-advertising and online commercial enterprise specialists. Examination of the user behavior can deliver a comprehension of

information to adjust and personalize consumer's web association. The world's main e-trade corporations have also started using this approach of web mining.

The data which is acquired from web utilization mining is very important for further research. It facilitates to discover hidden data of visited consumer from web browsing sample and examine the activities done by them. It includes numerous levels; Firstly, pre-processing of data. Secondly, Pattern discovery from obtained data. Lastly, The analysis of discovered data. Basically the server log report is cleaned and converted into standard format by the method of pre-processing and then algorithms of data mining are implemented to extract the hidden knowledge and data. After the preprocessing stage, it leads in finding the user interest.

The pre-processing stage of the data will be redefined and split in set of consumer actions which shows the moves of every single consumer when they have visited the website.

In the pattern discovery stage, the classification operations are performed to get the concealed data of users which shows activity of a particular user. Apart from this it also gives a detailed statistic of a user session and scheme.

In the pattern analysis stage, the acquired data and hidden knowledge information are further improved, clarified, reformed. The data which is acquired after this process will be used for various approaches in web analytics. In figure 1 gives the phases of web utilization mining.

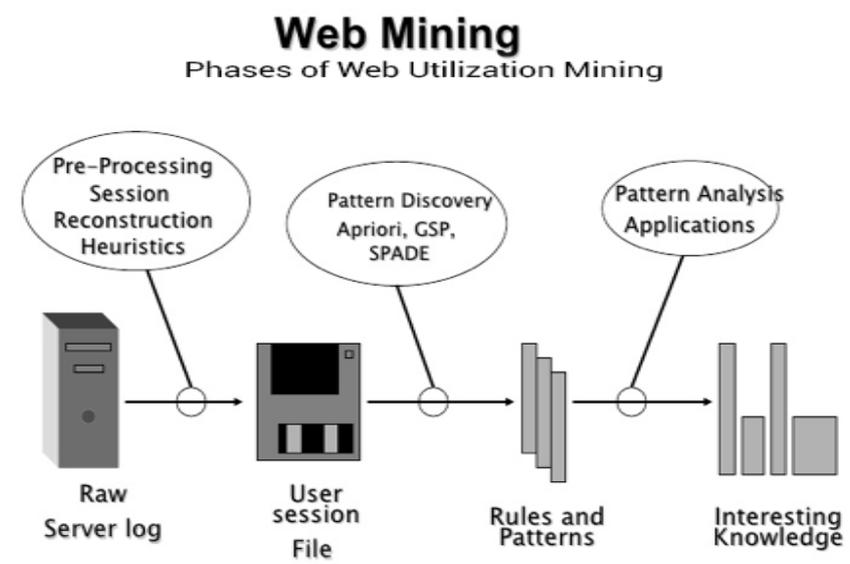


Figure 1 Web utilization mining Phases

2. METHODOLOGY AND LITERATURE REVIEW

[1] The paper “Internet Marketing research: opportunity and problems”, the author Oliver Furrer and D. Sudharshan et al., depicted that world wide web is an enforcing place to do research, it gives several possibilities for some issues. The creators have constructed up a structure to assist the analyst to apprehend the potential possibilities and issues which might be recognized with internet marketing. In addition, this article has mentioned three analyzing units in a web promoting review; site pages, sites and web client.

First of all, a typology of internet promoting look at configuration has been produced, which has 8 distinct plans in mild of the goals of the specialist; the unit testing, and the statistics accumulation. Then, 7 levels of the testing system of web are associated. After that,

researchers have dealt with primary issues which is totally centered on web surfing. Finally, this paper infers that web offers a remarkable chance to advertising and marketing experts.

In this paper the authors concluded that web site and it's page are essential to any research in the world of web. Therefore, it will help to find the hidden data and information of the users, who surf's webpages.

[2] In the paper "Integrating E-Commerce and Data Mining: Architecture and Challenges" by Ansari and Kohavi, et al., it is far portrayed that the area like internet business can deliver amend strategies which can be required for proficient data mining. Creators have defined a consolidated architecture that can colossally reduce the endeavors for the expertise.

Here authors had successfully designed a model which joins the 2 Data mining and a web-primarily based business framework. The proposed version is divided into 3 parts: enterprise information, consumer communication, and consumer evaluation. The coordination of three 3 sections, permits the automated development of a statistics block within evaluation segments which exhibits the acquisition pattern of customers in numerous age gatherings.

[3] The author Ahmad tasnim et al., In their paper "Web mining techniques in E-Commerce Applications" instructed that internet is tremendous for doing massive commercial enterprise inside the marketplace. It is moreover portrayed that the huge corporations around the globe have valued that web-based commercial enterprise isn't just a website for trading of goods online, although it enhances the effectiveness to contest the world of internet market. For realizing more data from the internet, they've applied the web mining methods.

They have developed the framework which carries four massive apparatuses; accumulating of data through consumer's interaction, commercial enterprise records, evaluation, and distribution center. The technologies which are used in the development of the framework is Microsoft.Net with the implementation of base class library like: HttpWebRequest and HttpWebResponse. These libraries help in the process of request from the website.

Here, this research paper prescribes that in the field of data mining we need to work on semantic web and ontology domain, content conveyance and creation so that we can identify and understand the appropriate attributes and data which is generated via Web utilization mining.

[4] From the above research paper we learned web mining and it's three major fields and realized, Web Utilization Mining is the proficient subject for performing our experiment and implementation. In the article "Overview on web mining and the different technique for web personalization"; author Pradnyesh et al., focuses on 6 typical errands of web utilization mining and actualization for customization web by using web mining strategies.

This paper taught us the preprocessing of the dataset by actualizing the accurate characteristic which are needed. From this article, we have also understood to perform analysis on the data which helps to lead in finding the hidden information of the visited user. The major focus is on "Personalization on the web mining" which is an incredible procedure for marketing as defined by the authors. It needs verifiably & expressly an accumulation of the guest data to actualize knowledge into the framework, which response in what data we need to demonstrate for dynamic clients.

At last, this research paper taught us the model of web utilization mining, which will provide efficiency in mining the important data from web log file to extracts the hidden data

of guest user which is very important resource to value for further Web utilization mining structure.

[5] We have figured out till now that how to pre-process the data and change it in the form of important necessary knowledge by the system of web utilization mining. Here in the research paper by Shanti et al., “An efficient web mining algorithms to mine web log information” manages different algorithms of web mining.

In this work, the creators find and recognize the co-related data of web user by applying web mining algorithms which is based on e-commerce web interface. They have done the investigation of various algorithms in the field of web utilization mining.

[5,6] The study is done by comparing the time complexity of algorithms. The authors have also provided the graph, which is based on performance with the related properties like number of transaction and items of the algorithms.

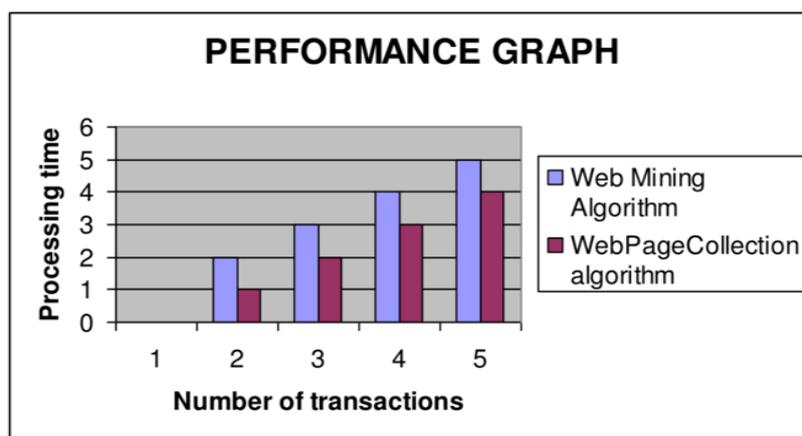


Figure 2 Performance Graph.

3. WEB SERVER LOG FILE

We need to familiarize ourselves with the kinds of information outlines available for setting apart the patron activities. Web utilization mining takes contribution via log file. A demand from patron's application to a web server, a response which is introduced from server is known as weblog record, server log or raw server log. Hence, the raw server log files consist the facts of internet website user.

Web server log file reaction will be in single coded line which is in the form of an ASCII characters.

The following is the format of standard log-file:

```
[IP Address; Date/Time; GET Request; Visited Link; status code; Bytes Transferred]
```

- **IP Address:** An Internet Protocol address is a numerical value allotted to every device associated with the computer network that uses the Internet Protocol for interaction over the internet.
- **Date/Time:** This field has been designed in such pattern of [Date: Month: Year: Hours: Minutes: Seconds] whereas in webservice log file it seems to be in this format; [DD: MM: YYYY: HH:MM: SS].

- **HTTP request:** The Hypertext Transfer Protocol is intended to empower communication between client System and servers. It's an intermediate between client and server and delivers the request-response protocol. Here client has been treated as a web browser and the application which host a website treated as a server. Therefore, this field handles the client's data which has been requested from the web server.

The GET is the most broadly asked for methodology which addresses a demand to get the data which is recognized by the URL.

- **Visited Link:** This field contains the link which was visited by a particular user from a particular IP address.
- **Status code:** It is a field which provides the response through reaction number, There is three digit in a number of status code which is generated from the server when the user performs some action on the website.

The server sends four type of status code in a response, they are:

200 series (Transmission)

200 series status code represents the request was received from the client and it is understood. Whereas 200 represents the “Success”, 201 represents the “Created”, 202 represents the “Accepted” and 204 represents the “No content”.

300 series (Redirection)

300 series status code represents the action should be taken in order to complete the request of client. Whereas, 301 represents the “Permanently Moved”, 302 represents the “Temporarily Moved”, 303 represents the “Not Modified” and 304 represents the “Cache document used”.

400 series (Client-Side error)

400 series status code represents that the client request cannot be permitted because of some technical issue. Whereas, 400 represents the “Bad request”, 401 represents the “Unauthorized”, 403 represents the “Forbidden” and 404 represents the “Not found”.

500 series (Server-side error)

500 series status code represents that the execution is has been failed by the server. Whereas, 500 represents the “Internal server error”, 501 represents the “Not implemented” and 503 represents the “Service unavailable”.

- **Bytes Transferred:** It is a field which represents the amount of volume transferred or exchanged between server and clients.

The below figure is the basic sample of server log file

```
64.242.88.10 -- [07/Mar/2004:16:05:49 -0800] "GET /wiki/bin/edit/Main/Double_bounce_sender?topicparent=Main.ConfigurationVariables HTTP/1.1" 401 12846
64.242.88.10 -- [07/Mar/2004:16:06:51 -0800] "GET /wiki/bin/rdiff/TWiki/NewUserTemplate?rev1=1.3&rev2=1.2 HTTP/1.1" 200 4523
64.242.88.10 -- [07/Mar/2004:16:10:02 -0800] "GET /mailman/listinfo/hsdivision HTTP/1.1" 200 6291
64.242.88.10 -- [07/Mar/2004:16:11:58 -0800] "GET /wiki/bin/view/TWiki/WikiSyntax HTTP/1.1" 200 7352
64.242.88.10 -- [07/Mar/2004:16:20:55 -0800] "GET /wiki/bin/view/Main/DCCAndPostFix HTTP/1.1" 200 5253
64.242.88.10 -- [07/Mar/2004:16:23:12 -0800] "GET /wiki/bin/oops/TWiki/AppendixFileSystem?template=oopsmore&param1=1.12&param2=1.12 HTTP/1.1" 200 11382
64.242.88.10 -- [07/Mar/2004:16:24:16 -0800] "GET /wiki/bin/view/Main/PeterTheony HTTP/1.1" 200 4924
64.242.88.10 -- [07/Mar/2004:16:29:16 -0800] "GET /wiki/bin/edit/Main/Header_checks?topicparent=Main.ConfigurationVariables HTTP/1.1" 401 12851
64.242.88.10 -- [07/Mar/2004:16:30:29 -0800] "GET /wiki/bin/attach/Main/OfficeLocations HTTP/1.1" 401 12851
64.242.88.10 -- [07/Mar/2004:16:31:48 -0800] "GET /wiki/bin/view/TWiki/WebTopicEditTemplate HTTP/1.1" 200 3732
64.242.88.10 -- [07/Mar/2004:16:32:50 -0800] "GET /wiki/bin/view/Main/WebChanges HTTP/1.1" 200 40520
64.242.88.10 -- [07/Mar/2004:16:33:53 -0800] "GET /wiki/bin/edit/Main/Smtpd_etrn_restrictions?topicparent=Main.ConfigurationVariables HTTP/1.1" 401 12851
64.242.88.10 -- [07/Mar/2004:16:35:19 -0800] "GET /mailman/listinfo/business HTTP/1.1" 200 6379
64.242.88.10 -- [07/Mar/2004:16:36:22 -0800] "GET /wiki/bin/rdiff/Main/WebIndex?rev1=1.2&rev2=1.1 HTTP/1.1" 200 46373
64.242.88.10 -- [07/Mar/2004:16:37:27 -0800] "GET /wiki/bin/view/TWiki/DontNotify HTTP/1.1" 200 4140
64.242.88.10 -- [07/Mar/2004:16:39:24 -0800] "GET /wiki/bin/view/Main/TokyoOffice HTTP/1.1" 200 3853
64.242.88.10 -- [07/Mar/2004:16:43:54 -0800] "GET /wiki/bin/view/Main/MikeMannix HTTP/1.1" 200 3686
64.242.88.10 -- [07/Mar/2004:16:45:56 -0800] "GET /wiki/bin/attach/Main/PostfixCommands HTTP/1.1" 401 12846
64.242.88.10 -- [07/Mar/2004:16:47:12 -0800] "GET /robots.txt HTTP/1.1" 200 68
64.242.88.10 -- [07/Mar/2004:16:47:46 -0800] "GET /wiki/bin/rdiff/Know/ReadmeFirst?rev1=1.5&rev2=1.4 HTTP/1.1" 200 5724
64.242.88.10 -- [07/Mar/2004:16:49:04 -0800] "GET /wiki/bin/view/Main/TWikiGroups?rev=1.2 HTTP/1.1" 200 5162
64.242.88.10 -- [07/Mar/2004:16:50:54 -0800] "GET /wiki/bin/rdiff/Main/ConfigurationVariables HTTP/1.1" 200 59679
64.242.88.10 -- [07/Mar/2004:16:52:35 -0800] "GET /wiki/bin/edit/Main/Flush_service_name?topicparent=Main.ConfigurationVariables HTTP/1.1" 401 12851
64.242.88.10 -- [07/Mar/2004:16:53:46 -0800] "GET /wiki/bin/rdiff/TWiki/TWikiRegistration HTTP/1.1" 200 34395
64.242.88.10 -- [07/Mar/2004:16:54:55 -0800] "GET /wiki/bin/rdiff/Main/NicholasLee HTTP/1.1" 200 7235
```

Figure 3 Web server log-file.

In this section we have pre-processed the data-set by using unix based scripting language that is SHELL. It is a Unix based programming language, designed to be accessed through terminal that is command line interpreter. By the use of this language, we have performed essential operations which are needed in data pre-processing like file manipulation.

We have used the in-built function; SORT, PIPE, CUT & UNIQ for Pre-Processing of data.

The Commands which are used for Pre-Processing:

Typed in command line interpreter.

```
cut -d',' -f1 amazon.csv | sort | uniq > ip.txt
```

In above syntax, the “cut -d ‘,’ -f1” will select first column of dataset that is IP address. And then by using pipe (|) function the execution is transferred to SORT and UNIQ function for sorting of data in appropriate manner. Lastly, we have stored the output in ip.txt file by using greater than operator (>).

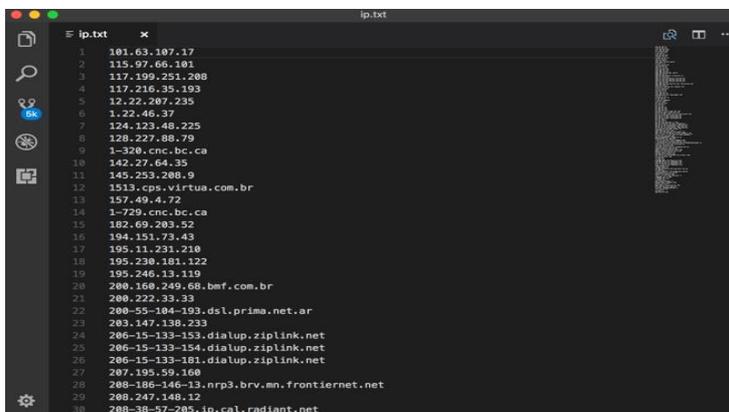


Figure 6 Various IP addresses filtered from server log file.

Typed in command line interpreter.

```
cut -d',' -f4 amazon.csv | sort | uniq > addresses.txt
```

In above syntax, the “cut -d ‘,’ -f4” will select fourth column of dataset that is Visited Links. And then by using pipe (|) function the execution is transferred to SORT and UNIQ function for sorting of data in appropriate manner. Lastly, we have stored the output in addresses.txt file by using greater than operator (>).

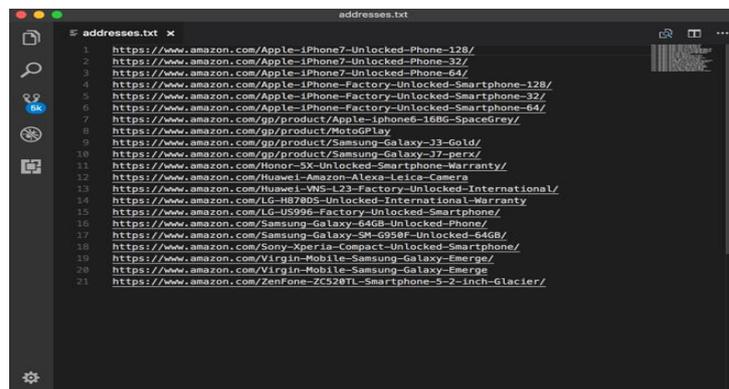


Figure 7 Various visited links filtered from server log file.

4.2. Classification Technique

To discover the anguish relationship, design & pattern between web user and website, we have used Apriori Algorithm. It is an established field of data mining calculations and algorithms to find similar Item Set. Let’s make it more understandable with an example.

Table 1 IP addresses and their visited links.

User IP Address	Visited Links
IP_Ad1	V-L1, V-L2, V-L3, V-L1, V-L4, V-L1, V-L2
IP_Ad2	V-L4, V-L2, V-L2, V-L3, V-L6, V-L1
IP_Ad3	V-L2, V-L3, V-L1, V-L1, V-L1, V-L1, V-L2
IP_Ad4	V-L1, V-L2, V-L3, V-L4, V-L4, V-L6, V-L4

We have created a table to demonstration the dataset and idea behind the experiment. Here in Table 1, all IP-Address (IP_Ad) are stored in a first column named as “User IP Address” and all Visited links (V-L) are stored in a second column of the table named as “Visited Links”.

Considering Apriori algorithm in Table 2, IP_Ad1 in User IP Address column has visited the link L1 four times, whereas IP_Ad2 IP Address has visited the link L2 twice, IP_Ad3 IP address has visited the link L1 four times and IP_Ad4 IP address has visited the link L4 thrice. By this assumption, we can predict most visited and favorable link of a user IP address.

Table 2 Most Visited Link by Individual IP address

User IP Address	Visited Links	Most Visited Link
IP_Ad1	V-L1, V-L2, V-L3, V-L1, V-L4, V-L1, V-L2	V-L1
IP_Ad2	V-L4, V-L2, V-L2, V-L3, V-L6, V-L1	V-L2
IP_Ad3	V-L2, V-L3, V-L1, V-L1, V-L1, V-L1, V-L2	V-L1
IP_Ad4	V-L1, V-L2, V-L3, V-L4, V-L4, V-L6, V-L4	V-L4

Now, in Table 3 considering our implementation and algorithm. we have added new column in a previous table named as “Count of visited link”, Here this column contains the number of count of a link which is most visited by a user IP Address. So, we have done the prediction on the basis of highest number of count variable. It will help e-business administrator to retain e-customers by visualizing them an advertisement independently.

Table 3 Count of Visited Links by an IP address

User IP Address	Most Visited link	Count of Visited Link
IP_Ad1	V-L1	4
IP_Ad2	V-L2	2
IP_Ad3	V-L1	4
IP_Ad4	V-L4	3

4.3. Implementation of Experiment and Results

This section explains, the implementation and experiment which is done with the help of python programming language. As in figure 8, We have used sorted server log file as an input one, Sorted and Filtered IP address text file as input two which was generated at the of pre-processing of data-set and non-repetitive visited links text file as input three, which is also generated at the time of pre-processing of data-set that is web server log file. When we execute this algorithm, it will take one IP address one by one from Ip.txt text file and it will compare all visited links from each other of a particular IP address, apart from that it will also verify the most visited links from addresses.txt text file and during comparison if any visited links found to be same, then visited link count variable will be increment by one of a particular user.

```

1 import time
2 def main():
3     f = open("addresses.txt", "r")
4     f1 = open("ip.txt", "r")
5     f2 = open("amazon.csv", "r")
6     addresses = f.readlines()
7     ip = f1.readlines()
8     visits = f2.readlines()
9     links_array = []
10    ip_array = []
11    max_visit = []
12
13
14    for line in addresses:
15        links_array.append(line.partition(',') [0])
16
17    for line in ip:
18        ip_array.append(line.partition(',') [0])
19
20    for i in range(0, len(ip_array)):
21        max_count = 0
22        max_j = 0
23        for j in range(0, len(links_array)):
24            count = 0
25            for lines in visits:
26                ip, link = lines.split('\t')
27                if ip.strip() == ip_array[i].strip() and link.strip() == links_array[j].strip():
28                    count += 1
29            if count == max_count:
30                max_count = count
31                max_j = j
32
33

```

Figure 8 Code Snippet

The figure 9, shows the result of our proposed algorithm which was executed in sequence manner and the result was stored and written in linkcount.csv file. The output record shows the link which is visited by an IP address with the maximum number of count and via this, we have anticipated that most favorable link of a person; that is the one who has the maximum number of count variable.

IP Address	Link	Count
101.63.107.17	https://www.amazon.com/ZenFone-ZC520TL-Smartphone-5-2-inch-Glacier/	6
115.97.66.101	https://www.amazon.com/Apple-iPhone-Factory-Unlocked-Smartphone-128/	4
117.199.251.208	https://www.amazon.com/Apple-iPhone-Factory-Unlocked-Smartphone-128/	2
117.216.35.193	https://www.amazon.com/ZenFone-ZC520TL-Smartphone-5-2-inch-Glacier/	28
12.22.207.235	https://www.amazon.com/Apple-iPhone7-Unlocked-Phone-32/	1
1.22.46.37	https://www.amazon.com/ZenFone-ZC520TL-Smartphone-5-2-inch-Glacier/	8
124.123.48.225	https://www.amazon.com/Apple-iPhone-Factory-Unlocked-Smartphone-128/	3
128.227.88.79	https://www.amazon.com/Virgin-Mobile-Samsung-Galaxy-Emerge	2
1-320.cnc.bc.ca	https://www.amazon.com/ZenFone-ZC520TL-Smartphone-5-2-inch-Glacier/	1

Figure 9 Result.

The figure 10 shows the snippet of the code, here we've got used time function in our code to generate and discover the time complexity of our algorithm. As in [5], the authors have achieved a comparison of the diverse algorithms of web utilization mining on basis of time complexity. So in figure 10, our calculation is effective in time complexity because it takes 1.94 seconds to reveal in the informational index of length 1600 records, whereas time

complexity of authors algorithm took most excessive 5 seconds and least 2 seconds to finish execution.

```
start_time = time.time()
if __name__ == "__main__":
    main()
    print("--- %s seconds ---" % (time.time() - start_time))
```

Figure 10 Time function ().

```
-----
[Satyams-MacBook-Pro:Impl satyamramawat$ python count_links.py
--- 1.94122290611 seconds ---
```

Figure 11 Complexity in term of time.

5. FUTURE IMPLEMENTATION

Firstly, the dataset which consists of server log file is organized. Hence, it can be actualized with the field of machine learning. As we are aware that machine learning is the next big thing in future of software technology. Hence it will provide greater adaptability if website is based on artificial intelligence specially e-commerce business websites. Our experiment should be actualized with machine learning concept based classification strategy than the outcome will be in more productive in terms of time complexity. By our calculation and algorithms, we can also make web application module based on retargeting internet business client. Secondly, there is one more field of the problem which should be actualized; how we can recognize the user interest of particular user from different MAC address. Our work is concentrating on favorable ads on basis of IP address but not on MAC address. Hence, there is a need to find the user interest of the common user based on MAC address. Thus this research work can be extended for finding user interest based on MAC address.

6. CONCLUSIONS

In this article, we have produced the semantic system for understanding the user behavior on the web; Additionally, we have attempted to provide the learning to retarget the online business client by utilizing web mining strategy. So we come up with a solution with an algorithm which is designed and tested by us. The algorithm is designed on Apriori algorithm, focused to target the most visited product of the consumer; By actualizing the IP address on basis of most visited links with the highest number of a count. Therefore, our algorithm is demonstrated highly productive from the other algorithms.

REFERENCES

- [1] Furrer and Sudharshan, (2001) "Internet marketing research: opportunities and problems", Vol. 4 Issue: 3, pp.123-129.
- [2] Ansari, Suhail and Kohavi, Ron and Mason, Llew and Zheng, Zijian, Integrating E-Commerce and Data Mining: Architecture and Challenges, ISBN 0-7695-1119-8, PAGES 27—34.
- [3] Siddiqui and Aljahdali, Web Mining Techniques in E-Commerce Applications, ISSN: 0975-8887, Volume 69– No.8.
- [4] Bhisikara and Sahu, Overview on Web Mining and Different Technique for Web Personalization, ISSN: 2248-9622.

- [5] Shanthi and Rajagopalan, IJIRCCE, Vol. 1, Issue-7, ISSN :2320-9801. An Efficient Web Mining Algorithm to Mine Web Log Information.
- [6] Sandro Araya, Mariano Silva, Richard Weber, A methodology for web usage mining and its application to target group identification, Fuzzy Sets and Systems, Volume 148, Issue 1, 2004, Pages 139-152, ISSN 0165-0114.
- [7] Orit Raphaeli, Anat Goldstein, Lior Fink, Analyzing online consumer behavior in mobile and PC devices: A novel web usage mining approach, Electronic Commerce Research and Applications, Volume 26,2017,Pages 1-12,ISSN 1567-4223.
- [8] [8] G. Shivaprasad, N.V. Subba Reddy, U. Dinesh Acharya, Prakash K. Aithal, Neuro-Fuzzy Based Hybrid Model for Web Usage Mining, Procedia Computer Science, Volume 54, 2015, Pages 327-334, ISSN 1877-0509.
- [9] C.J. Carmona, S. Ramírez-Gallego, F. Torres, E. Bernal, M.J. del Jesus, S. García, Web usage mining to improve the design of an e-commerce website: OrOliveSur.com, Expert Systems with Applications, Volume 39, Issue 12, 2012, Pages 11243-11249, ISSN 0957-4174.
- [10] B. Naveena Devi, Y. Rama Devi, B. Padmaja Rani, R. Rajeshwar Rao, Design and Implementation of Web Usage Mining Intelligent System in the Field of e-commerce, Procedia Engineering, Volume 30, 2012, Pages 20-27, ISSN 1877-7058.
- [11] Prajyoti Lopes, Bidisha Roy, Dynamic Recommendation System Using Web Usage Mining for E-commerce Users, Procedia Computer Science, Volume 45, 2015, Pages 60-69, ISSN 1877-0509.
- [12] Stephen G. Matthews, Mario A. Gongora, Adrian A. Hopgood, Samad Ahmadi, Web usage mining with evolutionary extraction of temporal fuzzy association rules, Knowledge-Based Systems, Volume 54, 2013, Pages 66-72, ISSN 0950-7051.
- [13] Yu-Hui Tao, Tzung-Pei Hong, Yu-Ming Su, Web usage mining with intentional browsing data, Expert Systems with Applications, Volume 34, Issue 3, 2008, Pages 1893-1904, ISSN 0957-4174.
- [14] D.A. Adeniyi, Z. Wei, Y. Yongquan, Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method, Applied Computing and Informatics, Volume 12, Issue 1, 2016, Pages 90-108, ISSN 2210-8327.
- [15] Tzung-Pei Hong, Cheng-Ming Huang, Shi-Jinn Horng, Linguistic object-oriented web-usage mining, International Journal of Approximate Reasoning, Volume 48, Issue 1, 2008, Pages 47-61, ISSN 0888-613X.
- [16] Yoon Ho Cho, Jae Kyeong Kim, Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce, Expert Systems with Applications, Volume 26, Issue 2, 2004, Pages 233-246, ISSN 0957-4174.
- [17] A. Sangeetha and Dr. C. Nalini , A Novel Approach: Various Classifications in Web Mining using Diagnosis of Faults in Electro - Mechanical Devices from Vibration Measurements, International Journal of Mechanical Engineering and Technology 8(10), 2017, p p. 65 – 70
- [18] Prof. Sindhu P Menon and Dr. Nagaratna P Hegde, Research On Classification Algorithms And Its Impact On Web Mining. International Journal of Computer Engineering and Technology, 4 (4), 2013, pp. 495 - 504.