Web Log Analysis Using Association Rule Mining Algorithms

Ms.E.Suganya
Ph.D Research Scholar
Department of Computer Science
Bharathiar University
Coimbatore

Dr.S.Vijayarani
Assistant Professor
Department of Computer Science
Bharathiar University
Coimbatore

Ms. V.Eswary
Student
Department of Computer Science
Bharathiar University
Coimbatore

Abstract: The Web has recently become a powerful source for retrieval of information and discovering knowledge from web data. Web Usage Mining is one of the applications of data mining techniques to depict the knowledge out of web log data. Web Usage Mining consists as the process of data cleaning, data preprocessing, and pattern discovery and pattern analysis. The main aim of this research work is to analyze the user's web log and to find which website is seen by the users at a longer period. Session information and association rule mining algorithms like Apriori, FP-Growth and Improved FP-Growth are used for web log analysis. The synthetic dataset is generated which includes 232 instances and 3 attributes, namely IP address, URL and time taken. From the observation based on the performance analysis, it is found that the improved FP Growth algorithm gives better results while comparing other algorithms.

Keywords: Web Usage Mining, Web Log Analysis, Association Rule Mining, Apriori, FP-Growth

1. INTRODUCTION

Web Mining is one of the areas of data mining which can handle an unstructured or semi structured data. It can be defined as the discovery and analysis of useful information from World Wide Web. Based on the significance, web mining is classified into three different kinds; web content mining, web structure mining and web usage mining. Web content mining is the process of extracting useful information from the web documents which have the collection of data such as text, image, audio, video, HTML pages and animation. Web structure mining is the study of data interconnected to the structure of a particular website. It consists of web graph which contains the web pages or web documents as nodes and hyperlinks as those edges are connected between two related pages. Web structure is a useful source for extracting information. Web usage mining is also called as web log mining, which is used to analyze the behavior of online users. Web usage mining is to extract the data which are stored in server access logs, referrer logs, agent logs and error logs. Web usage mining generally uses basic data mining algorithms such as association rule mining, sequential rule mining, clustering, and classification. It has several tools to analyze the behavior of the user.

Association rule mining is the process of finding the association rules that satisfy the predefined minimum support and confidence from a given database [3]. Association rule mining finds an interesting relations and connections along with a large set of data items. Association rules show attribute value conditions that occur frequently together in a given dataset with help of candidate generation. A typical and widely-used example of an association rule mining is Market Basket Analysis. For example, data are collected using bar-code scanners in a shopping mall. Such 'market basket' databases consist of a large number of transaction records. Each record lists all items bought by a customer on a single purchase transaction. Managers would be interested to know if certain groups of items are consistently purchased together. They could use this data for adjusting store layouts (placing items optimally with respect to each other), for cross-selling, for promotions, for catalog design and to identify customer segments based on buying patterns. Based on the types of values, the association rules can be classified into two categories:

1. Boolean Association Rules
2. Quantitative Association Rules

1. **Boolean Association Rules -** Association rule mining is very important research topic in the field of data mining. Boolean vector "relational calculus" method is used to discover frequent itemsets.
2. **Quantitative Association Rules** A rule template is a preset format of a quantitative association rule and it is used as a starting point in the quantitative mining process. It is defined by the set of attributes occurring in the left hand side and the right hand side or both sides of the rule.

1.1 Rule basic Measures

There are two basic measures are used in association rule mining. These are

- **Support:** Denotes the frequency of the rule within transactions. A high value means that the rule involves a great part of the database.
- **Confidence:** Denotes the percentage of transactions containing item A also contain item B. It is an estimation of conditioned probability.

- **Itemset:** A set of items is referred to as itemset. An itemset containing k items is called k itemset. An itemset can also be seen as a conjunction of items (or a predicate)
- **Frequent Itemsets:** Suppose min_sup is the minimum support threshold. An itemset satisfies minimum support means the occurrence of the itemset is greater than or equal to min_sup. If an itemset satisfies minimum support, then it is a frequent itemset.
- **Strong Rules:** Rules that satisfy both minimum support threshold and minimum confidence threshold are called strong rules. Two important steps of Association Rule Mining is to find all the frequent itemsets and to generate strong association rules from the frequent itemsets.
- **Itemset Generation:** A frequent itemset is an itemset whose support is greater than some user-specified minimum support (denoted L_k, where k is the size of the itemset). A candidate itemset is potentially a frequent itemset (denoted C_k, where k is the size of the itemset).

1.2 Advanced Techniques

- Multiple-Level Association Rules
- Multi -Dimensional Association Rules
- Quantitative Association Rules
- Mining Sequential Patterns
- Sequential Pattern Mining

2. RELATED WORKS

Author	Title	Techniques	Inference
Pooja Sharma, Rupali Bhartiya	An Efficient Algorithm for Improved Web Usage Mining	Apriori Algorithm	The authors proposed the improved Apriori algorithm which deals with a small number of candidate sets. The authors observed that the proposed algorithm has produced good results.
Pamnani.R & Chawan.P	Web Usage Mining: A Research Area In Web Mining	Pattern Analysis Techniques	The authors proposed a novel integrated technique for analyzing the user behavior in digital learning in an education environment. The proposed technique is initiated with pattern analysis techniques followed by web usage analysis with the final implementation of the resource analysis algorithm for an E-learning system.
Mr. Rahul Mishra, Ms. AbhaChoubey	Discovery of Frequent Patterns from Web Log Data by using FP-Growth algorithm for Web Usage Mining	Web Data Preprocessing Techniques	The steps used by the authors are cleaning the web data, user identification and session identification. User identification is done by considering the IP address and agent used. The author described that missing references are present, which are removed using path completion method.
Hongzhou Sha, Tingwen Liub, Peng Qin, Yong Sun and Qingyun Liu	EPLogCleaner: Improving Data Quality of Enterprise Proxy Logs for Efficient Web Usage Mining	EPLogCleaner	They made an evaluation of EPLogCleaner with a real network traffic trace captured from one enterprise proxy. Experimental results showed that EPLogCleaner improved the data quality of enterprise proxy logs by further filtering out more than 30% URL requests comparing with traditional data cleaning methods.
Rakesh Agrawal, Ramakrishnan Srikant	Fast Algorithms for Mining Association Rules	AIS, SETM, Apriori and AprioriTid	The authors compared the algorithms to the previously known algorithms, the AIS and SETM algorithms. The experimental results proved that the proposed algorithms always outperform AIS and SETM. The performance gap increased with the problem size, and ranged from a factor of three for small problems to more than an order of magnitude

			for large problems.
Mahajan, R., Sodhi, J., & Mahajan	Usage Pattern Discovery From A Web Log in an Indian E-Learning Site	Frequent Access patterns	The authors described the research analysis of web log data shows how the e-content was accessed. Topic wise frequent access patterns indicate what topics in Biology, by learners of a particular grade.
Shipra Khare, Vivek Jain, Manoj Ramaiya	Implementation of Web Usage Mining with Customized Web Log Using FP Growth Algorithms	FP Growth and Apriori algorithm	The authors proposed and analyzed the two association rule based algorithms, namely FP Growth and Apriori algorithm, which meet the needs of various web service providers and various viewers, users, business analysts, etc.

3. METHODOLOGY

This research work is mainly concentrated on web log analysis. In web usage mining, the data preprocessing techniques are data cleaning, user identification, session identification and path completion are available to process the data. The main objective of this research work is to identify the session information from student web log file which is collected and analyzed using association rule mining algorithms. Two existing algorithms are compared such as Apriori and FP-Growth and one algorithm are proposed namely improved FP-Growth algorithm. Figure 3.1 shows the overall system architecture of the proposed system.

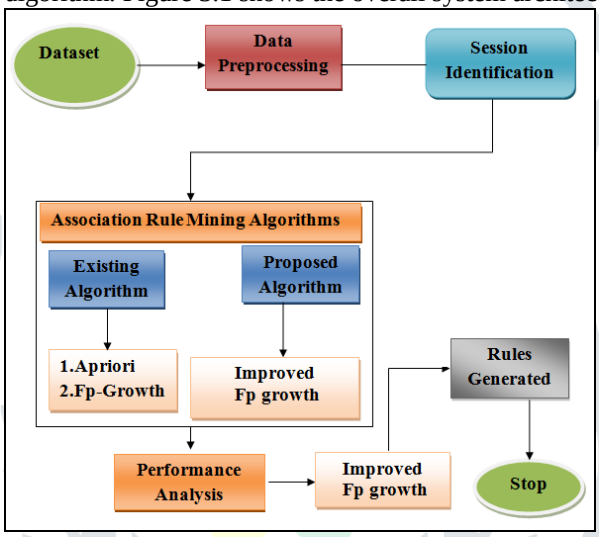

Figure 3.1 System Architecture

3.1 ACCESS LOG FILE

An access log is a list of all the requests for individual files that people have requested from a Web site. These files will include the HTML files and their imbedded graphic images and any other associated files that get transmitted.

3.2 DATA PREPROCESSING

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issue. There are four stages: Data Cleaning, User Identification, Session Identification and Path Completion.

3.2.1 Data Cleaning

The web log file is the source of data to the process of data cleaning. Data cleaning is often web sites specific. The data cleaning purpose is to remove an irrelevant record or items from the log file. The data cleaning task is to delete unrelated data to mining namely images in the format of JPEG, jpg and GIF, error response, JavaScript, Cascading Style Sheets and robot request. Data cleaning also involves the removals of references due to crawler information navigation.

3.2.2 User Identification

A user identification is performed after data cleaning. The website visited by users is identified by the process of user identification. This is performed with the support of user agent and IP address. The web usage analysis does not require knowing about the identity of the user. However, it is essential to distinguish among different users. Since a user may visit a site more than once, the server logs register numerous sessions for every user. The user activity record phrase is used to identify the logged activities sequence of similar user. In the absence of authentication mechanisms, the most widespread approach to differentiate the distinct visitors is done by using client side cookies. It is not possible to use cookies for entire websites due to the concerns of

privacy where the client side cookies are disabled by the users. Without client side cookies or user authentication is not possible to recognize distinct users accurately. Another approach can be the use of IP addresses for user identification. Users with different IP addresses can be treated as unique but it is not the perfect solution as ISPs assign rotating addresses to users.

3.2.3 Session Identification

The session is a time period between user's login and logout process. The user visits several pages during this time. The session is used to predict the page sequences and trace the activity of the user. A user session can be defined as the set of pages visited by a similar user within particular duration of one specific visit to a website. A user may have multiple or single sessions during the time period. Once the user has been recognized, then every user click stream is divided into the logical clusters. The portioning method in the sessions is known as session reconstruction or sessionization. Session reconstruction can be categorized into two main approaches and they are navigation oriented approach and time oriented approach. The methods used for user identification to certain extent, can be used for session identification. To group the activities of a single user from the web log files is called a session. As long as the user is connected to the website, it is called the session of that particular user. Most of the time, 30 minutes time-out was taken as a default session time-out. A session is a set of page references from one source site during one logical period. Historically a session would be identified by a user logging into a computer, performing work and then logging off. The login and logoff represent the logical start and end of the session. Following rules to identify user sessions.

1. If there is a new user there is a new session.

2. In one user session, if the referrer page is null, there is a new session.

3. If the time between page requests exceeds a certain limit (2 minutes). It is assumed that user is starting a new session.

The reference page is estimated by access time of this page and the next one i.e. the reference length of an accessed page equals the difference between the access time of the next and the present page. If this time is few seconds than that page can be considered as an auxiliary page and otherwise that page can be considered as a content page.

3.2.4 Path Completion

The purpose of path completion is to identify user's travel pattern and also the missing pages in path where the user access must to be appended. It is possible to recognize several missed pages by cached versions and proxy servers of pages used by client. So the step of path completion is undertaken to recognize missing pages. The path set is incomplete accessed pages in a session of user and it is retrieved from each set of user session.

3.2 Association Rule Mining

Association rules are usually required to satisfy a user-specified minimum support and a user-specified minimum confidence at the same time. Association rule generation is usually split up into two separate steps:

- First, minimum support is applied to find all frequent itemsets in a database.
- Second, these frequent itemsets and the minimum confidence constraint are used to form rules.

While the second step is straightforward, the first step needs more attention.

Finding all frequent itemsets in a database is difficult since it involves searching all possible itemsets (item combinations). The set of possible itemsets is the power set over I and has size $2^n - 1$ (excluding the empty set which is not a valid itemset). Although the size of the power set grows exponentially in the number of items n in I, efficient search is possible using the downward-closure property of support (also called *anti*-monotonicity [2] which guarantees that for a frequent itemset, all its subsets are also frequent and thus for an infrequent itemset, all its supersets must also be infrequent. Exploiting this property, efficient algorithms (e.g., Ankit R Kharwar, Viral Kapadia, Nilesh Prajapati, Premal Patel, "Implementing APRIORI Algorithm on Web server log", [5] can find all frequent itemsets.

Framework: Find all the rules that satisfy both a minimum support (min_sup) and a minimum confidence (min_conf) threshold.

• Association rule mining:

– Find all frequent patterns (with support ≥ min_sup).

– Generate strong rules from the frequent patterns.

• The second step is straightforward:

– For each frequent pattern p, generate all nonempty subsets.

– For every non-empty subset s, output the rule s ⇒(B-C) if conf=sup(B)/sup(C) ≥ min_conf.

• The first step is much more difficult. Hence, focus on frequent pattern mining.

3.1.1 Apriori Algorithm

Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data, in order to understand and better serve the needs of Web-based applications. Usage data capture the identity or origin of Web users along with their browsing behavior on a Web site. Web server data correspond to the user logs that are collected in Web server. In

order to produce the usage patterns and user behaviors, this paper implements the high level process of Web Usage Mining using basic Association Rules algorithm call Apriori Algorithm. Web Usage Mining consists of three main phases, namely Data Pre-processing, Pattern Discovering and Pattern Analysis. Server log files become a set of raw data where it's must go through with all the Web Usage Mining phases to producing the final results. Here, Web Usage Mining, approach has been combined with the basic Association Rules, Apriori Algorithm to optimize the content of the serve log data. Finally, this paper will present a finding association Rule from server log which are useful in many applications like a cache for web page, Marketing, Targeted Advertising etc.

3.1.2 FP-Growth Algorithm

Tree FP-Growth frequent pattern mining is used in the development of association rule mining. FP-Tree algorithm overcomes the problem found in Apriori algorithm. By avoiding the candidate generation process and fewer passes over the database, FP-Tree was found to be faster than the Apriori algorithm [9]. It adopts a divide and conquer strategy. Firstly, it compresses the database representing frequent items into a frequent –pattern tree or FP-tree. It retains the item set association information and compressed databases are divided into a set of conditional databases, each one associated with a frequent item. It takes the help of prefix tree representation of the given database of transactions (called FP tree), which saves a considerable amount of memory for storing the transactions. An FP-Tree is a prefix tree for transactions. Every node in the tree represents one item and each path represents the set of transactions that involve with the particular item. All nodes referring to the same item are linked together in a list, so that all the transactions that containing the same item can be easily found and counted. Large databases are compressed into a compact FP tree structure. FP tree structure stores necessary information about frequent item sets in a database[14].

Improved FP-Growth Algorithm

The proposed methodology also extracts the database information from the given set of database based on the given problem. The effective association rule mining is done by making the relationship between the quantitative FP- growth set of data derived from the earlier conversion.

FP-Growth is generating the many rules so we have used to the quantitative FP-Growth. The quantitative FP-Growth is used to the lexicography techniques. This techniques following steps and preprocessing as like that. Now the values of three attributes transaction item, transaction id and time taken. So here taken time we consider 2 mins below reduced from the dataset, 2 mins above we considered generating rule and comparing the FP-Growth and quantitative FP-Growth.

Improved FP-Growth Algorithm

Input: A dataset composed of N, MinConf, and the number of attributes

Output: Quantitative FP-Growth algorithm

Step 1: Select a set of attributes from the given database R

Step 2: Let Rt a set of constraints defined on these attributes.

Step 3: Generate concept rule of FP-Growth for the attributes

Step 4: For each record in the database substitute the numerical value

Step 5: Call Procedure quantitative _relation (database)

Step 6: For each tuple in N find the relationship using the combined number for the datasets

Step 7: Generate the rules from the transformed dataset.

Step 8: Filter the rules that satisfy the min-conf.

The IFP algorithm finds the most appropriate item in the transaction database system.

Step 1: shows the transaction database in its original form.

Step 2: The frequencies of individual items are determined from this input in order to be able to store infrequent items into the hash table immediately. Assume a maximum support of three transactions for the example, there are no frequent items, so all items are kept.

Step 3: The (frequent and infrequent) items in each transaction are sorted according to their frequency in the transaction database, since it is well known that processing the items in the order of increasing frequency usually leads to the shortest execution times. For frequency the algorithm reads the specific index and for infrequent it reads negative index.

Step 4: The transactions are sorted lexicographically into ascending order initially for frequent items and descending for infrequent items, with item comparisons again being decided by the item frequencies, although here the item with the higher frequency precedes the item with the higher frequency.

Step 5: The data structure on which IFP operates is built by combining equal transactions and setting up an array, in which each element consists of two fields: an occurrence counter and a pointer to the sorted transaction (array of contained items).

4. RESULTS AND DISCUSSIONS

4.1 DATASET DESCRIPTION

In order to perform the experiments, the real time dataset is collected from different web pages. The main objective of this research is to automatically extract the information from the various websites in the educational Institute. In this research paper find the association rule using comparing two algorithms for session identification. We have taken real time dataset. Hence, here two fields are considered; i,e. IP address considered as transaction id, URL is considered as transaction items and time taken (minutes) is considered as time. This dataset has the client IP address, URL (location), time taken (Mintues). From the observation, based on the performance analysis FP-growth algorithm gives the best results while comparing with Apriori algorithm. Table 4.1 shows the details about three different data sets in three different formats.

Table 4.1 Dataset

IP Address	URL	Transaction Item (URL)
1 to 110	https://mail.google.com/mail	A
	http://www.b-u.ac.in/	B
	http://www.ugcnetonline.in/	C
	http://tnpscexams.net/	D
	http://www.conferencealerts.com/	E
	https://mail.yahoo.com/mail	F
	https://www.tutorialspoint.com/r/	G

Hence, we transformed the dataset as transaction id and transaction item and execution time i.e. IP address is considered as transaction id and URL is considered as transaction items and execution time is considered as time taken(seconds). Here the same user i.e IP address visits one or more website so 232 instances are transformed into 110 instances. For example, IP address 1 visits A, B, C URL. From the observation, based on the performance analysis improved FP Growth algorithm gives better results while comparing other algorithms.

4.3 Association Rule Mining Algorithms

4.3.1 Existing Algorithms

Table 4.3 and Figure 4.1 shows the performance analysis of Apriori and FP-Growth algorithms. Based on this observation FP-Growth algorithm generates best rule for identifying the session information from the web log file.

Table 4.2 Best Rules

Algorithm	Rules Generated
Apriori	50
FP-Growth	**223**

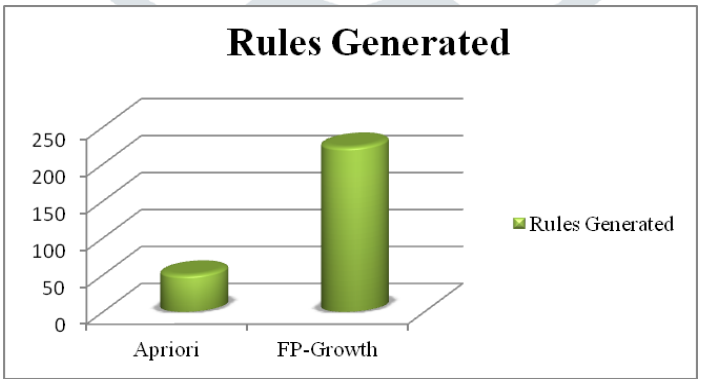

Figure 4.1 Best Rules Generated

Table 4.4 and figure 4.2 shows the execution time for performing association rule mining algorithms such as Apriori and FP-Growth. Here, we analyzed that FP-Growth algorithm takes minimum execution time while comparing Apriori algorithm.

Table 4.3 Execution Time

Algorithm	Time (Milli Secs)
Apriori	2000

FP-Growth	**1000**

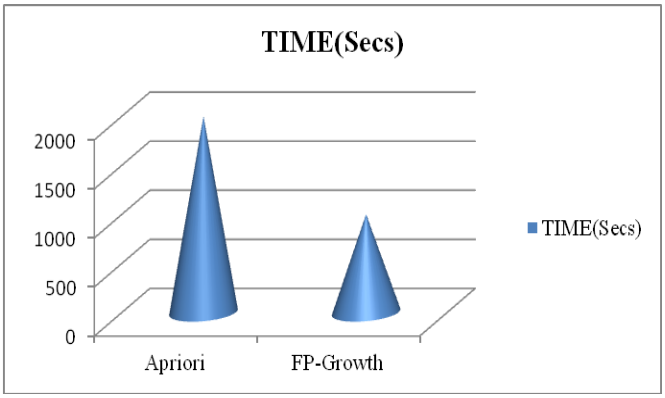

Figure 4.2 Execution Time

4.3.2 Improved FP-Growth Algorithm

Table 4.5 and Figure 4.3 shows the performance analysis of Apriori and FP-Growth and Improved FP-Growth algorithms. Based on this observation Improved FP-Growth algorithms generate best rule for identifying the session information from the web log file.

Table 4.4 Best Rule

Algorithm	Rules Generated
FP-Growth	**223**
Improved FP-Growth	**113**

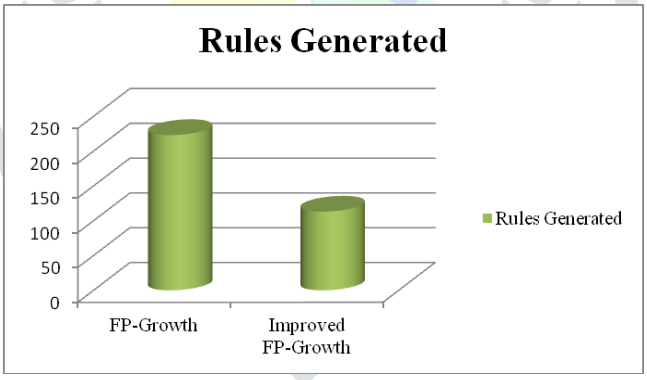

Figure 4.3 Best Rule Generated

Table 4.6 and figure 4.4 shows the execution time for performing association rule mining algorithms such as FP-Growth and Improved FP-Growth . Here, we analyzed that Improved FP-Growth algorithm takes minimum execution time while comparing FP-Growth algorithm.

Table 4.5 Execution Time

Algorithms	Time (secs)
FP-Growth	**1000**
Improved FP-Growth	**300**

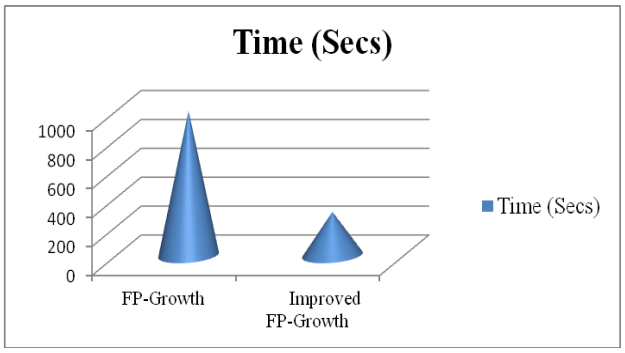

Figure 4.4 Execution Time

In this research paper work is access web log file from this we identified that the session information i.e most of the users visit a UGC website which considered as transaction item C from the web log file. Table 4.7 and figure 4.5 shows the most of users visit UGC website considered as transaction item C from the web log file.

Table 4.6 Visited Websites

Items	URL	Hits
A	https://mail.google.com/mail	30
B	http://www.b-u.ac.in/	28
C	**http://www.ugcnetonline.in/**	**80**
D	http://tnpscexams.net/	22
E	http://www.conferencealerts.com/	31
F	https://mail.yahoo.com/mail	20
G	https://www.tutorialspoint.com/r/	21

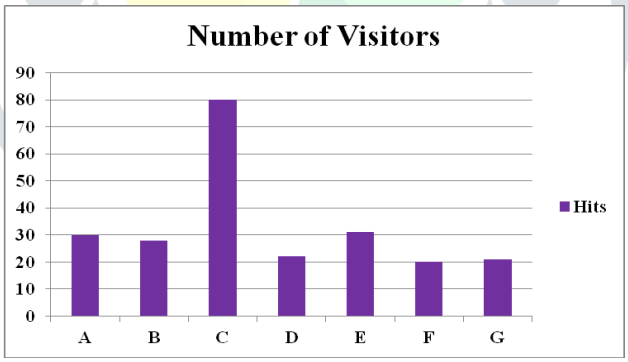

Figure 4.5 Visited Websites

5. CONCLUSION

The main objective of this research work is to find session identification using the association rule mining algorithms. The real time dataset set is collected from Bharathiar university computer science department client system's log files. It contains three months of information which include 232 instances and 3 attributes, namely IP address, URL and time taken. Hence, here IP address considered as transaction id and URL considered as transaction items. To identify the session information from web log file using association rule mining algorithms i.e. two existing algorithms such as Apriori and FP-Growth algorithm and one proposed algorithm namely improved FP-Growth algorithm. In this research work, two association rule mining algorithms are compared and proposed improved FP-Growth algorithm. Experimental results shows that the proposed algorithm performs well than other algorithms. The proposed system finds the frequently used website based on its usage.

REFERENCES

[1] Pooja Sharma, Rupali Bhartiya, "An Efficient Algorithm for Improved Web Usage Mining", International Journal of Computer Technology &Applications, Vol. 3 (2), 766-769, ISSN:2229-6093

[2] Rakesh Agrawal, Ramakrishnan Srikant, "Fast Algorithms for Mining Association Rules", Proceedings of the 20th VLDB Conference, Santiago, Chile, 1994

[3] Sandeep Singh Rawat, Lakshmi Rajamani, "Discovering Potential User Browsing Behaviors Using Custom-Built Apriori Algorithm", International Journal of Computer Science & Information Technology (IJCSIT), Vol.2, No.4, August 2010
[4] Goswami D.N., Chaturvedi Anshu, Raghuvanshi C.S., "An Algorithm for Frequent Pattern Mining Based On Apriori", (IJCSE) International Journal on Computer Science and Engineering, Vol. 02, No. 04, 2010, 942-947
[5] Ankit R Kharwar, Viral Kapadia, Nilesh Prajapati, Premal Patel, "Implementing APRIORI Algorithm on Web server log", National Conference on Recent Trends in Engineering & Technology, 13-14 May 2011
[6] Harendra Singh, Ashish Kumar Srivastava, Sitendra Tamrakar, "A Modified FP-Tree Algorithm for Generating Frequent Access Patterns", JECET, June – August-2013; Vol.2.No.3, 730-740, E-ISSN: 2278–179X
[7] Shipra Khare, Prof Vivek Jain, Prof Manoj Ramaiya, "Implementation of Web Usage Mining with Customized Web Log Using FP Growth Algorithms", International Journal of Engineering & Managerial Innovations (IJEMI), ISSN: 2321-693X,Volume I (II), September 2013
[8] Ankit R Kharwar, Viral Kapadia, NileshPrajapati, Premal Patel, "Implementing APRIORI Algorithm on Web server log", National Conference on Recent Trends in Engineering & Technology, 2011.
[9] Suneetha K R, Krishnamoorti R, "Web Log Mining using Improved Version of Apriori Algorithm", International Journal of Computer Applications, Volume 29 (6), September 2011,pp. 23-27.
[10] R. Suguna, D. Sharmila, "An Overview of Web Usage Mining", International Journal of Computer Applications, Volume 39 (13), February 2012.
[11] Sudhamathy and JothiVenkateswaran, "Hierarchical frequent pattern analysis of web logs for efficient interestingness prediction", Indian Journal of Education Info Management, Vol. 1 (5), May 2012, pp. 29-35.
[12] Mr. Rahul Mishra, Ms. AbhaChoubey, "Discovery of Frequent Patterns from Web Log Data by using FP-Growth algorithm for Web Usage Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2(9), September 2012.
[13] T. Sunil Kumar, Dr. K. Suvarchala, "A Study: Web Data Mining Challenges and Application for Information Extraction", IOSR Journal of Computer Engineering (IOSRJCE), Vol. 7(3) ,2012, pp 24-29.
[14] RakeshAgrawal, RamakrishnanSrikant, "Fast Algorithms for Mining Association Rules", Proceedings of the 20th VLDB Conference, Santiago, Chile, 1994
[15] I-Hsien Ting, "Web Mining Techniques for On-line Social Networks Analysis", IEEE 2008. Pp. 696-700.
[16] Anupam Joshi, Tim Finin, Akshay Java, Anubhav Kale, and PranamKolari, "Web (2.0) Mining: Analyzing Social Media", IEEE 2008.
[17] K. R. Suneetha, Dr. R. Krishnamoorthi, "Identifying User Behavior by Analyzing Web Server Access Log File", IJCSNS International Journal of Computer Science and Network Security. VOL.9 No.4, April 2009
[18] Mehdi Heydari, Raed Ali Helal, Khairil Imran Ghauth, "A Graph-Based Web Usage Mining Method Considering Client Side Data", 2009 International Conference on Electrical Engineering and Informatics, 2009.
[19] HuipingPeng, "Discovery of Interesting Association Rules Based on Web Usage Mining", 2010 - International Conference on Multimedia Communications.
[20] Maja Dimitrijevic, TanjaKrunic, "Association rules for improving website effectiveness: case analysis", Online Journal of Applied Knowledge Management Volume 1, Issue 2, 2013
[21] Kirti S. Patil, Sandip S. Patil, "Sequential Pattern Mining Using Apriori Algorithm & Frequent Pattern Tree Algorithm", IOSR Journal of Engineering (IOSRJEN) Vol. 3, Issue 1 (Jan. 2013), Pp. 26-30
[22] A. Smeaton and T. Ianeva, "TRECVID-2006 Search Task," TREC Video Retrieval Evaluation Online Proc., 2006. http://www.computer.org/csdl/trans/tk/2010/08/ttk2010081191-abs.html
[23] Aviral Nigam," Web Crawling Algorithms", International Journal of Computer Science and Artificial Intelligence, Sept. 2014, Vol. 4 Issue. 3, PP. 63-67
[24] Ameesha Reddy, B. Sivaiah, Rajani Badi, Venkateshwararaju Ramaraju, "Feature Based Fusion Approach for Video Search", International Journal of Advances in Engineering & Technology, Sept 2012. ©IJAET ISSN: 2231-1963
[25] Personalization: Collaborative Filtering vs Prediction Based on Benefit Theory, November 05,2007, http://myshoppal.typepad.com/blog/2007/11/personalization.html
[26] Applying Web Usage Mining to a University Website Access Domain, International Journal of Applied Information Systems (IJAIS) – ISSN: 2249-0868, Foundation of Computer Science FCS, New York, USA
[27] New Approaches to Web Personalization, Ph.D. Thesis, Magdalini P. Eirinak, Athens University of Economics and Business Dept. of Informatics May 2006
[28] Study of Pre-processing Methods in Web Server Logs, Dr.Sanjeev Dhawan, Mamta Lathwal, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 5, May2013
[29] Identifying User Behavior by Analyzing Web Server Access Log File, K. R. Suneetha, Dr. R. Krishnamoorthi, IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.4, April 2009
[30] Ms. Dipa Dixit, Fr.CRIT, Vashi, M Kiruthika," PREPROCESSING OF WEB LOGS", (IJCSE) International Journal on Computer Science And Engineering, Vol. 02No. 07, 2010, 2447-2452.
[31] Miller, J. (2005). Usability in e-learning.
[32] Castro, F., Vellido, A., Nebot, À., & Mugica, F. (2007). Applying data mining techniques to e-learning problems. Evolution of Teaching and Learning Paradigms in Intelligent Environment, 62, 183-221.
[33] Mobasher, B., Cooley, R., & Srivastava, J. (2000). Automatic personalization based on web usage mining. Communications of the ACM, 43(8), 142–151.

[34] Mahajan, R., Sodhi, J., & Mahajan, V. (2014). Usage pattern discovery from a web log in an Indian e-learning site. Education and Information Technologies.
[35] Etzioni. O. (1996). The world wide web: Quagmire or gold mine. Communications of the ACM, 65-68.
[36] Pamnani, R., & Chawan, P. (2010). Web Usage mining: a research area in web mining. Proceedings of ISCET 2010. Punjab, India.
[37] R. Agrawal, Tomasz Imielinski, Arun Swami," Mining association rules between sets of items in larger databases," Proceedings of ACM SIGMOD International Conference of Management of Data, Washington DC, May 1993, 207-216.
[38] L. Wang, J. Fan, L. Liu, H. Zhao," Mining Data Association Based on a Revised FP-Growth Algorithm," Proc. of the 2012 International Conference on Machine Learning and Cybernetics, Xian, July 2012.
[39] R. Narayanan, D. Honbo, G. Memik, A. Choudhary and J. Zambreno. "An FPGA Implementation of Decision Tree Classification," Proc. Conf. Design, Automation and Test in Europe (DATE), pp. 189-194, April 2007.
[40] P. Sampath, C. Ramesh, T. Kalaiyarasi, S. Sumaiya Banu and G. Arul Selvan," An Efficient Weighted Rule Mining For Web Logs Using Systolic Tree," in IEEE International Conf. of Advances in Engineering, Science and Management (ICAESM-2012), March 2012.
[41] Song Sun and J. Zambreno," Mining Association Rules With Systolic Trees."
[42] Liu Jian and Wang Yan-Qing." Web Log Data Mining Based on Association Rule," in Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2011.Liu Kewen, "Analysis of Preprocessing Methods for Web Usage Data", IEEE 2012
[43] Sheetal A. Raiyani, Shailendra Jain and Ashwin G. Raiyani, "Advanced Preprocessing using Distinct User Identification in web log usage data", ISSN: 2278 – 1021, IJARCCE, Vol. 1, Issue 6, August 2012
[44] V. Sujatha and Punithavalli, "Improved User Navigation Pattern Prediction Technique From Web Log Data", ELSEVIER-2012
[45] Hongzhou Sha, Tingwen Liub, Peng Qin, Yong Sun and Qingyun Liu, "EPLogCleaner: Improving Data Quality of Enterprise Proxy Logs for Efficient Web Usage Mining", ELSEVIER-2013