An Enhanced Integrated Approach To Increase The Hit Ratio In Web Usage Mining

Pushpraj Singh Chauhan
PhD Scholar
Mewar University, Chittorgarh

Dr. Sarvottam Dixit
Professor
Mewar University, Chittorgarh

Dr. Suresh Jain
Prof. & Head CSE
Medicaps University Indore

Abstract:- : In today's era we have seen that there is a huge impact of Internet and the World Wide Web (WWW) in our Society and this impact is increasing day by day. Because of the increase in the use of internet, the business of E-Commerce is also increasing. WWW is increasing day by day where user interaction with the web is increased rapidly which provides the user perceived latency for accessing of resources from the web servers. This user behaviour is described in terms of web log data. Accessing of information from web log is accomplished by web usage mining. Through web usage mining useful information as identity of user, web user clustering, pattern discovery and analysis, pre-fetching and caching is extracted. To improve the web server performance prefetching and caching is used where most frequently access pages are prefetched and cached. In this paper, a better approach for predicting the most accessed web pages is proposed. This paper also represents the K-means clustering according to user location and then applying Apriori and FP growth algorithm for accessing prefetched pages. Later on we have also applied LRU and LFU to further increase the Hit Ratio.

Keywords—Web Log, Web Usage Mining, Apriori, FP growth, LRU, LFU, Web Caching, Web Prefetching.

I. INTRODUCTION

Today our society is vastly dependent on various kinds of Web-based transaction systems such as shopping, banking, and reservation. Typical Web-based transaction systems are guest book systems, shopping cart systems, seminar room booking systems etc. These Web-based transaction system generators can be used to construct Web-based transaction systems by composition of Web transition diagrams without any programming ability. A complete Transaction is a set of operations that is performed on a Web and it is treated as a Single Event.

1.1. Web Mining

Web Mining is the application of data mining techniques to extract patterns from the WWW. This is the information that we can extract from Web. There are

various tool that are available and can be used to extract information from Weblogs which we can obtained from various Web Servers. As the name itself suggests these informations extracted by web Mining.

1.2. Web Usage Mining

To serve the needs of web based applications web mining is used which generates patterns from the web log file. [5]. Web usage mainly consist of 3 phases like Data preprocessing,Pattern discovery and pattern analysis.In data preprocessing unwanted data are removedfrom the web log file through many data preprocessing techniques.In pattern generation data mining algorithms are used to generate most accessed patterns and help them in prefetching and cachingAfter completion of these steps user can use this information according to their specific needs.It has application in various fields, such as web personalization, e-commerce, user recommendation, pre-fetching and caching pages for user navigation, improvement of Web design increase the performance of web server.

1.3. Web Caching

Web caching plays an important role in performance enhancement. is used to store those object in the cache which is likely to be requested in the near future[8].To manage the cache properly various cache replacement policies are used as size of cache is minimum to store all objects accessed by the user.

Mechanism to apply caching is of three levels- Main server level, proxy server level and client level [3],[4].Implementing proxy server decrease the response time of the user with less network bandwidth. So cache replacement policies are applied on a proxy cache to improve the better hit ratio [1],[2]. In many caching scheme hit ratio varies in a limited range from 30-40% in respect to other replacement policies [6],[7],[8]. In conventional policies web caching causes cache pollution problem where large number of objects are stored in a cache which are never requested by the user. In order to store hose objects which are likely to be requested by the user is first fetched to gain maximum hit ratio, decrease network traffic and loads on the server. Integration of web prefetching and caching is a necessary step.

1.4. Web Pre-fetching

User's latency can be reduced by using Web caching. This can be done by storing the objects already visited in a place closer to the end user. For improving the performance of a System web caching is not sufficient alone, It should be combined with other mechanism like Web pre fetching. Web pre-fetching is used to predicts the web pages that are likely to be accessed in the future. This is done by studying web log (user's behavior). The data from a proxy server is collected and then preprocessed. Later on clustering is done by using
K-Means algorithm based on location. After clustering, Apriori and FPgrowth algorithm is applied to generate the association rules from the clusters.

II. RELATED WORK
2.1. LITERATURE REVIEW

The promise of the internet and e-commerce has led to the increasing use of the internet for group action process. In today's era as the use of internet is increased rapidly, there is a need to improve the user satisfaction and requirement for business, for this proper study of web user behaviour is needed. So there is a concept of web usage mining which is studied and analysed by researchers.

2.1.1. Review of Web log Mining

Internet users are increasing day by day with the growth of technologies. User perceived latency becomes a serious concern for the web service providers. Researchers have been done tremendous efforts to combine various techniques from different domains to resolve this issue. To overcome network latency, prefetching of popular documents is focused by authors. Improvement of web server is increased by integration of prefetching and caching concept. It also enhances the running time of application 50% [9].

Cochrane & Markey et al. [10] describes the combination of two approaches. They suggests that TLA cab be combined with another types of analysis like questionnaire. This helps in providing a more clear and complete picture having advantages of both the approaches. Much literature was reviewed to identify the factors that may influence using Web Enabled Transaction Processing.

Young et al. [11] shown that Web Transaction Logs can be used as a tool of collection Management. The transaction monitoring of an information system is defined as "the automatic logging of the type, content, or time made by a person from a terminal with that system".

Garofalakis et al. [12] describes about hypertext, discovering web structure and hyperlink through data mining algorithms and techniques. Also clustering approach, based on generalization is introduced which incorporates attribute oriented induction.
Peters, T. A., Kaske, N. K., & Kurth et al. **[13]** describes that the Web logs can be used to predict the user's behavior. They also explains that Web logs can also be used to describe what searches patrons enter and when they enter. But Web Logs can not explain the reason that why they have entered.

2.1.2. Review of Pre-fetching and Caching
Waleed Ali , Siti Mariyam Shamsuddin, and Abdul Samad Ismail et al. [14] proposed that how Neuro-fuzzy system has been used for classifying web objects into either cacheable or uncacheable objects. Then, the trained neuro-fuzzy system has also been used in the past in order to predict Web objects that can be reaccessed

They have also explained how prefetching and caching approaches can be used for improving the performance of web proxy server. Web caching and prefetching complement each other by working independently or integrated. This paper also presents the web prefetching as spatial locality to predict next objects of requested objects while web caching represents temporal locality for revisited objects. Researchers have also discussed conventional and investigated approach to integrate web caching and prefetching.

Waleed Ali et al. describes that in Web caching a copy of frequently used Web pages is kept in the Proxy Server which is nearer to the user. Mainly, a cache is located in proxy server (proxy cache), client (browser cache) and origin server (cache server). A cache setting of any modern web browsers like Internet explorer, Crome, Mozila, safari can be noticed.

Farhan et al. [15] The author proposed an integrated approach, in which ANN and LRU both are used. ANN is used for caching decision policies and LRU is used for replacement policies for data objects.
Foong et al. [16] describes a model which is based on logistic regression.
Greeshma G. Vijayan1 and Jayasudha J. S et al. [17] describes about the issues in web traffic and network bottlenecks. As users grow rapidly day by day to access the internet so internet causes severe overloading of many sites and network links because many users didn't have patience to wait for downloading a web page. For accessing the web sites efficiently reduction of web traffic is necessary. This paper presents the web prefetching and web caching techniques to reduce the web latency. Web pre-fetching techniques and web caching reduces the web latency to predict web objects requested.

Later. Cobb & Elaarag et al. [18] describes how ANN can be used in making cache replacement policies. Here ANN is used to rate the objects and on the basis of these ratings objects get selected for replacement.

III. PROPOSED WORK

In the previous works, we have seen that the work has been done by using Markov model,Apriori algorithm and FP growth algorithm. These algorithms are used for prediction of user's behavior. These algorithms helps us in improving the performance of Web Server. In some paper Data mining algorithms are mined with location on the basis of user request.To overcome the limitation of pattern discovery algorithms,in this paper both are integrated with each other to give the better result.In the proposed work two different Pattern discovery algorithms are used to pgenerate frequent pages on the basis of association rules.We consider FP growth algorithm generates more patterns than Apriori algorithm but when they are used combine better result is evaluated.In the existing work users are clustered according to group, interest and time taken by the server to respond back also with location .In this work we map IP address to location to cluster it with k-means to improve the performance, we also used cache replacement algorithms LRU,LFU and Optimal replacement so that from proxy cache when user request for pages then it will be served through replacement policies to achieve more hit ratio.We map IP addresss with location because when user search for shopping mall then there is achance for searching restaurant also in that location, so accessing pages are likely to improve when we add location with pattern generation. this in our work we integrate both prefetching and caching technique to increase the precision by generating more patterns through FP growth algorithm, which is helpful to prefetch more pages and put them in the proxy cache. In addition we used LRU, LFU policies to properly manage the cache by replacing the pages which are not used by the user in the future and compare them on the basis of hit ratio with cache size.

We can formulate the complete work in the following pseudo code.

Step1 Web log data is collected from the proxy server log files.

Step 2 Data preprocessing is performed by removing unwanted data like .gif,.jpeg,.css etc with removal of   URLs generated by agents and web spiders also with status 200.

Step 3 Map the IP address with the user location and add it to Step 2.

Step 4 Apply K-means algorithms to cluster user on the basis of location.

Repeat Step 4 until all clusters are formed.

Step 5 Apriori and FP growth algorithms are applied to generate patterns based on the association rules.

      a. Generated rules predict the most accessed pages are likely to be requested.

      b. Pre-fetch the most accessed pages and store them in the proxy cache.

Step 6 Apply Cache replacement algorithms to generate more hit ratio and store those pages only in cache which can be request more likely in future and compare the hit ratio by two of them.

      a. Apply LRU which discard least recently accessed pages.

      b. Apply LFU which discard least frequently accessed pages.

Step 7 Compare LRU and LFU on the basis of future page request.

Step 8 Ends.

3.1. WEB LOG FILES

Whenever a user gets interacted with the Server then there is a generation of text file known as Web Log file. Web Log file contain useful information about the user like IP address, URL accessed, time stamp, number of bytes used, protocol details which are used and requesting method. Web log also contain some irrelevant information, removal of such kind of information and filtered pre-processed data is a necessary step in web mining.

Where web log contains:

- 114.67.87.34.3-IP address
- "-"-it is anonymous user id
- 22/6/2018:05:16:45- Access time of web page
- -700-Yime zone
- GET/HTTP- it is a http request method by the user
- 200-http status code
- 4587-it is number of bytes transmitted

3.2. WEB LOG DATA SET

Web log file is a dataset which we used in this work.Web log files are that files which are generated when user and web gets interacted.It contains user behaviour information.Various types of web log files are generated like referrer logs, access logs error logs and with client side cookies also.These web logs are available in various sources and can be downloadable for experimenting.

3.3. KINDS OF WEB LOG FILES

    a.   Web Server Logs(Server side)

    b.   Proxy Server Logs(Proxy side)

    c.   Browser Logs(Client side)

3.4. DIFFERENT WEB LOG FORMATS

Web log files which are generated have some file extensions which are listed below.

    a.   Common log file format:

It is comma separated files which is a straight line record in a standard text log file format.

b.   Microsoft IIS log format:

It is a standard format for storing web log files and contains more information than NCSA file.It is a non-customizable format of ASCII.

3.5. DATA PRE-PROCESSING

Data preprocessing is first step of web usage mining where data is processed by removing unwanted and irrelevant information from the web log file.There are various methods like data cleaning,session identification,data integration,data transformation involve in data preprocessing step.

3.5.1 Data Cleaning

Data cleaning is a step which removes unwanted data like .gif,.jpeg.jpg,.css,.png etc file extensions.After removal of these extension lines are filtered and represented in GUI format.

3.5.2 Session Identification

When user interact with the web this information is stored in a file.Record stored in the log is a session and represents a user request.

3.5.3 Data Integration

Data integration is a process of extracting useful information of user like IP address, URL requested by user and user ID.

3.5.4 Data Transformation

Data transformation is a process where web log file is converted to a standard format for experimenting because web log file contains long string which is difficult to processed.

3.5.5 Location Annotation:

Location annotation is that where user location of user is found in terms of IP address.For k-means step where cluster is formed on the basis of user location,but in a log file location information is not available.So to map IP address to location this is a very necessary step.

3.6. K-MEANS CLUSTERING

Macqueen (1967) determine a process for partitioning an N-dimensional species into k sets on the basis of a sample and the method of processing is termed as 'k-means', which tends to cluster similar type of data in one cluster.. It is a very simple algorithm to implement which classify data in a useful manner.Main aim of this algorithm is to cluster similar data and grouped them on the basis of similar characteristics. In this we have cluster user on the basis of location. Grouping of user on the basis of same location is identified and clustered.Cluster formation is done by given no of clusters. As understood k-means concept gives a generalization of the standard sample mean , which tend to study the relevant asymptotic behavior, and the object being to represent some sort of rules of large numbers for the k-means algorithm.

K-means can be represented as iterative algorithm which is refined to minimize a squared error within the cluster. Then each cluster is first initialized to a random point by setting mean in the data set.In k-means each step is processed by performing number of steps as first data points are classified as a member of the cluster with the nearest cluster mean and secondly the cluster means are updated to be the actual mean of the data points used in each cluster and this process is repeated until no points membership changes , for some maximum number of repeated iterationss. k-means run many times with many different initializations. After cluster formation on the basis of user location is given for pattern generation algorithm.

3.7. APRIORI ALGORITHM

To find frequent items from the data set apriori algorithm is used on the basis of association rules applied. Association rules are rules which are applied when generating patterns.In our work we have used apriori algorithm to generate user access pages, pages which are frequently accessed by the user is discovered for prefetching and caching mechanism.Apriori algorithm generates frequent pages in a less amount as compared to FP growth algorithm, so we represent both apriori and FP growth algorithm for pattern generation.Time taken to generate frequent pages is less in apriori.It possess some applications on the basis of market basket problem like google search engine, shopping malls inherit the property of association rule to place items which are related to each other so that user can buy them easily,many online sites are applying reviews on the basis of product purchased and also recommend item similar to that item which user buy.

3.8 FP GROWTH ALGORITHM

FP growth algorithm is most popular and useful for generating frequent patterns on the basis of association rule applied.FP algorithm generates more items as more association rule is applied to it but it takes more time.It works in divide and conquer manner as it performs two database scan,in first scan frequent items are found and sorted in descending order.In second scan for each of the frequent item algorithm generates FP tree where support value is greater or same as threshold value.In our work we generate user access frequent pages from FP growth which is helpful for prefetching and caching the most accessed pages.As user access page "a" and "b" at the same session then there is a chance for him to visit "c"page also based on the association rule applied. After this we can Fetch Most Occurring Pages and then cache those Web pages.

3.9 CACHE REPLACEMENT ALGORITHMS

Cache replacement algorithms are those algorithms which are mainly used to discard those pages which are no longer in use and manage the cache to store those pages that are used in the near future. If page which is requested is not in the cache then page

fault occurs then that page is to be replaced with the existing page to put it in the cache. various cache replacement policies are used to manage the pages in the cache. In this paper LRU and LFU Page Replacement Policies are used so that pages which are requested by the user is properly stored in the cache with minimum number of page faults and maximum hit ratio.

3.9.1 LRU
LRU is least recently used algorithm which replace the page that are least used and manage the cache properly with min page faults and max hit ratio. LRU is based on greedy approach where pages are constantly replaced until optimal replacement is achieved. It is a best algorithm to achieve hit ratio in terms of cache size as compare to LFU and Optimal.

3.9.2 LFU
LFU is a page replacement algorithm that replace the page which less frequent, so that proper management of cache is done.
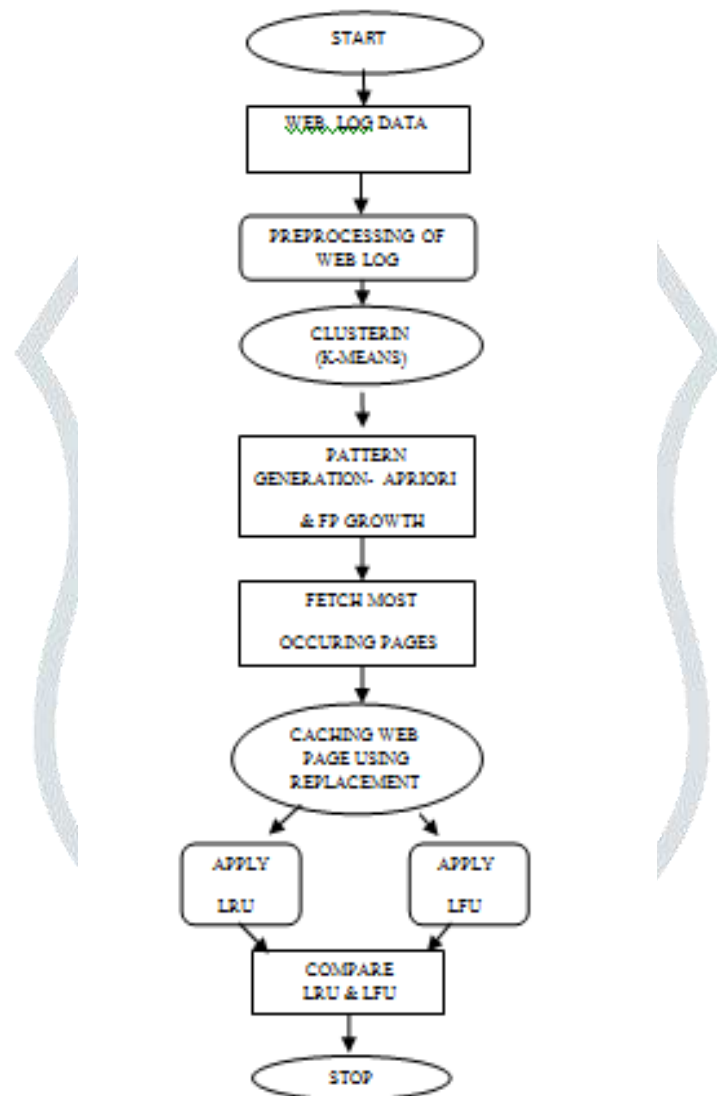
Proposed Algorithm

IV. EXPERIMENTAL WORK AND RESULTS

Here we have describe about the experimental result of our work and how it is achieved. WWW has become an emerged technology for storing, retrieving, distributing and sharing of data. Due to user's accessibility web becomes a dominant platform to predict user behaviour. When users interact with the web then all the information is stored collectively in a text file called as weblog. Information which is collected is having some unwanted and inappropriate data so, we want to process these irrelevant data. After processing these data is mined this is termed as web mining. In this thesis we have described about web usage mining where pre-processing, pattern generation & pattern analysis is studied and also to discuss about prefetching and caching of frequent data which is useful in E-commerce, web personalization.

4.1. INTRODUCTION
For the experimental work dataset from the ircache.net website is collected and is used for experiment which is available at the proxy server installation ftp://ircache.net. In the proposed framework, first web log data is pre-processed by different pre-

processing schemes by removing unwanted data like .gif, .jpeg etc and remove the URLs generated by agents and web spiders also with status 200.

In our proposed framework we collect the raw data from the log file initially which is processed later.
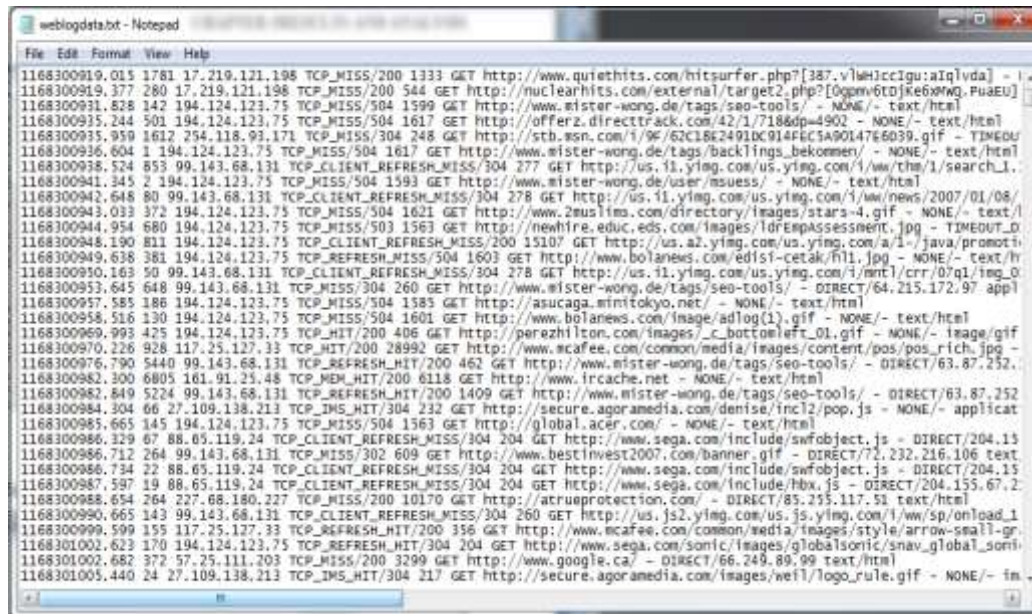Fig 1 shows the sample of the raw proxy log file.

Figure 1: Web Log File

Fig 2 Interface of our work is implemented in Java platform

Fig 3 shows the snapshot of the starting phase where we pass the file for pre-processing. The file in stored in text format. We have applied rules to pre-process the file and remove the unwanted data. The text file is processed one line at a time.

web log data is pre-processed by different pre-processing schemes by removing unwanted data like .gif, .jpeg etc and remove the URLs generated by agents and web spiders also with status 200.
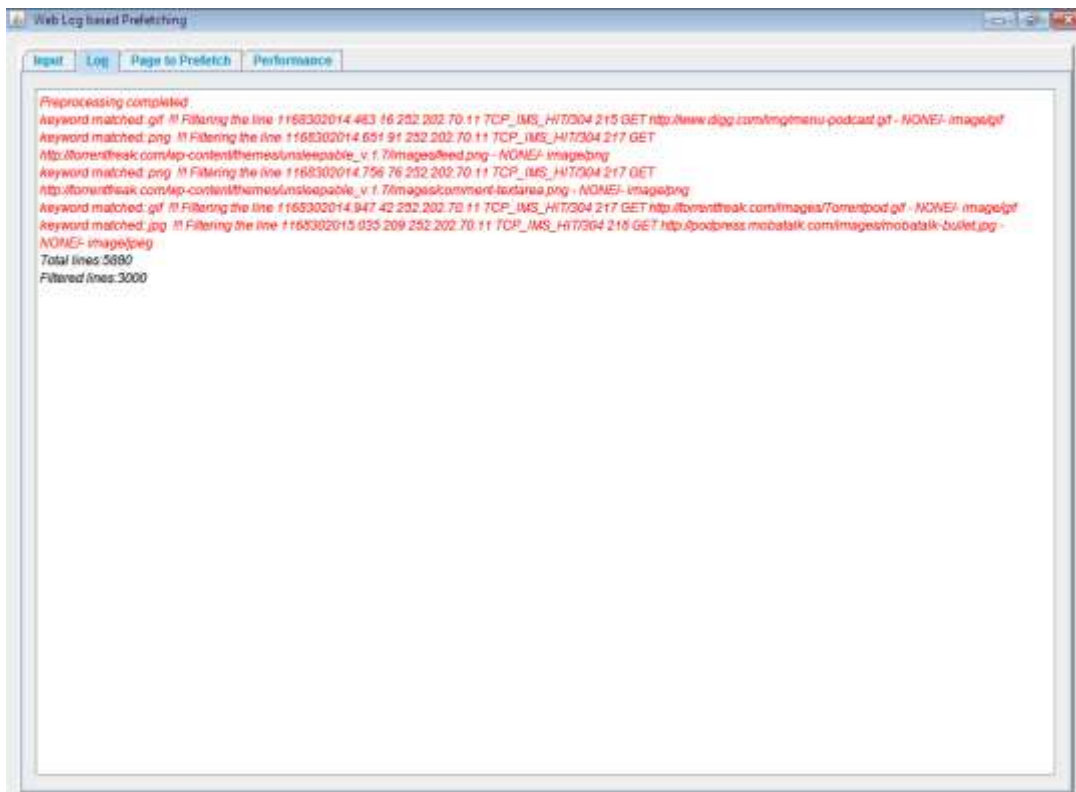
Figure 3: Web Log File Preprocessing

Fig 4 shows pre-processing where lines are filtered from 5500 to 3000.Then through k-means clustering url access by the users are clustered on the basis of location by mapping IP address to location. Then each url is assigned unique number for identification.
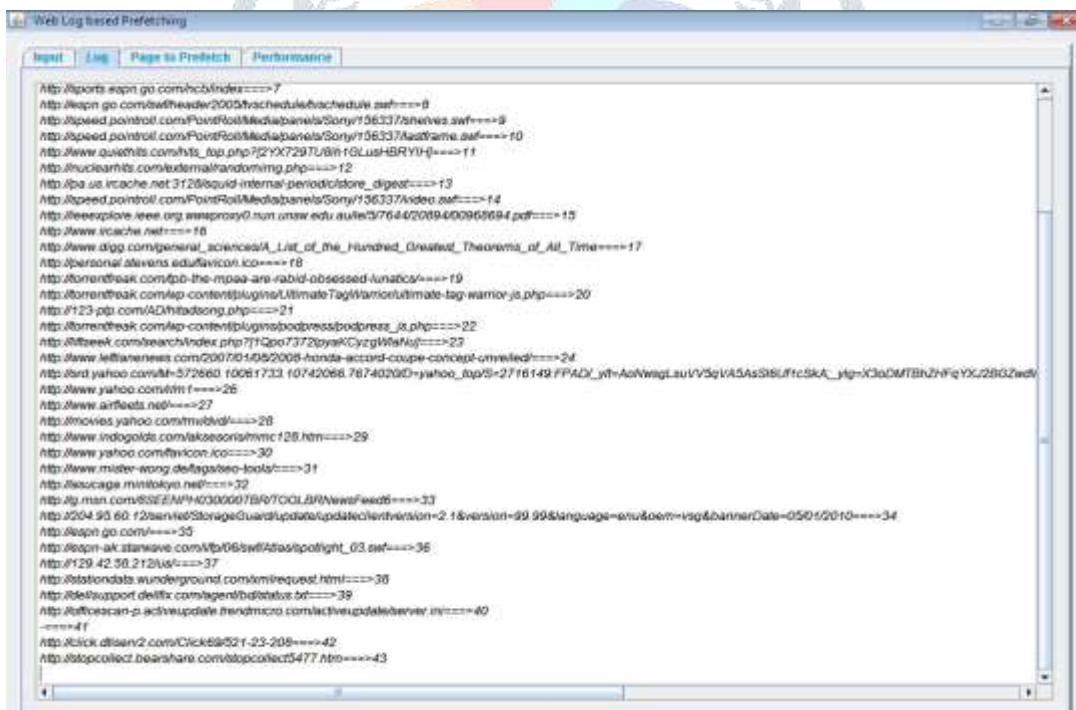

Figure 4 : URL Cluster formation

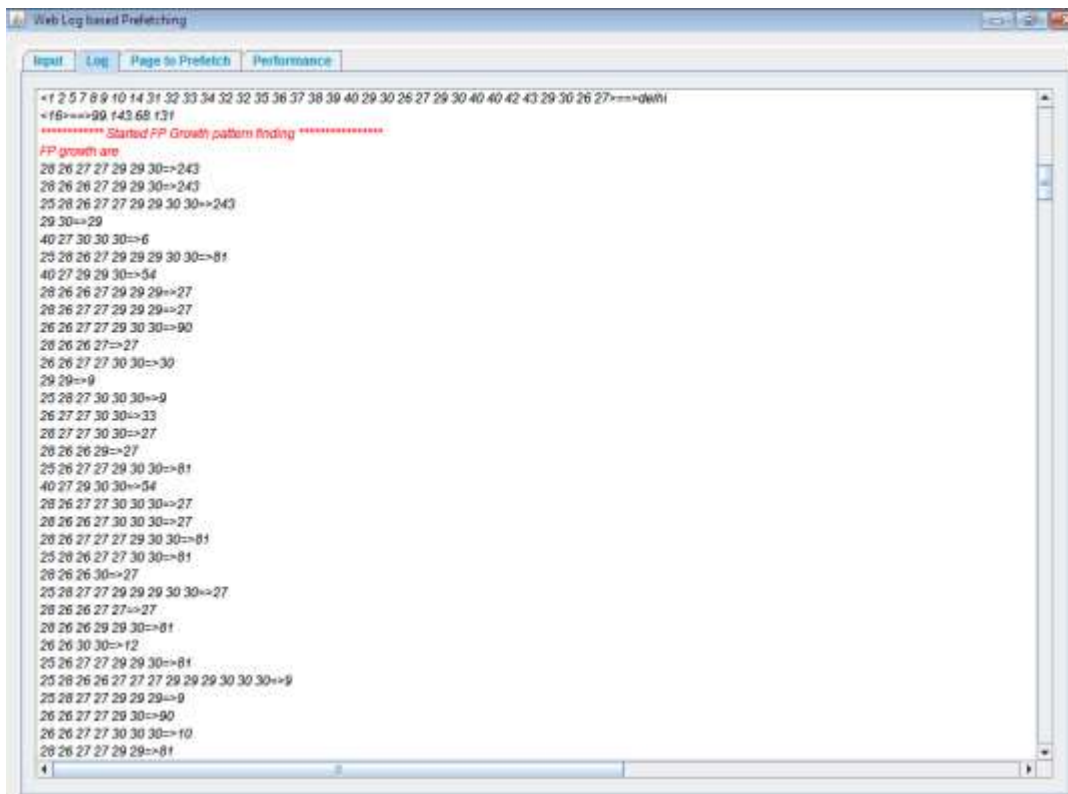Fig 5 shows FP growth algorithm is applied to generate frequent urls.

Figure 5 : FP growth Pattern Generation

Fig 6 represents FP growth generates patterns on the basis of user location.
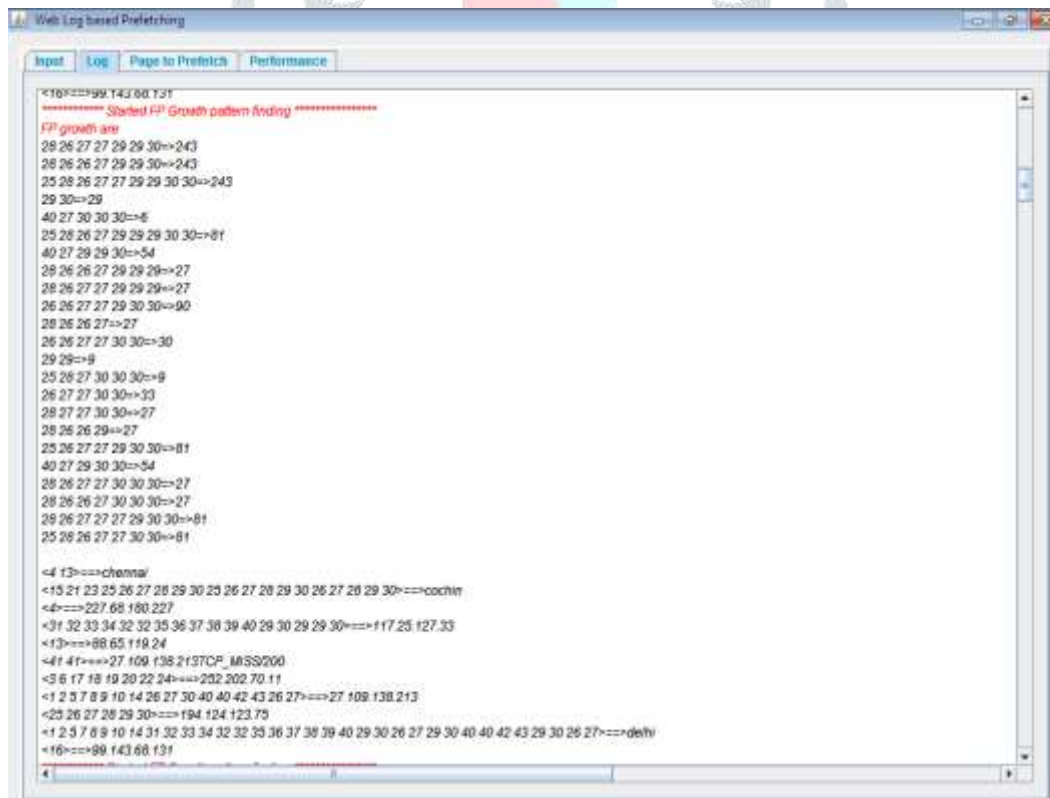
Figure 6 : FP growth pattern generation with user location

FP growth generates more patterns then Apriori algorithm. Thus FP growth is best as compared to Apriori for pre-fetching frequent pages.

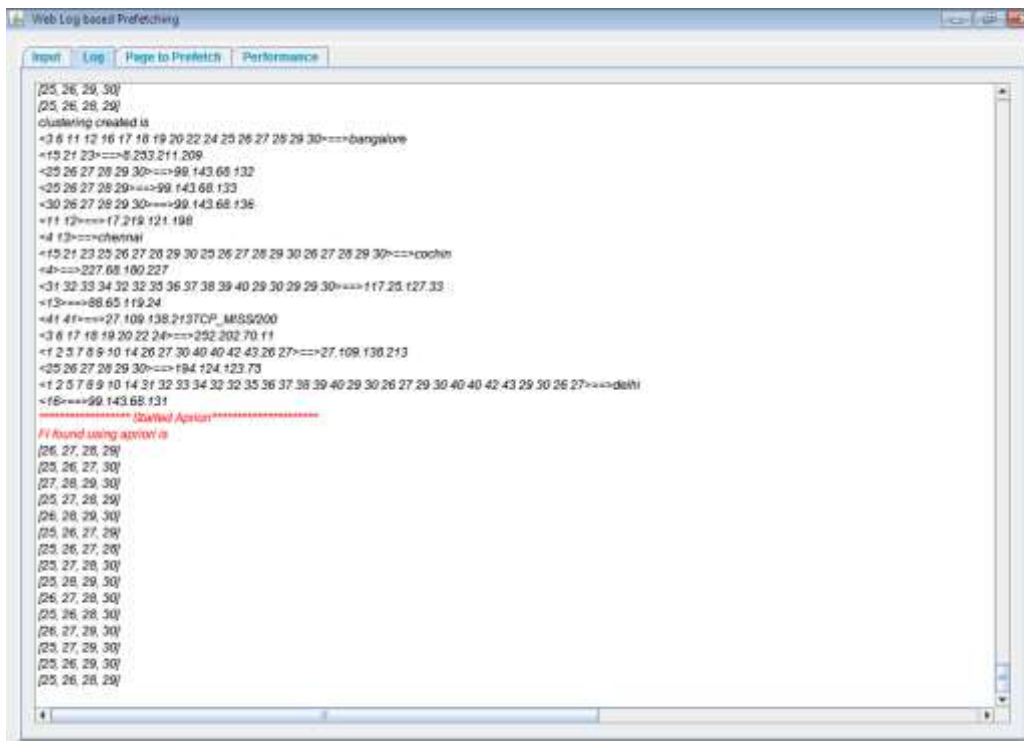Fig 7 shows Apriori algorithm to generate patterns.

Figure 7: Apriori algorithm pattern generation with user location

Fig 8-After this the predicted pages are fetched from the server and the most accessed pages are stored in cache of proxy server.

Figure 8: Prefetched Pages

In FP growth more association rules is applied as compare to apriori so it generates more number of patterns but access time is more in FP growth.
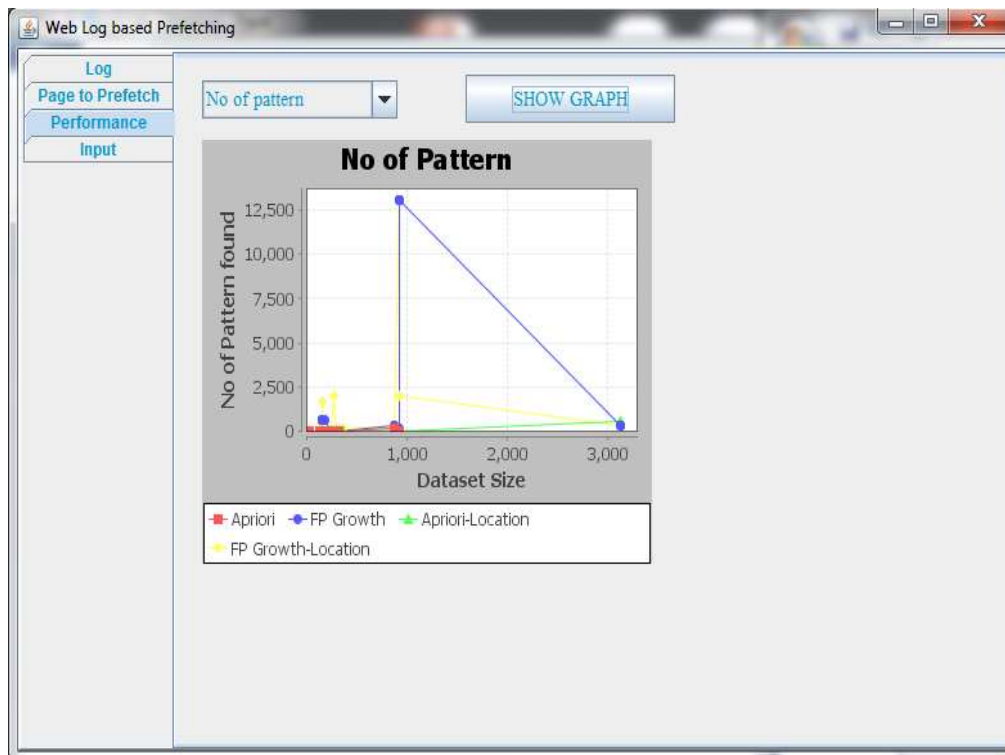
Figure 9: Pattern Generation Graph

Fig 9 It shows FP generates more patterns .After pattern generation pages are pre-fetched and stored in a cache.
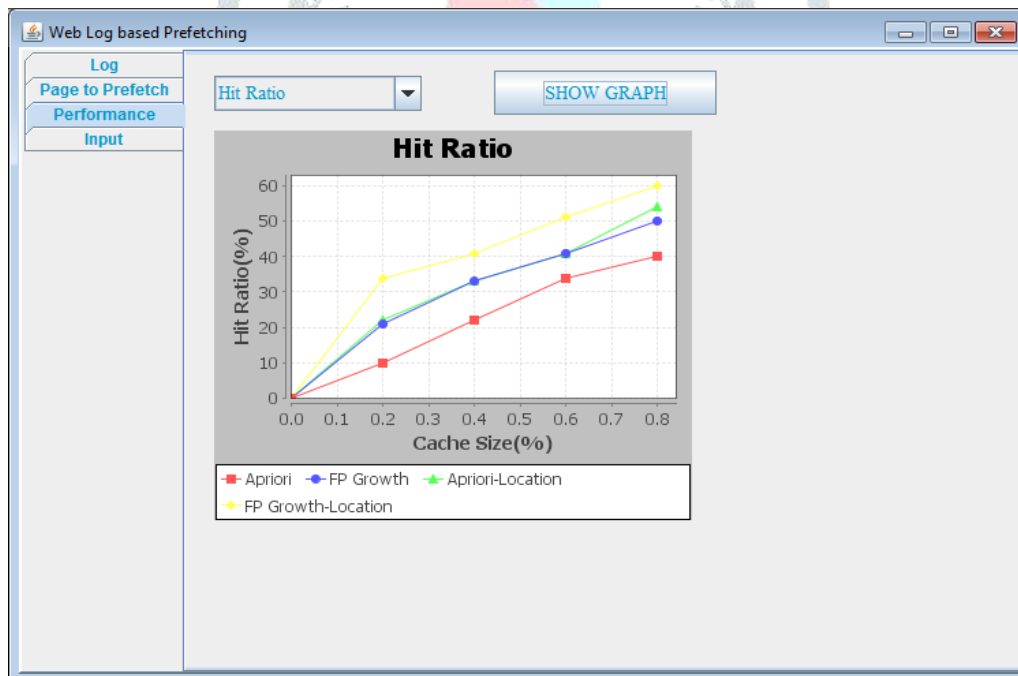

Figure 10 Hit Ratio Generation Graph

Figure 10 shows the hit ratio of the algorithms. Hit ratio is increased by 5% to 8% when clustering is done on the basis of location. Above graph also shows that FP growth performs better than Apriori.

To improve the performance of caching various Cache Replacement algorithms like LRU and LFU are implemented and compared. When most access pages are stored in proxy server of cache then that pages are applied to replacement algorithms to gain the more hit ratio because when user request the same pages then it would be cached through LRU and LFU algorithms.

In our work first we have used web log which gives user information, after applying pre-processing on web log it is filtered to get useful information. Then cluster is formed on the basis of IP location of user, that location is found by mapping IP address to user

location. Now frequent patterns are generated through mining algorithms on he basis of location. After this prefetched urls are stored in a proxy cache, that url which is prefetched is given specific number to have unique identification. This urls is map to unique number which is stored in a cache.

Now this unique number is passed to cache replacement policies to manage the cache according to request made by the user in the future and compare the result by calculating hit ratio in terms of cache size.
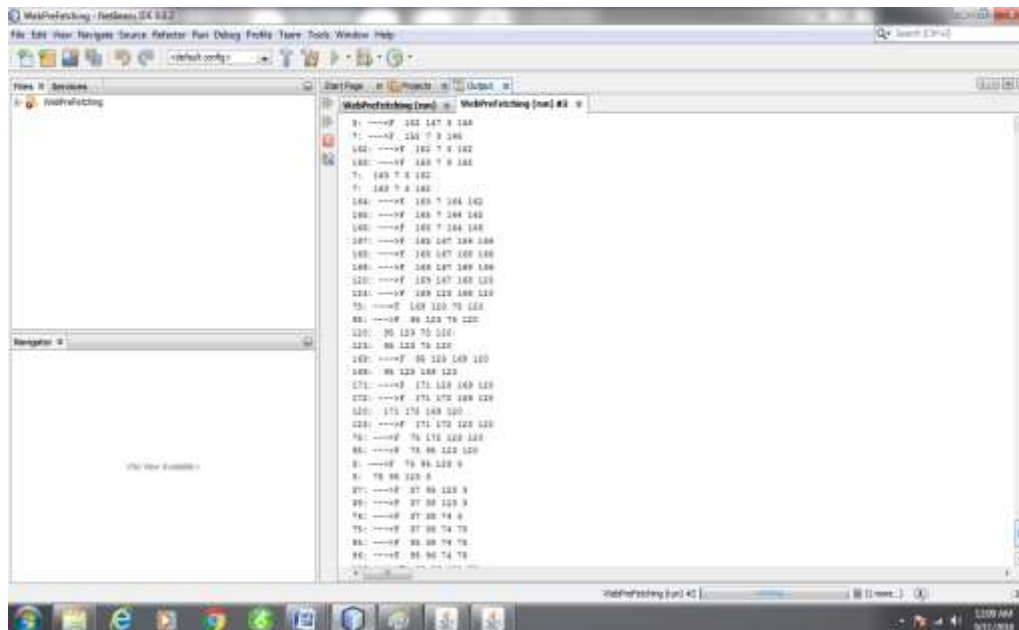

Figure 11 LRU Page Faults

Fig 12-LRU(Least Recently Used) used to discard urls which are least recently used and replace them with the new request .Graph to represent the hit ratio by LRU.
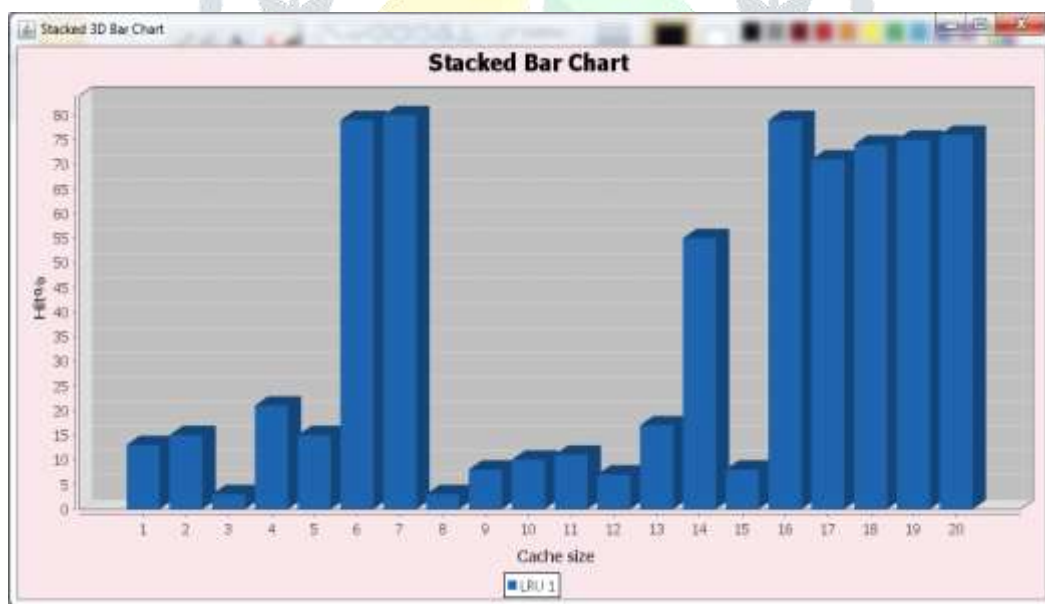

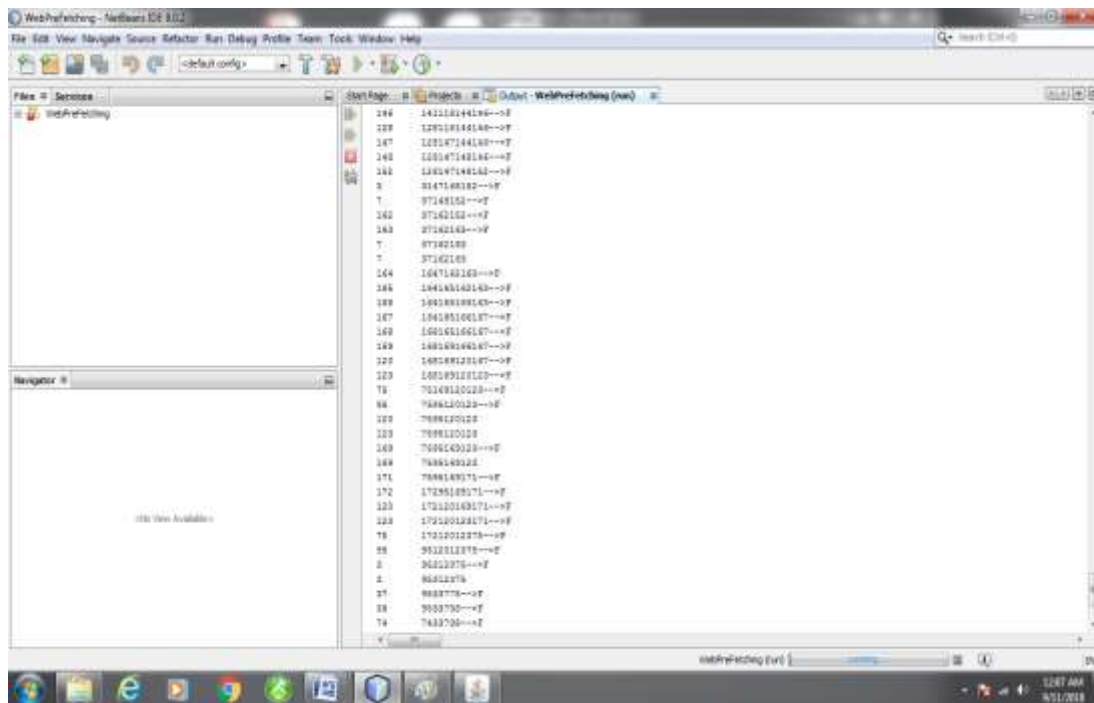**Figure 12 : LRU Hit Ratio Generation Graph**

Figure 13 LFU Page faults

Fig 14-LFU (Least frequently used) is used to discard urls which are least frequently used and replace them with the new request made by the user. Graph to represent hit ratio by LFU.
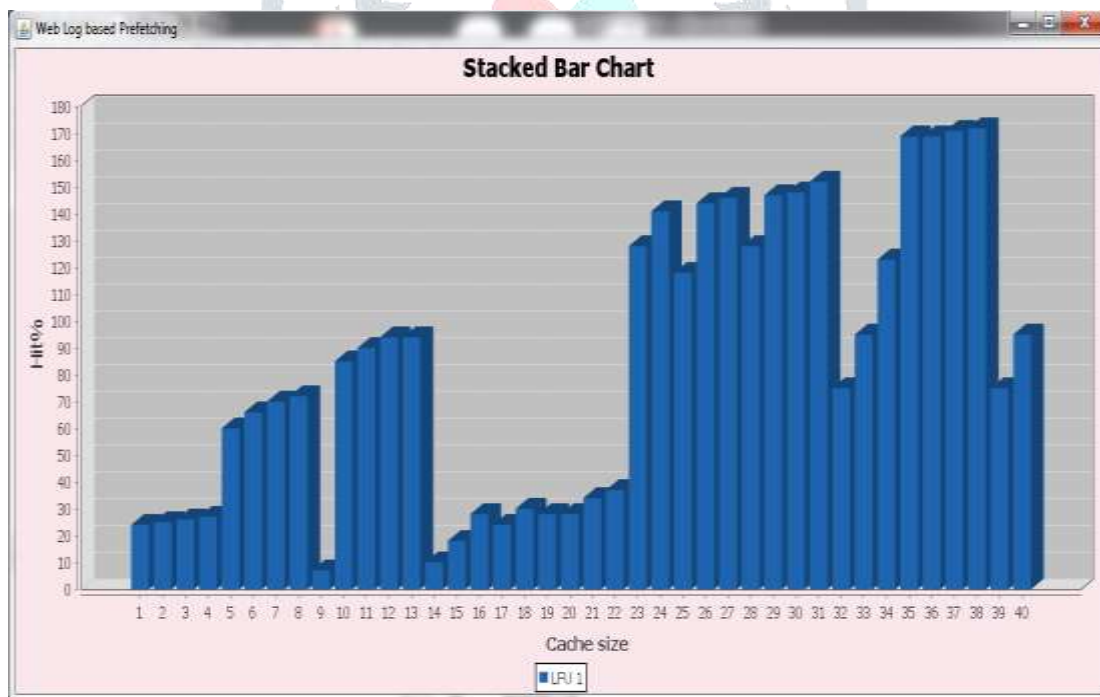

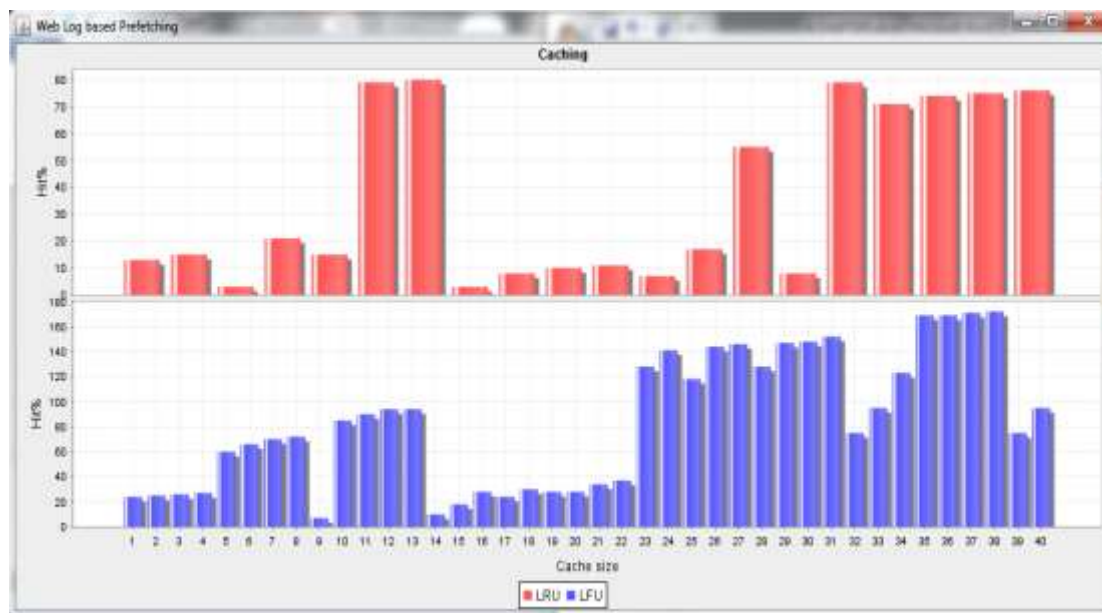**Figure 14: LFU Hit Ratio Generation Graph**

FIGURE 15 COMPARISON BETWEEN LRU AND LFU

V. SUMMARY

Here we have shown how web usage mining is implemented and results are evaluated by graph. Also cache replacement policies are used to manage the cache properly with the help of prefetching technique.

VI. FUTURE WORK

In the future scope Fuzzy c-means clustering can be used and other caching policies can be applied to increase the web server performance and hit ratio.

VII. REFERENCES

[1] T. Koskela, J. Heikkonen, ,and K. Kaski, (2003). "Web cache optimization with nonlinear model using object feature", Computer Networks journal, elsevier , 43(6), (2003), pp. 805-817.

[2] J. Cobb, and H. Elaarag, "Web proxy cache replacement scheme based on back-propagation neural network", Journal of System and Software, 81(9), (2008), pp. 1539-1558.

[3] R. Ayani, Y.M. Teo, and Y.S. Ng, "Cache pollution in Web proxy servers", International Parallel and Distributed Processing Symposium (IPDPS'03), 22-26 April 2003, pp.7.

[4] A.K.Y. Wong, " Web Cache Replacement Policies: A Pragmatic Approach", IEEE Network magazine, 20(1), (2006), pp.28–34

[5] Web Usage Mining: Discovery and Applicaions of Usage Patterns from Web Data, Jaideep Srivastava, Robert Cooleyz , Mukund Deshpande, Pang-Ning Tan, ACM SIGKDD, Jan 2000.

[6] I. R. Chiang, P. B.Goes, and Z. Zhang, "Periodic cache replacement policy for dynamic content at application server", Decision Support Systems, Elsevier, 43 (2), (2007), pp. 336- 348.

[7] H.k. Lee, B.S. An, and E.J. Kim, "Adaptive Prefetching Scheme Using Web Log Mining in Cluster-Based Web Systems", 2009 IEEE International Conference on Web Services (ICWS), (2009), pp.903-910.

[8] L. Jianhui, X. Tianshu, Y. Chao. "Research on WEB Cache Prediction Recommend Mechanism Based on Usage Pattern", First International Workshop on Knowledge Discovery and Data Mining(WKDD), (2008), pp.473-476.

[9] K. Chinen and S.Yamaguchi.. An Interactive Prefetching Proxy Server for Improvement of WWW Latency. In Proceedings of the Seventh Annual Conference of the Internet Society (INEt'97), Kuala Lumpur, June 1997.

[10]. Cochrane, P. A. & Markey, K. (1983). Catalog use studies since the introduction of online   interactive catalogs: impact on design for subject access. Library & Information Science   Research, 5(4), 337-363.

[11] Young, I. R. (1992). The Use of a general periodicals bibliographic database transaction log as   a serials collection management tool. Serials Review, 18(4), 49-60.

[12] Garofalakis M. N., Rastogi R., Sheshadri S., and Shim K.,"Data mining and the Web: past, present and future." InProceedings of the second international workshopon Web information and data management, ACM, 1999.

[13] Peters, T. A., Kaske, N. K., & Kurth, M. (1993). Transaction log analysis. Library Hi Tech Bibliography, 8, 151-183.

[14] W. Ali, and S.M. Shamsuddin, "Intelligent Client-side Web Caching Scheme Based on Least recently Used Algorithm and Neuro-Fuzzy System", The sixth International Symposium on Neural Networks(ISNN 2009), Lecture Notes in Computer Science (LNCS), Springer-Verlag Berlin Heidelberg , 5552, (2009), pp. 70–79.

[15] Farhan, Intelligent Web Caching Architecture. Master thesis. Faculty of Computer Science and Information System, UTM University, Johor, Malaysia, (2007).

[16] A. P. Foong, Y.-H. Hu, and D. M.Heisey, "Logistic regression in an adaptive web cache", IEEE Internet Computing, 3,

(1999), 27-36.

[17] A Survey On Web Pre-Fetching and Web Caching Techniques in a Mobile
Environment"(Greeshma G. Vijayan1 and Jayasudha J. S.) (2012)

[18] J. Cobb, and H. Elaarag, "Web proxy cache replacement scheme based on back-propagation neural network", Journal of System and Software, 81(9), (2008), pp. 1539-1558.

[19] Nanhay Singh, Arvind Panwar and Ram Shringar Raw. Enhancing the performance of Web Proxy Server using Cluster Based Pre-fetching technique. IEEE 2013.

[20] Toufiq Hossain Kazi, Wenying Feng and Gongzhu Hu "Web Object Prefetching: Approaches and a New Algorithm", IEEE 2010, pp 115-120.

[21] J. B. Patil and B. V. Pawar, "Improving Performance on WWW using Intelligent Predictive Caching for Web Proxy Servers," IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011.

[22] C. Kumar and J.B. Norris, "A new approach for a proxy-level Web caching mechanism", Decision Support Systems, Elsevier, 46(1), (2008), pp.52-60.

[23] R. Chiang, P. B. Goes, and Z. Zhang, "Periodic cache replacement policy for dynamic content at application server", Decision Support Systems, Elsevier, 43 (2), (2007), pp. 336- 348.

[24] H.T. Chen, Pre-fetching and Re-fetching in Web caching systems: Algorithms and Simulation, Master Thesis, TRENT UNIVESITY, Peterborough, Ontario, Canada(2008).

[25] T.Chen, "Obtaining the optimal cache document replacement policy for the caching system of an EC Website", European Journal of Operational Research.181(2),(2007), pp. 828. Amsterdam.

[26] Koskela, J. Heikkonen, ,and K. Kaski, (2003). "Web cache optimization with nonlinear model using object feature", Computer Networks journal, elsevier , 43(6), (2003), pp. 805-817

[27] H.k. Lee, B.S. An, and E.J. Kim, "Adaptive Prefetching Scheme Using Web Log Mining in Cluster-Based Web Systems", 2009 IEEE International Conference on Web Services (ICWS), (2009), pp.903-910.

[28] L. Jianhui, X. Tianshu, Y. Chao. "Research on WEB Cache Prediction Recommend Mechanism Based on Usage Pattern", First International Workshop on Knowledge Discovery and Data Mining(WKDD), (2008), pp.473- 476.

[29] Abhari, S. P. Dandamudi, and S. Majumdar , "Web Object-Based Storage Management in Proxy Caches", Future Generation Computer Systems Journal , 22(1-2), (2006). pp. 16-33.

[30] Seung Won Shin, Byeong Hag Seong & Daeyeon Park, (2000)"Improving World-Wide- Web Performance Using Domain-Top Approach to Prefetching", Fourth International Conference on High- Performance Computing in the Asia-Pacific Region vol. 2, pp. 738-746.

[31] Sandhaya Gawade and Hitesh Gupta, "Review of algorithms for prefetching and caching" 2012 IJARCCE

[32] HeunKi Lee, Baik Song an, and Eun Jung Kim, "AdaptivePrefetching Scheme Using Web Log MininginClusterbasedWeb", InternationalConference on Web Services, IEEE, 2009, pp. 903-910.

[33] Wenying Feng and Hua Chen ,"A Matrix Algorithm for Web Cache Prefetching", International Conference IEEE/ACIS, 2007, pp. 788-794.

[34] J. B. Patil and B. V. Pawar, "Improving Performance on WWW using Intelligent Predictive Caching for Web Proxy Servers," IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011.

[35] MukeshDawar, Charanjit Singh, "A Review on Web CachingTechniques",International Journal of Advanced Research inComputer Science and Software Engineering, ISSN: 2277 128X,Volume 4, Issue 3, March 2014.

[36] B.lalithadevi, A.Merry Ida, W.AncyBreen "A new Approach for improving world wide web techniques in data mining".IJARCSSE,2013.

[37] Suvarnatemgire, poonam Gupta, "Review on web prefetchingtechniques", IJTEEE,2013

[38] Anuradha,kiran jain, "Survey on an Integrated Approach for Web Pre-Fetching and Caching," Vol. 2, Issue 02, 2014 | ISSN (online): 2321-0613.

[39] Bhumika garg,Ravinder singh , "Hybrid Approach for Performance of Web Page Response through Web Usage Mining "
Vol. 4, Issue 07, july 2014 | ISSN (online): 227 7-128X.