An Efficient Cluster Based Web Object Filters From Web Pre-Fetching And Web Caching On Web User Navigation

Dr. A. K. Santra[1], S. Jayasudha[2]
[1] **Dean, CARE School of Computer Applications ,**
Trichy – 620 009, India.
[2] **Research Scholar, Bharathiar University,**
Coimbatore – 638401, India.

Abstract

The World Wide Web is a distributed internet system, which provides dynamic and interactive services includes on line tutoring, video/audio conferencing, e-commerce, and etc., which generated heavy demand on network resources and web servers. It increase over the past few year at a very rapidly rate, due to which the amount of traffic over the internet is increasing. As a result, the network performance has now become very slow. Web Pre-fetching and Caching is one of the effective solutions to reduce the web access latency and improve the quality of service. The existing model presented a Cluster based pre-fetching scheme identified clusters of correlated Web pages based on users access patterns. Web Pre-fetching and Caching cause significant improvements on the performance of Web infrastructure.

In this paper, we present an efficient Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme to evaluate the web user navigation patterns and user references of product search. Clustering of web page objects obtained from pre-fetched and web cached contents. User Navigation is evaluated from the web cluster objects with similarity retrieval in subsequent user sessions. Web Object Filters are built with the interpretation of the cluster web pages related to the unique users by discarding redundant pages. Ranking is done on user's web page product preferences at multiple sessions of each individual user. The performance is measured in terms of Objective function, Number of clusters and cluster accuracy.

Keywords: Web usage mining, web mining, web log files, Web Proxy

1. Introduction

The Web has growing speedily from a trouble-free communication method presenting only static objects to a prosperous collection of dynamic and interactive services. The volatile development of the Web has forced an intense demand on networking objects and Webservers. Web users frequently have seen lengthy and impulsive delays when recovering Web pages from remote sites. Therefore, clear result in order to develop the quality of Web services might be the enhancement of bandwidth, although such a preference engages rising financial cost. Nevertheless, the greater bandwidth might solve provisionally the troubles since it will ease the web users to make more and more web objects-need applications, grouping again the network. Therefore, the network restrictions can stay or get worse unless efficient software results are also given.

Caching verified itself as a significant procedure to optimize the method the Web is used. Specially, most of the Web caching features are initiated from the caching technique developed in a variety of network systems while Web caching presents new distresses in Web objects organization and recovery across the network. Particularly, Web caching is created by proxy server applications implemented to maintain several users. Proxy applications work as an intermediary between Web users and servers. Web users create their link to proxy applications executing on their hosts. The proxy attaches the web server and relays data between the web user and the web server. At every request, the proxy server is gone primary to discover whether it contains a suitable copy of the requested object. When the proxy contains the requested web object this is assumed as a cache hit, otherwise a cache ignore happens and the proxy server should promote the request on behalf of the web user. Upon getting a new web object, the proxy services a copy to the final web user and maintains one more copy to its local storage.

Web caching minimizes bandwidth utilization, network jamming, and network traffic due to the storage of the recurrently requested web object nearer to users. Also, since it distributes cached web contents from proxy servers, it decreases external latency which means the time taken to convey contents from the origin web server to proxy servers. At last, caching increases consistency since users attain a cached copy even if the remote server is occupied. The performance of a Web proxy caching scheme, it is mostly dependent on the cache alternate technique which discover the objects to be restored in a cache upon a request entrance which has been improved by the fundamental proxy server. Though, cache hit rates have not enhanced greatly with these method. If the proxy is not correctly updated, a user will obtain decayed data, and, when the number of users grows, original web servers normally become bottlenecks.

In this paper, we propose an efficient Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme which evaluates the web user navigation patterns and user references of product search. Particularly, the probable major advantage of accepting Pre-fetching policies over a proxy cache server is that we handle efficiently the Web object by using both the temporal and the spatial locality of contents.

2. Related Work

Pre-fetching and caching are the famous techniques for enhancing the performance of the Web. They have become important objects the Web infrastructure. The advantages of these approaches have provided go up to new manufacturing which includes equipment and service dealers that provide cache servers. It presents caching and pre-fetching services to clients and suppliers of Web resources [1]. These days, a number of profitable systems apply various type of pre-fetching. For illustration, a number of browser additions for FireFox [2], Netscape and Microsoft Internet Explorer as well as a number of individual proxies that performs pre-fetching of associations of the present page [3].

The Web pre-fetching methodologies have been distinguished corresponding to its temporary and lasting remuneration [4]. In this framework, we classify the presented pre-fetching policies as follows: temporary pre-fetching policies: prospect requests are guessed to the cache's new admission account. Based on these forecast, clusters of Web objects are pre-fetched [5]. In this situation, the temporary pre-fetching schemes employ Dependency Graph (DG) [6], wherever the patterns of contacts are detained by a graph and Prediction by Partial Matching (PPM) [7], where a system is employed, accepted from the text compression domain [8]. Further, several temporary pre-fetching policies [9] are depended on Markov models, which are employed for modeling and forecasting user's browsing performance above the Web.

Lasting pre-fetching policies [10], worldwide content admission pattern statistics for example, objects' regard, objects' steadiness is employed to discover precious (clusters of) contents for pre-fetching. In this kind of system, the contents with advanced contact frequencies and no longer inform time periods are additional probable to be pre-fetched [11]. The previous Web caching schemes employ the temporary pre-fetching strategies, by pre-fetching contents which are probable to be referenced in the near future. Whereas, the lasting pre-fetching scheme will be related on duplication schemes includes a Content Distribution Network (CDN) platform [12] as well as on mobile computing settings. Now, we compact with temporary pre-fetching policies that are applied on a Web caching environment.

3. Efficient Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme

Efficient Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme is proposed to evaluate the web user navigation patterns and user references of product search. Web page objects are clustered by using pre-fetched and web cached contents. The web cluster objects with similarity retrieval in subsequent user sessions evaluate the User Navigation. Web Object Filters are developed with the interpretation of the cluster web pages associated to the distinctive users by the removal redundant pages.

3.1. Web Pre-fetching and Web caching scheme

Web pre-fetching is the method of presuming user's upcoming requests for Web objects by placing admired requested objects into the cache proceeding to precise request for web user. Contrasting Web data caching that uses the temporal area, the Web pre-fetching schemes are depended on the spatial area of Web contents. Specifically, the temporal area describes to frequent users' admissions to the similar object within short time periods, while, the spatial one describes to web users' requests where admits to a few web objects recurrently involve accesses to definite further objects. Normally, the major gain of using pre-fetching is that it avoids bandwidth underutilization and decreases the latency. Hence, bottlenecks and traffic congestion on the Web are bypassed and web contents are conveyed earlier. Therefore, the proxies will efficiently

provide more users' requests, decreasing the workload from the beginning servers. Therefore, the beginning servers are protected from the flash crowd events as an important component of the Web traffic is detached over the proxy servers. Whereas the main disadvantage of systems which have improved Pre-fetching methods is that some Pre-fetched web contents may not be ultimately requested by the web users. In such a case, the Pre-fetching scheme enhances the network traffic as well as the Web servers' content. In order to defeat this drawback, high accuracy forecast methods have been employed. From the above it happens that caching and pre-fetching balance each other in order to decrease the obvious response time apparent by web users.

3.2 Cluster based Web Object Filters

Our proposed model, Cluster based Web Object Filters Scheme Clustering the web page objects which is obtained from pre-fetched and web cached contents.

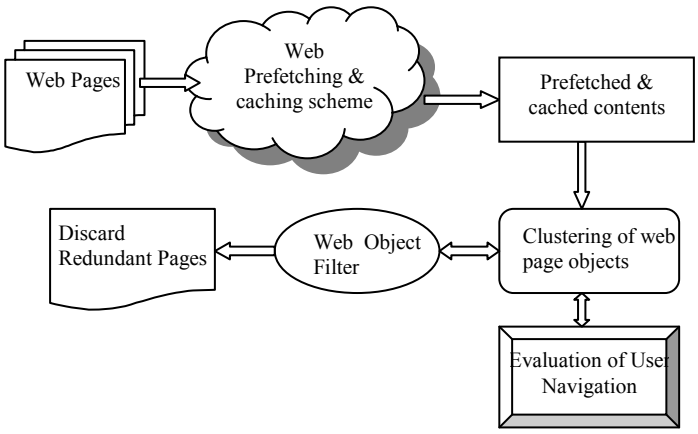

Fig 1: Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme

Web users visit a number of Web sites from time to time and utilize random amount of time between repeated visits. Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme examine the Web proxy log file for the changeable nature of Web browsing. Specially, the Web proxy access log is a chronological file with each and every user contact record per line. Each Internet Service Provider (ISP) contains a proxy server cache. The Web proxy log files offer information about web user activities from the instant the web user logs in to the ISP to the instant the unique user logs out from it.

After that, the contents generated by Web Pre-fetching and Web caching scheme are undergone a confident pre-processing, such as redundant object filter. Web Object filtering discards web pages in which are obviously generated by unique user. In addition, we filter the uncacheable requests. Then the IPs is grouped into several classes, according to their domains. The method of grouping the IPs into classes enhances the web data management. Further, it reduces the complication of the fundamental crisis as the number of domains is lesser than the number of single IPs. Moreover, with such a technique, we avoid incorrect hypothesis because in proxy logs, the real-world individual web users cannot be distinctively recognized and therefore multiple users are mapped into one IP address. Hence, our proposed Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme evaluates the web user navigation patterns and user references of product search efficiently.

3.3. Algorithm

Cluster based Web Object Filters from Web Pre-fetching and Web caching

Step 1: Initialization Phase

Initialize web pages objects $O = (h_i, b_i, s_i)$

Step 2: Web Pre-Fetching & Caching Phase

For each i Є O do

Evaluate the factor $f(i) = \frac{h_i b_i}{h_i b_i + 1}$

End for

Calculate the overall hit rate = $\sum$ hi (fi) where i Є O

Calculate the overall bandwidth rate = $\sum \frac{s_i}{b_i}$ *(fi)

where i Є O

For each i Є O do

Compute the increase factor $inc(i) = \frac{h_i pref}{b_i pref}$

End for

While (n objects) do

Select largest increase factor
Highlight it as "pre-fetched"

End while

Highlight others as "Non-re-fetched"

While (n objects) do

Select from n objects having longest lifetime
Highlight the object as "cache"

End while

Highlight the other objects as "Non-cache"

End while

Step 3: Cluster Formation

- Initialize the number of cluster
- Initialize the objects pre-fetched results obtained from second phase
- Let us use two objects A and B as the first centroids
- Calculate the distance using Euclidean model

$$d(A,B) = \Sigma\ (Bi - Ai)$$

Where i = 1,2,..,n

Assign each object based on the minimum distance

$$Min(dist) = (A \mid B)$$

The output obtained is cluster based web object

Step 4: Filtering and Evaluation Phase

- Cluster based web object is given as input set
- Select j most similar items $(j_1, j_2, \ldots j_k)$
- Compute corresponding similarities $(s_{j1}, s_{j2}, \ldots s_{jk})$ using cosine based similarity

$$\text{Similarity } S(i,j) = \frac{i * j}{i^2 + j^2}$$

The objective of Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme is to evaluate the web user navigation patterns and user references of product search. The overall process consists of five phases.

The first phase consists of initialization phase. The phase initializes the web object. Web objects are initialized where O denotes the set of web objects, hi denotes hit value of object i and bi denotes bandwidth object i. Let n denote the number of objects to be pre-fetched.

The second phase consists of web pre-fetching and caching phase. During this phase, pre-fetching is performed using hit rate and bandwidth rate and caching is performed using lifetime. The overall factor is evaluated using hit bandwidth rate. The next step is to evaluate the increase factor in order to determine the pre-fetched object and non pre-fetched object. The objects having largest increase factor and the objects that have maximum life time are determined as pre-fetched and cached objects.

The third phase involves with the cluster formation. During this phase Pre-fetched and cached documents obtained from web pre-fetching and caching phase are given as input to form cluster. Clusters are forming using the Euclidean distance. Based on the minimum distance the objects are assigned and accordingly the web objects clusters are formed.

The last phase consists of filtering phase and evaluation. Here cluster based web objects are given as input. Similarity items are evaluated which in turn discards web pages generated by unique users and also filter uncacheable requests. The output obtained is the filtered web object.

User Navigation of web object is evaluated from the web cluster objects with similarity retrieval in subsequent user sessions using cosine based similarity. Web Object Filters are built with the interpretation of the cluster web pages related to the unique users by discarding redundant pages.

4. Performance Evaluation on Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme

To evaluate the proposed Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme, we carried out experimentation that the proposed scheme progress the clustering result in terms of the values of objective functions. The result of the proposed Cluster based Web Object Filter scheme are compared with the previous methods and discussed in order to provide an imminent on their applicability and significance. The experimentation was conducted on the log file, resultant from the NASA Web server, maintains records over a one-month period, and its size is about 200 Mbytes.

Web Pre-fetching and Web caching scheme precedes the clustering process and involves data filtering, which filters any log entry that is having the redundant web object. Thus, the preliminary log entries have been significantly decreased so as to work and be helpful for the clustering information.

Experimental evaluation of proposed Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme is carried out using the java platform. The performance is measured in terms of Objective function, Number of clusters and cluster accuracy.

5. Result and Discussions on Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme

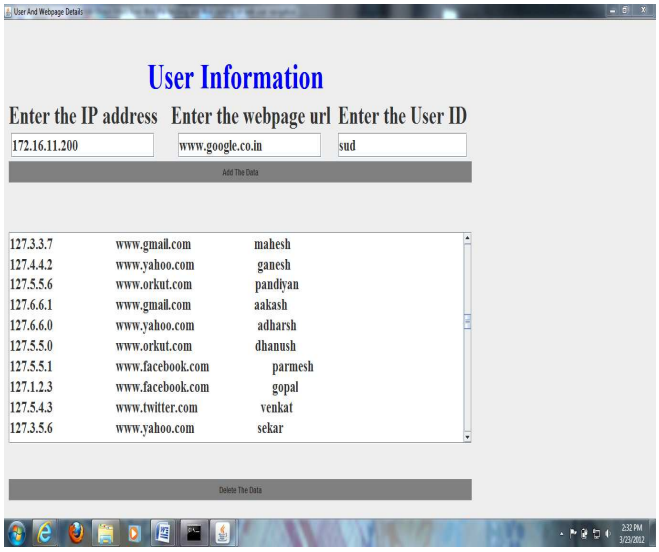

Fig: 2 User Log Data Information

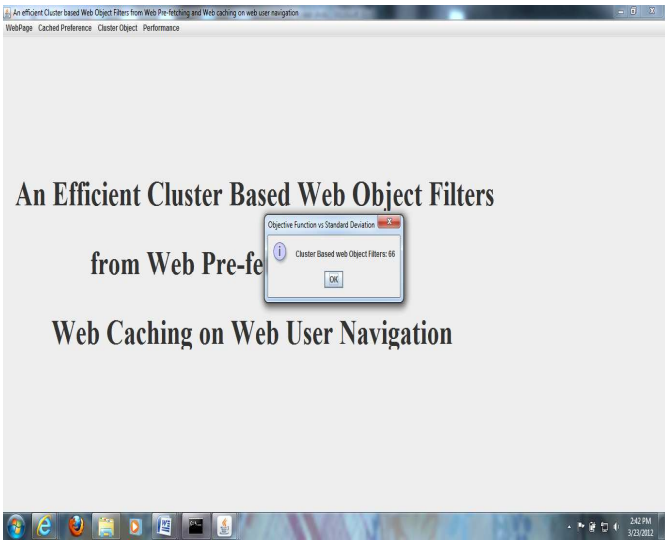

Fig:3 Experimental result for Number of Clusters vs Objective Function of the user in web object filters from pre-fetching and Web Caching.

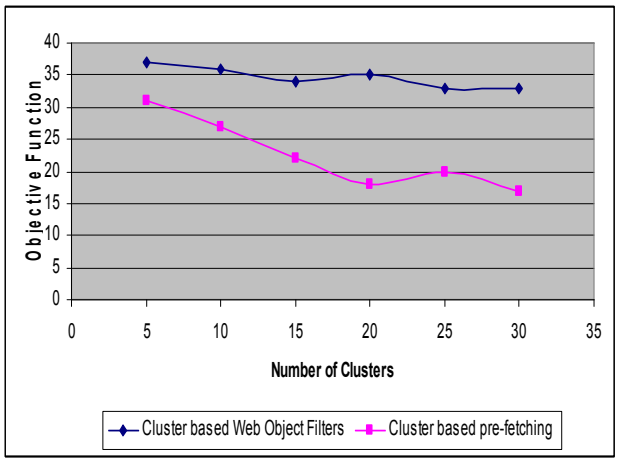

Fig 4: Number of clusters vs. Objective function

The improvements observed in Cluster based pre-fetching scheme are lower than the in the proposed Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme. This is demonstrated in Fig. 4, where the improvements of objective function are presented as a function of the number of clusters as 5, 10, 15, 20, 25 and 30.

The proposed Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme improved up to 30% compare to existing Cluster based pre-fetching scheme.

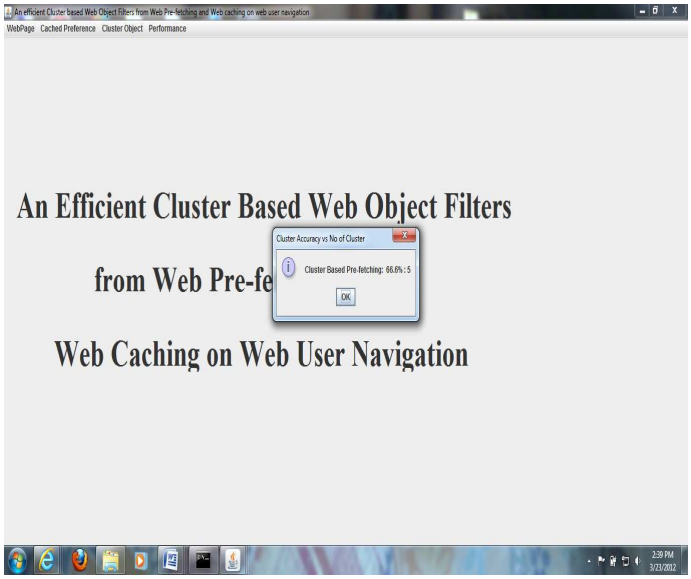

Fig: 5 Experimental results that the cluster accuracy 66.6 % in Web Pre-fetching and Web caching scheme.

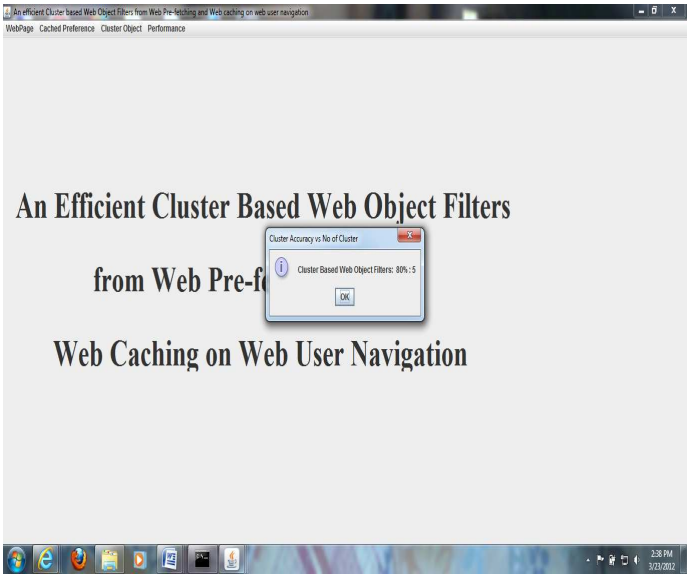

Fig:6 Experimental result that 80% cluster accuracy in Cluster Based Web Object Filters from Web Pre-fetching and Web Caching scheme.

From the result, we observe that the average performance on cluster accuracy of our proposed Cluster based Web Object Filters method is better than that of the previous Cluster based pre-fetching method. Here, proposed Cluster based Web Object Filters method achieves 80%, where as Cluster based pre-fetching scheme is 66.6%. Which is comparatively outstanding result because it can improve clustering robustness, deal with filtered web data objects and make use of multiple clustering criteria.

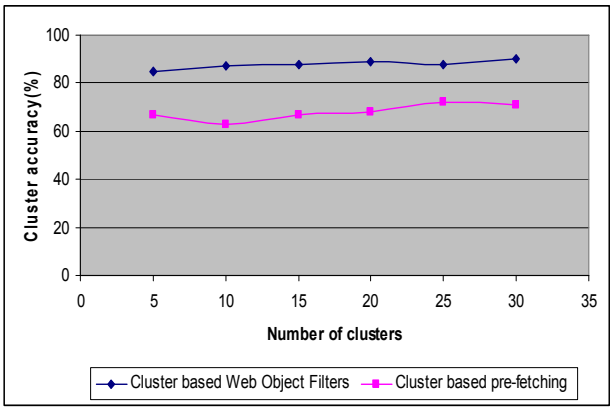

Fig 7: Cluster accuracy

6. Conclusion

In this paper, we have presented an efficient Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme to evaluate the web user navigation patterns and user references of product search. The proposed Web Pre-fetching and Web caching scheme efficiently integrates Web caching and pre-fetching contents. Clustering was done with the requested web page objects obtained from pre-fetched and web cached contents. Web Object Filters have built with the interpretation of the cluster web pages related to the unique users by discarding redundant pages. Cluster based Web Object Filter scheme evaluates the User Navigation from the web cluster objects with similarity retrieval in subsequent user sessions. Experimental evaluation of proposed Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme is carried out using the java platform. The performance is measured in terms of Objective function, Number of clusters and cluster accuracy which is efficient than existing model. This proposed method can be used for on line tutoring, video/audio conferencing and e-commerce applications.

References

[1] Teng WG, Chang CY, Chen MS. Integrating Web caching and Web Pre-fetching in client-side proxies. IEEE Trans Parallel Distributed Syst 2005;16(5):444–55.

[2] Pallis G, Vakali A, Sidiropoulos E. FRES-CAR: An adaptive cache replacement policy. In: Proceedings of the 1st IEEE international workshop on challenges in Web information retrieval and integration (WIRI'05) in cooperation with the 21st IEEE conference on data engineering ICDE 2005. Tokyo, Japan; 2005.

[3] Berendt B, Spiliopoulou M. Analysing navigation behaviour in web sites integrating multiple information systems. VLDB J Special Issue on Databases and the Web 2009(1):56–75.

[4] A. Bianco, G. Mardente, M. Mellia, M. Munafo, and L. Muscariello, "Web User Session Characterization via Clustering Techniques," Proc. IEEE GLOBECOM '05, p. 6, Dec. 2005.

[5] S. Petridou, V. Koutsonikola, A. Vakali, and G. Papadimitriou, "A Divergence-Oriented Approach for Web Users Clustering," Proc. Int'l Conf. Computational Science and Its Applications (ICCSA '06), pp. 1229-1238, May 2006.

[6] G. Pallis, L. Angelis, and A. Vakali, "Model-Based Cluster Analysis for Web Users Sessions," Proc. 15th Int'l Symp. Methodologies for Intelligent Systems (ISMIS '05), pp. 219-227, May 2005.

[7] Yue Suo, Naoki Miyata, Hiroki Morikawa, Toru Ishida, and Yuanchun Shi, "Open Smart Classroom: Extensible and Scalable Learning System in Smart Space

Using Web Service Technology", IEEE transactions on knowledge and data engineering, vol. 21, no. 6, june 2009.

[8] M. Albanese, A. Picariello, C. Sansone, and L. Sansone, "Web Personalization Based on Static Information and Dynamic User Behavior," Proc. Sixth ACM Int'l Workshop Web Information and Data Management (WIDM '04), pp. 80-87, Nov. 2004.

[9] P. Lingras and C. West, "Interval Set Clustering of Web Users with Rough-Means," J. Intelligent Information Systems, vol. 23, no. 1, pp. 5-16, 2004.

[10] Y. Zhao and G. Karypis, "Topic-Driven Clustering for Document Datasets," Proc. SIAM Int'l Conf. Data Mining, pp. 358-369, Apr. 2005.

[11] W. Yang, Z. Wang, and M. You, "An Improved Collaborative Filtering Method for Recommendations' Generation," Proc. Int'l Conf. Systems, Man and Cybernetics (SMC '04), pp. 4135-4139, Oct.2004.

[12] J. Xu, J. Liu, B. Li, and X. Jia, "Caching and Pre-fetching for WebContent Distribution," Computing in Science and Eng., vol. 6, no. 4, pp. 54-59, 2004.

A. K. Santra received the P. G. degree and Doctorate degree from I.I.T., Kharagpur in the year 1975 and 1981 respectively. He has got 20 years of Teaching Experience and 19 years of Industrial (Research) Experience. His area of interest includes Artificial Intelligence, Neural Networks, Process Modeling, Optimization and Control. He has got to his credit (i) 45 Technical Research Papers which are published in National / International Journals and Seminars of repute, (ii) 20 Research Projects have been completed in varied application areas, (iii) 2 Copy Rights for Software Development have been obtained in the area of Artificial Neural Networks (ANN) and (iv) he is the contributor of the book entitled **"Mathematics and its Applications in Industry and Business"**, Narosa Publishing House, **New Delhi.** He is the recognized Supervisor for guiding Ph. D. / M. S. (By Research) Scholars of Anna University-Chennai, Anna University-Coimbatore, Bharathiyar University, Coimbatore and Mother Teresa University, Kodaikanal. Currently he is guiding 12 Ph. D. Research Scholars in the Department. He is a Life member of CSI and a Life member of ISTE.

S. Jayasudha received her M. C. A., from Periyar University, Salem, M.Phil. from Bharathidasan University, Trichy. Currently she is working as Asst.Professor in Bannari Amman Institute of Technology, Sathyamangalam. Her area of interest includes Web Mining, Text Mining. She has presented one national and one International Conference paper. She has published a research paper in International Journal (IJCSI). She is a Life member of Computer Society of India and a Life member of Indian Society for Technical Education.