# Comprehensive Survey of Framework for Web Personalization using Web Mining

Vikas Verma
M.M.I.C.T & B.M.
M.M.University
Mullana
Haryana-133203, India

A.K.Verma
Dept. of Comp. Sc. and Engg.
TIET, Thapar University
Patiala
Punjab-147004, India

S.S.Bhatia
SMCA
Thapar University
Patiala
Punjab-147004, India

## ABSTRACT
World Wide Web is a global village and a rich source of information. The number of users accessing web sites is increasing day by day. For effective and efficient handling, web mining coupled with recommendation techniques provides personalized contents at the disposal of users. Web Mining is an area of Data Mining dealing with the extraction of interesting knowledge from the World Wide Web. While surfing the web sites, users' interactions with web sites are recorded in web usage file. These Web Logs when mined properly are rich source for Web Personalization. Mining of these Web Logs is referred to as Web Usage Mining. This paper presents a comprehensive survey of over 100 research papers dealing with Web Mining framework.

## General Terms
Data Mining, Framework.

## Keywords
Web Mining, Web Log Mining, Web Personalization.

## 1. INTRODUCTION
The expansion of the World Wide Web has resulted in a large amount of data that is now in general freely available for user access. The different types of data have to be managed and organized such that they can be accessed by different users efficiently [Han and Kamber (2001)]. Therefore, the application of data mining techniques on the Web is now the focus of an increasing number of researchers. Several data mining methods are used to discover the hidden information in the Web. However, Web mining does not only mean applying data mining techniques to the data stored in the Web. The algorithms have to be modified such that they better suit the demands of the Web. In accordance with [Kosala *et al.* (2002)], the term 'Web Mining' is defined as the whole of data mining and related techniques that are used to automatically discover and extract information from web documents and services [Cowie and Lehner (1996)]**.** Web mining research, is an integrate research from several research communities such as: Database (DB), Information retrieval (IR), The sub-areas of machine learning (ML) and Natural language processing (NLP) [Etzioni (1996)].

Web mining, can be categorized into three types viz., Content Mining, Structure Mining and Usage Mining. Web content mining can be described as the process of extracting knowledge from the content or descriptions of web documents. In web content mining there are two dominant groups of strategies: Web page content mining and Search result mining. Web content mining has to do with the retrieval of information (content) available on the web into more structured forms as well as its indexing for easy tracking information locations.

Web structure mining is the process of extracting information from the hyperlink structure within the web itself. On any given website, pages can be classified into two conceptual categories [Linoff and Berry (2001)]: Navigation pages and Destination pages. The goal of Web structure mining is to categorize the Web pages and generate information such as the similarity and relationship between them, taking advantage of their hyperlink topology.

An important constituent category of Web Mining is Web Log mining also known as Web Usage mining, is the process of extracting interesting patterns from web access logs [Zaiane (2001)]. Web Usage mining imitates the actions of humans as they interact with the Internet [Vellingiri and Pandian (2011)]. It can also be defined as the process of identifying browsing patterns by analyzing the user's navigational behavior. This information takes as input the usage data, i.e. the data residing in the Web server logs, recording the visits of the users to a Web site. Extensive research in the area of Web usage mining led to the appearance of a related research area, that of Web personalization. Web personalization utilizes the results produced after performing Web usage mining, in order to dynamically provide recommendations to each user [Kosala and Blockeel (2000)].

Web Personalization is defined as the process of customizing the content and structure of a web site to the specific and individual needs of each user taking advantage of the user's navigational behavior. The methods employed to analyze the collected data includes Content-based filtering, Collaborative filtering, Rule-based filtering and Web Usage Mining as depicted in Fig. 1. The constituent phases contributing to basic framework of Web Mining are: Data Collection, Data Preprocessing, Data Analysis, and Pattern Analysis [Chakrabarti (2000)].

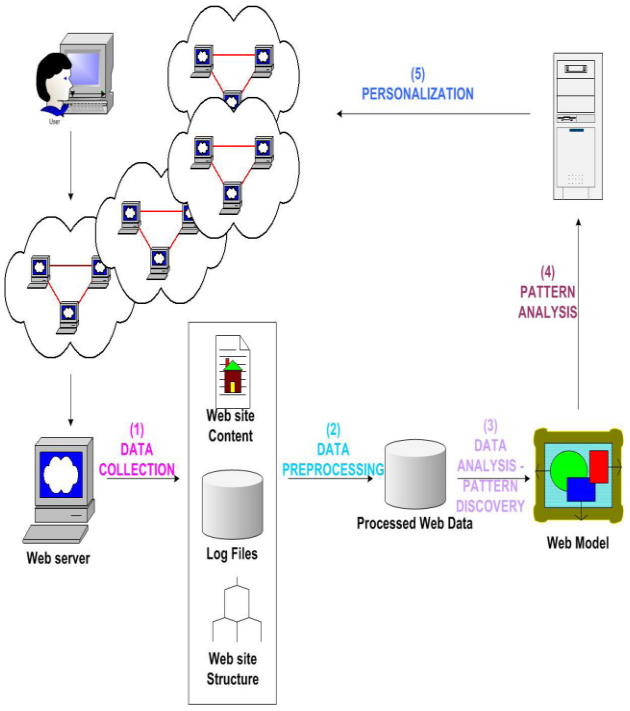


**Fig. 1: Web Personalization [Eirinaki and Vazirgiannis (2003)].**

## 2. LITERATURE REVIEW

It is worthwhile to mention here that, although the generalized framework of Web Personalization is shown above, yet different authors have given their contributions in various techniques / components as discussed below**.**

Most existing Web analysis tools provide mechanisms for reporting user activity in the servers and various forms of data filtering. Using such tools it is possible to determine the number of accesses to the server and to individual files, the times of visits, and the domain names and URLs of users. However, these tools are designed to handle low to moderate traffic servers, and usually provide little or no analysis of data relationships among the accessed files and directories within the Web space. More sophisticated systems and techniques for discovery and analysis of patterns are now emerging. For instance the WEBMINER system, developed by [Cooley *et al*. (1997)] introduces a general architecture for web usage mining. This system discovered association rules and sequential patterns from server access logs.

A lot of research work has been conducted on web personalization. This work was further extended by [Perkowitz *et al*. (1999)] for adaptive Sites use information about user access patterns to improve their organization and presentation for different types of users. [Mobasher *et al*. (1999, 2001)] formulated a technique for capturing common visitor profiles using association rule discovery and usage-based clustering of URLs. [Spiliopoulou *et al*. (2000)] developed a methodology to assess the quality of a web site based on the discovery and comparison of navigation patterns of customers and non-customers and also proposed a technique of dynamically adapting the site. An algorithm to reorganize a web site using page access frequency and classification of pages was proposed by [Fu *et al*. (2001)]. A tool was developed by [Masseglia *et al*. (1999)] for customizing the website dynamically. [Kitsuregawa *et al*. (2001)] performed association rule mining to discover interesting behavioral patterns of mobile device users. [Srikant and Yang (2001)] concluded that finding page locations that were different than where visitors expected them to be during their visit also helped to restructure website organization.

[Wu *et al*. (2004)] presented a novel integrated data warehousing and data mining framework for website management and pattern discovery to analyze web user behavior. This framework combines multidimensional Web databases to support online analytical processing for improving Web services.

[Fu and Shih (2002)] proposed a framework for Personal Web Usage Mining. In this framework Web Usage data was mined on client side as a complement to the server side usage data. By mining client side Web usage data, more complete knowledge about Web Usage was obtained.

[Jalali *et al.* (2007)] developed architecture for online predicting in Web Usage Mining System. This architecture of Web Usage Mining enhanced the accuracy of classification by interaction between classification, evaluation, current user activities and user profile in online phase.

[Yen *et al.* (2005)] framed an Efficient Incremental Algorithm. This algorithm uses the concept of Web traversal pattern mining by discovering most of users' access patterns from web logs. So, incremental mining utilizes the previous mining results and finds new patterns just from the inserted or deleted part of web logs such that mining time can be reduced.

[Petroumias *et al.* (2004)] framed a platform for web mining to refine browsing pattern. They used the concept of fuzzy sets while searching for the pages of interest. Another important characteristic that was addressed was the notion of time. Within this, one can identify the longest time periods within which frequencies of pages occur and also the periodicity with which these web pages are accessed. The framework also addresses several 'constraint-based' pre processing mining tasks to be performed prior to applying data mining algorithms to data collected from server logs. These constraints were taken from standard Web log files.

A click stream clustering based on the navigation behavior of users and the time spend at each page was taken into consideration for finding the number of clusters automatically through the utility of accelerated ant based clustering algorithm. This algorithm was framed by [Inbarani *et al*. (2006)].

An overview of using frequent pattern mining techniques for discovering different types of patterns in a web log database was presented by [Ivancsy and Vajk (2005 a)]. The three patterns searched were: Frequent itemsets, Sequences and Tree patterns, which was extended by [Ivancsy and Vajk (2005 b)].

The discovery of patterns from usage data by itself was not sufficient for performing the personalization tasks. The critical step was the effective derivation of good quality and useful (i.e.,

actionable) "aggregate usage profiles" from these patterns. [Mobasher *et al.* (2002)] experimentally evaluate two techniques, based on clustering of user transactions and clustering of pageviews, in order to discover overlapping aggregate profiles that can be effectively used by recommender systems for real-time Web personalization. These techniques were evaluated both in terms of the quality of the individual profiles generated, as well as in the context of providing recommendations as an integrated part of a personalization engine.

[Abraham (2003)] proposed a novel approach called "intelligent-miner" that could optimize the concurrent architecture of a fuzzy clustering algorithm to discover web data clusters and a fuzzy inference system to analyze the web site visitors' trends.

[Raju and Satyanarayana (2008)] proposed a complete preprocessing methodology, one of the important steps in Knowledge Discovery from Web Usage Data (KDWUD) process. Several heuristics have been proposed for cleaning the WUD which was then aggregated and recorded in the relational data model. Several experiments were conducted and the results showed that the proposed methodology reduces the size of Web access log files down to 73-82% of the initial size and offer richer logs that were structured for further stages of KDWUD.

WebTMS in Web was developed by [Mahanta (2008)] which adopts multi-agents architecture and combines text mining and multidimensional document analysis in order to help user in mining HTML documents on the Web effectively.

[Zaki *et al.* (2001)] proposed LOGML-Log Markup Language for Web Usage Mining wherein the web-log reports were presented from web log files and web graph. This proposed language showed the simplicity with which mining algorithms for extracting increasing complex frequent patterns were specified and implemented efficiently.

[Joshi *et al.* (2000)] showed how clustering can be done in case of discrepancy regarding two web sessions. A fuzzy clustering algorithm was developed that lead to the extraction of interesting user profiles.

[Wu *et al*. (2002)] represent user behaviors by sequences of consecutive web page accesses, derived from the access log of a proxy server. Frequent sequences were discovered and organized as an index. Based on the index, they proposed a scheme for predicting user requests and a proxy based framework for prefetching web page.

[Mobasher (1999)] presented a usage based Web personalization system, called WebPersonalizer, drawing heavily upon Web mining techniques, making the personalization process automatic, and dynamic. The system architecture separated the offline tasks of data preparation and Web usage mining, and the online recommendation engine. At the heart of the system was a technique based on clustering of user transactions which allowed for the discovery of effective aggregate usage profiles.

[Dai *et al.* (2002)] described an efficient framework for Web personalization based on sequential and non-sequential pattern discovery from usage data. Their experimental results performed on real usage data indicate that more restrictive patterns, such as contiguous sequential patterns (e.g., frequent navigational paths) were more suitable for predictive tasks, such as Web prefetching, which involved predicting which item was accessed next by a user), while less constrained patterns, such as frequent itemsets or general sequential patterns were more effective alternatives in the context of Web personalization and recommender systems.

[Borges *et al.* (2000)] proposed a data mining model that captured the user navigation behavior patterns. The user navigation sessions were modeled as a hypertext probabilistic grammar whose higher probability strings correspond to the user's preferred trails .In this algorithm they made use of the Ngram model which assumed that the last N pages browsed affected the probability of the next page to be visited.

[Patel *et al.* (2009)] uses Markov Model techniques for prediction of knowledge discovery in web. [Etminani *et al.* (2009)] proposed the discovery of the users' navigational patterns using Self Organized Maps by preprocessing Web logs for extracting the common patterns. [Zhang *et al.* (2009)] presented a Web usage mining technique based on fuzzy clustering in Identifying Target Group that share similar interests and behaviors by examining the data gathered in Web servers. [Nina *et al.* (2009)] suggests a complete idea for the pattern discovery of Web usage mining coupled with the Web warehousing. [Wu *et al*. (2010)] given a Web Usage Mining technique based on the sequences of clicking patterns in a grid computing environment. [Aghabozorgi *et al*. (2009)] proposed the usage of incremental fuzzy clustering to Web Usage Mining to produce a dynamic model from off-line model produced by fuzzy clustering.

[Maratea *et al.* (2009)] Personalized Web page recommendation is strictly restricted by the nature of web logs, the intrinsic complexity of the problem and the higher efficiency needs. The author designed a heuristic majority intelligence technique, which effortlessly adjusts to changing navigational patterns; with the low cost explicitly individuate them ahead of navigation.

[Inbarani *et al.* (2007)] proposed Rough set based feature selection for web usage mining involving different phases for Personalization. [Jalali *et al.* (2007)] put forth a web usage mining technique based on LCS algorithm for online predicting recommendation systems for categorizing user navigation behavior for forecasting users' future requests. [Shinde and Kulkarni (2008)] provides a architecture for online recommendation for predicting in Web Usage Mining System (OLRWMS) for improving the accuracy of classification by interaction between classifications, evaluation, and present user activates and user profile in online phase of this architecture.

[Zhang and Liang (2004)] give an intelligent algorithm of data pre-processing in Web usage mining called "USIA" was proposed and its merits and demerits were examined. The experimental evaluation of USIA indicates the better efficiency and also it determines the exact user and session. [Nasraoui *et al*. (2008)] provides a whole framework and findings in mining Web usage navigation from Web log files of a genuine Web site

which has every challenging characteristics of real-life by providing a technique for identifying and tracing growing user profiles.

[Hogo *et al.* (2003)] proposed the temporal Web usage mining of Web users on single educational Web site with the help of SOM based on rough set properties. [Dong (2009)] presented a SQL Server2000 based Web usage mining system and demonstrates with genuine instances on how to make use of DTS, T-SQL and other tools under SQL Server2000 to appreciate data transfer, data cleansing, user recognition, session recognition and further data pre-processing purpose, and demonstrates on how to utilize Online Analysis and Process (OLAP) and also the Data Mining (DM) under SQL Server2000 as to recognize mode discovery and mode analysis.

[Li *et al.* (2008)] developed a data preprocessing technique for Web usage mining using path completion algorithm. [Baraglia *et al.* (2002)] proposed a Web usage mining (WUM) system, called SUGGEST, which continuously creates the suggested connections to Web pages of probable importance for a user. [Chen *et al.* (2004)] uses the closure property of frequent item sets for the purpose of extracting the cross-transaction association rules from Web log databases.

A World Wide Web usage mining and examination tool called SpeedTracer, was created by [Wu *et al.* (1998)] in order to realize user browsing pattern by investigating the Web server log files with data mining procedures.

[Labroche *et al.* (2007)] proposed a novel technique for web usage mining and visualization that is based on the bio-mimetic relational clustering technique Leader Ant and the description of prototypes based on typicality computation for producing an efficient visualization of the activity of users on a website. The simulation result illustrates that it can effortlessly construct meaningful visualizations of typical user browsing patterns.

[Lee *et al.* (2008)] put forth a Web Usage Mining technique based on clustering of browsing characteristics for an E-commerce application using hierarchical agglomerative clustering to cluster users' navigation patterns. [Tzekou *et al.* (2007)] gives an effective site customization technique based on Web Semantics and Usage Mining for efficiently identifying the user likings that are secreted behind user browsing behavior and new recommendation technique is proposed that utilizes Web mining techniques for correlating the recognized importance to the sites semantic content, for the purpose of modifying them to certain users.

[Wang *et al.* (2004)] provides a technique that can discover users' frequent browsing patterns underlying users' browsing Web behaviors using an algorithm (FAP-Mining) based on the FP-tree technique for mining the common access patterns. [Adda *et al. (2007)*] presents a technique with the help of metadata about the content that they imagine is stored in domain ontology. This technique includes a dedicated pattern space constructed on top of the ontology, navigation primitives, mining procedure and recommendation methods.

[Suneetha and Krishnamoorthi (2009)] presents an algorithm for data preprocessing particularly regarding data cleaning user identification and session identification**.** [Dixit and Kiruthika (2010)] also presents preprocessing techniques based on XML and text file**.** [Bayir *et al.* (2006)] presents some heuristics like time and navigation along with site topology for evaluating actual sessions in pre processing phase for Web Personalization.

[Zhang *et al.* (2010)] presents comparison differences between the web server log and the enterprise proxy log, and then an incremental data cleaning algorithm based on the differences using a clustering algorithm involving hierarchical URL similarity. [Wang and She (2009)] presents some novel algorithms to mine users' interests. The algorithms are based on visit time and visit density which can be obtained from an analysis of web users' Web Log Data for finding user's interested domains.

## 3. CONCLUSIONS AND SCOPE

In this paper we have presented a survey of over 100 research papers analyzing framework / techniques / phases for Web mining, the application of data mining and Web Personalization techniques. A change in any of the constituent technique for Web Mining accounts for a change in paradigm consequently affecting the approach of web analyzer to examine the data for identifying the navigation pattern. So a good understanding of the data preparation technique and pattern discovery method is required.

Web usage mining in particular has an immense potential in itself for recommendations to users based on their preferences. This field has motivated various researchers to provide their ideas and thus rendering their services for "user-centric" approach. There are several techniques proposed by different researchers for the web usage mining in particular with its own merits and demerits.

The future work involves the development of autonomous agents that analyzes the discovered rules to provide meaningful courses of action or suggestions to users. Future scope of Web Mining includes predicting user needs in order to improve the usability, scalability, user retention, and framing an efficient framework for Web Personalization through efficient harnessing Web Log file.

## 4. REFERENCES

[1] Abraham, A. 2003. Business Intelligence from Web Usage Mining, Journal of Information & Knowledge Management, 2(4), 375-390.

[2] Adda, M., Valtchev, P., Missaoui, R. and Djeraba, C. 2007. Toward Recommendation Based on Ontology-Powered Web-Usage Mining, IEEE Internet Computing, 11(4), 45-52.

[3] Aghabozorgi, S. R. and Wah, T.Y. 2009. Using Incremental Fuzzy Clustering to Web Usage Mining, International Conference of Soft Computing and Pattern Recognition, 653-658.

[4] Baraglia, R. and Palmerini, P. 2002. SUGGEST: a Web usage mining system, International Conference on Information Technology: Coding and Computing, 282-287.

[5] Bayir, A. M., Toroslu, H. I. and Cosar, A. 2006. A New Approach for Reactive Web Usage Data Processing, Proceedings of the 22nd International Conference on Data Engineering Workshops, 44.

[6] Borges, J. and Levene, M. 2000. Data Mining of User Navigation Patterns, 92-111.

[7] Chakrabarti, S. 2000. Data Mining for Hypertext: A Tutorial Survey, SIGKDD Explorations, Knowledge Discovery & Data Mining, 1(2), 1–11.

[8] Chen, J., Yin, J., Tung, A. K. H. and Liu, B. 2004. Discovering Web usage patterns by mining cross-transaction association rules, International Conference on Machine Learning and Cybernetics, 5, 2655-2660.

[9] Cooley, R., Srivastava, J. and Mobasher, B. 1997. Web Mining: Information and Pattern Discovery on the WorldWideWeb, ICTAI'97.

[10] Cowie and Lehner, W. 1996. Information Extraction, ACM, 39(1), 80–91.

[11] Dai, H., Mobasher, B., Luo, T. and Nakagawa, M. 2002. Using Sequential and Non-Sequential Patterns in Predictive Web Usage Mining Tasks, Proceedings of the 2002 IEEE International Conference on Data Mining , 669-672.

[12] Dixit, D. and Kiruthika, M. 2010. Preprocessing of Web Logs, International Journal on Computer Science and Engineering , 2(7), 2447-2452.

[13] Dong, D. M. 2009. Exploration on Web Usage Mining and its Application, International Workshop on Intelligent Systems and Applications, 1-4.

[14] Eirinaki, M and Vazirgiannis, M. 2003. Web Mining for Web Personalization, ACM Transactions on Internet Technologies (ACM TOIT), 3(1), 1 – 27.

[15] Etminani, K., Delui, A. R., Yanehsari, N. R. and Rouhani, M. 2009. Web Usage Mining: Discovery of the Users' Navigational Patterns Using SOM, First International Conference on Networked Digital Technologies, 224-249.

[16] Etzioni 1996. The World-Wide Web: Quagmire or gold mine? ACM, 39(11), 65–68.

[17] Fu, Y., Creado, M. and Ju, C. 2001. Reorganizing websites based on user access patterns, ACM CIKM, 583-585.

[18] Fu, Y. and Shih, Y. M. 2002. A Framework for Personal Web Usage Mining, Proceedings of International Conference on Internet Computing , 595-600.

[19] Han, J. and Kamber, M. 2001. Data Mining: Concepts and techniques, 1st edn. Morgan Kaufmann.

[20] Hogo, M., Snorek, M. and Lingras, P. 2003. Temporal Web usage mining, International Conference on Web Intelligence, 450-453.

[21] Inbarani, H. H. and Thangavel, T. 2006. Clickstream Intelligent Clustering Using Accelerated Ant Colony Algorithm, ADCOM , 129-134.

[22] Inbarani, H. H., Thangavel, K. and Pethalakshmi, A . 2007. Rough Set Based Feature Selection for Web Usage Mining, International Conference on Conference on Computational Intelligence and Multimedia Applications, 1, 33-38.

[23] Iváncsy, R., and Vajk, I. 2005 a. Efficient Sequential Pattern Mining Algorithms, WSEAS Transactions on Computers, 4(2), 96-101.

[24] Iváncsy, R., and Vajk, I. 2005 b. PD-Tree: A New Approach to Subtree Discovery, WSEAS Transactions on Information Science and Applications, 2(11), 1772-1779.

[25] Jalali, M., Mustapha, N., Sulaiman, N. M., and Mamat, A. 2007. An Architecture for Online Predicting in Web Usage Mining System, The 2nd National Intelligent Systems And Information Technology Symposium, ISITS'07, ITMA - UPM, Malaysia, 33-38.

[26] Jalali, M., Mustapha, N., Sulaiman, N. B. and Mamat, A. 2008). A Web Usage Mining Approach Based on LCS Algorithm in Online Predicting Recommendation Systems, 12th International Conference Information Visualisation, 302-307.

[27] Joshi, A. and Krishnapuram, R. 2000. On Mining Web Access Logs, ACM SIGMOD, 63-69.

[28] Kitsuregawa, M., Shintani, T. and Pramudiono, I. 2001. Web mining and its SQL based parallel execution, IEEE Workshop on Information Technology for Virtual Enterprises, 128.

[29] Kosala, R., Blockeel, H. and Neven, F. 2002. An overview of Web Mining, J. Meij, Editor, Deling with Data Flood: Mining Data, Text and Multimedia, 480-497.

[30] Kosala, R. and Blockeel, H. 2000. Web Mining Research: A Survey, SIGKDD Explorations, 2(1), 1-15.

[31] Labroche, N., Lesot, M. J. and Yaffi, L. 2007. A New Web Usage Mining and Visualization Tool, 19th IEEE International Conference on Tools with Artificial Intelligence, 1, 321-328.

[32] Lee, H. C. and Fu, H. Y. 2008. Web Usage Mining Based on Clustering of Browsing Features, Eighth International Conference on Intelligent Systems Design and Applications, 1, 281-286.

[33] Li, Y., Feng, B. and Mao, Q. 2008. Research on Path Completion Technique in Web Usage Mining, International Symposium on Computer Science and Computational Technology, 1, 554-559.

[34] Linoff, G. S. and Berry, A. M. 2001. Mining the Web, 1st edn. John Wiley and Sons.

[35] Mahanta, N.A. 2008. Web Mining: Application of Data Mining, NCKM, 111-116.

[36] Maratea, A. and Petrosino, A. 2009. An Heuristic Approach to Page Recommendation in Web Usage Mining, Ninth International Conference on Intelligent Systems Design and Applications, 1043-1048.

[37] Masseglia, F., Poncelet, P. and Teisseire, M. 1999. Using data mining techniques on web access logs to dynamically improve hypertext structure, ACM SigWeb Letters, 8(3), 13-19.

[38] Mobasher, B. 1999. WebPersonalizer: A Server-Side Recommender System Based on Web Usage Mining, 9th Workshop on Information Technologies and Systems, 2003-2004.

[39] Mobasher, B., Cooley, R. and Srivastava, J. 1999. Creating adaptive web sites through usage based clustering of URLs, KDEX'99, 32-37.

[40] Mobasher, B., Dai, H., Luo,T. and Nakagawa, M. 2001. Elective personalization based on association rule discovery from web usage data, WIDM01, Atlanta, 9-15.

[41] Mobasher, B., Dai, H., Luo, T. and Nakagawa, M. 2002. Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization, Journal of Data Mining and Knowledge Discovery, 6, 61-82.

[42] Nasraoui, O., Soliman, M., Saka, E., Badia, A. and Germain, R. 2008. A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites, IEEE Transactions on Knowledge and Data Engineering, 20(2), 202-215.

[43] Nina, S. P., Rahman, M., Bhuiyan, K. I. and Ahmed, K. 2009. Pattern Discovery of Web Usage Mining, International Conference on Computer Technology and Development, 1, 499-503.

[44] Patel, K. R., Patel, S. A., Patel, R. N. 2009. Markov Model Analysis: Knowledge Discovery in Web, PIMT Journal of Research.

[45] Perkowitz, M. and Etzioni, O. 1999. Towards Adaptive Web sites: Conceptual framework and case study, Artificial Intelligence, 118, 245–275.

[46] Petrounias, I., Tseng, A., Kolev, B., Chountas, P. and Kodogiannis, V. 2004. An Intuitionistic Fuzzy Component Based Approach for Identifying Web Usage Patterns, Second IEEE International Conference on Intelligent Systems.

[47] Raju, T. G., Satyanarayana, S. P. 2008. Knowledge Discovery from Web Usage Data: Complete Preprocessing Methodology, IJCSNS, International Journal of Computer Science and Network Security, 8(1).

[48] Shinde, S. K. and Kulkarni, U. V. 2008. A New Approach for on Line Recommender System in Web Usage Mining, International Conference on Advanced Computer Theory and Engineering, 973- 977.

[49] Spiliopoulou, M., Pohle, C., and Faulstich, C. L. 2000. Improving the effectiveness of a website with web usage mining, In Advances in Web Usage Analysis and User Profiling, Berlin, Springer, 14162, 2000.

[50] Srikant, R., Yang, Y. 2001. Mining web logs to improve website organization, The 10th International World Wide Web Conference, Hong Kong, 430-437.

[51] Suneetha K. R. and Krishnamoorthi, R. 2009. Data Preprocessing and Easy Access Retrieval of Data through Data Ware House, World Congress on Engineering and Computer Science, 1, 306-311.

[52] Tzekou, P., Stamou, S., Kozanidis, L. and Zotos, N. 2007. Effective Site Customization Based on Web Semantics and Usage Mining, Third International IEEE Conference on Signal-Image Technologies and Internet-Based System, 51-59.

[53] Vellingiri, J. and Pandian, C. S. 2011. A Survey on Web Usage Mining, Global Journal of Computer Science and Technology, 11(4), 67-72.

[54] Wang, S. and She, L. 2009. Algorithm Research on User Interests Extracting via Web Log Data, International Conference on Web Information Systems and Mining, 93-97.

[55] Wang, X., Ouyang, Y., Hu, X. and Zhang, Y. 2004. Discovery of user frequent access patterns on Web usage mining, The 8th International Conference on Computer Supported Cooperative Work in Design, 1, 765-769.

[56] Wu, H. C., Wu, L.Y., Chang, M.Y. and Hung, H.M. 2010: Web Usage Mining on the Sequences of Clicking Patterns in a Grid Computing Environment, ICMLC, 6, 2909-2914.

[57] Wu, H. Y. and Chen, P. L. A. 2002. Prediction of Web Page Accesses by Proxy Server Log, World Wide Web: Internet and Web Information Systems, 5, 67–88.

[58] Wu, H. E., Ng, K. M. and Huang, Z.J. 2004. A Data Warehousing and Data Mining Framework for Web Usage Management, Communications in Information and Systems International Press, 4(4), 301-324.

[59] Wu, K. L., Yu, P. S. and Ballman, A. 1998. SpeedTracer: A Web usage mining and analysis tool, IBM Systems Journal, 37(1), 89-105.

[60] Yen, J. S., Lee, S. Y. and Hsieh, C. M. 2005. An Efficient Incremental Algorithm for Mining Web Traversal Patterns, ICEBE, 274-281.

[61] Zaıane, R. 2001. Building Virtual Web Views, Data and Knowledge Engineering, 39, 143–163.

[62] Zaki, J. M., Punin, R. J., Krishnamoorthy, S. M. 2001. LOGML: Log Markup Language for Web Usage Mining, WEBKDD, 88-112.

[63] Zhang, J., Zhao, P., Shang, L. and Wang, L. 2009: Web Usage Mining Based On Fuzzy Clustering in Identifying Target Group, International Colloquium on Computing, Communication, Control, and Management, 4, 209-212.

[64] Zhang, H. and Liang, W. 2004. An intelligent algorithm of data pre-processing in Web usage mining, Fifth World Congress on Intelligent Control and Automation, 4, 3119-3123.

[65] Zhang, Y., Lidai, Zhou, J. Z. 2010. A New Perspective of Web Usage Mining: Using Enterprise Proxy Log, International Conference on Web Information Systems and Mining, 38-42.