Web Usage Mining Tools & Techniques: A Survey

Kamika Chaudhary, Santosh Kumar Gupta

Abstract— Previous decade has proved itself to be a witness of day to day inventions and discoveries that leads to amelioration of various technologies. As a result World Wide Web has emerged and become information superhighway. The availability of information becomes just a click away. As time passed World Wide Web has grown into myriad of information and it became difficult to extract the data according to the requirement. Web Mining came as a boon to provide solution of above problem. Web Usage Mining is a category of Web Mining. Web Usage Mining mainly deals with discovery and analyzing of usage patterns in order to serve the needs of web based applications. The process of Web Usage Mining mainly consists of three inter-dependent stages: data preprocessing, pattern discovery and pattern analysis. This paper is concerned with the study of different tools and techniques for Web Usage Mining. This survey concentrates in usage mining area and its applications in various fields.

Index Terms— Data Preprocessing, Pattern Analysis, Pattern Discovery, Web Usage Mining.

1 INTRODUCTION

With the growth of technology at a faster pace, World Wide Web has also grown exponentially. It has spread its arms along the entire world and being used in every field of day to day life. WWW became ubiquitous and acts as a tool to collect, share and disseminate information at any place and at any time. With the increase in popularity of web, its complexity also increases because of the presence of large amount of data. Web turns into huge repository of information and it became difficult for the users to retrieve the information of his interest from such a huge nugget. Getting the information from web has become like taking a drink from fire hydrant. The Web Mining technique acts as a device to bring out of this trouble. It helps in automatic discovery and retrieval of information from the internet [14]. Web Mining is divided into three categories: Web Content mining, Web Structure Mining and Web Usage Mining [2]. Extracting useful information from the structured or unstructured contents of web document is described in Web Content Mining [3]. Web Structure Mining is performed on hyperlink structure. It focuses on improving the structural design of the website. In Web Usage Mining, mining is performed on the usage pattern of the data which is stored in web log. It mainly deals with getting information about the navigational pattern of the user.

The structure of the paper is as follows: Section 2 describe about Web Mining Overview, Web Usage Mining, its procedures and various techniques used. Section 3 deals with the literature survey and gives a brief of the recent researches done in the field of web usage mining. Section 4 enlightens the privacy issues related to web usage mining, section 5 gives the description of web usage mining tools and their features, section 6 and 7 describes the conclusion and future scope of the Field and section 7 is about the references used.

- *Kamika Chaudhary, Department of Computer Science & Engineering, Krishna Institute of Engineering & Technology, Ghaziabad-201206, E-mail: kamika.agrohi@gmail.com*
- *Santosh Kumar Gupta, Department of Computer Science & Engineering, Krishna Institute of Engineering & Technology, Ghaziabad-201206, E-mail: santosh_k25@rediffmail.com*

2 WEB MINING

2.1 Overview

Web Mining [14] is defined as the application area of data mining which deals with the extraction of useful and interesting facts from World Wide Web. It is used to provide the solution of various problems such as finding relevant information, creating information from the data available on web, learning about consumers or individual users, personalization of information and so on. Web mining makes use of various data mining techniques to automatically discover web and retrieve information from the web documents [4]. It is mainly decomposed into four subtasks namely: resource finding, information selection and preprocessing, generalization and discovery of patterns and pattern analysis. Web Mining uses several data mining techniques to retrieve the useful facts from internet. However, along with Data mining techniques various other techniques such as artificial intelligence, information retrieval, natural language processing, information extraction, machine learning can also be applied. All these techniques are compared with Data mining. Most often Web Mining is associated with information retrieval or information extraction. Information retrieval works by indexing text and searching for useful document. It ended up in retrieving all relevant documents and some irrelevant also [5].

. Transformation of a collection of documents into information is done by Information Extraction with the help of Information Retrieval. Main aim of Information Extraction is to extract relevant facts from documents while that of Information Retrieval is to select relevant documents [6].Web Mining is a part of Information retrieval and Information Extraction system. Information extraction acts a preprocessing stage in the process of Web mining and is used for indexing text which is a part of Information Retrieval process. IE can be performed on both unstructured or semi structured text [7]. Web mining shares a close relationship between machine learning techniques. They can improve text classification process in a much better way as compared to IR techniques [8].

Web mining is categorized into three areas of interest: Web Content Mining, Web Structure Mining and Web Usage Mining [2].Web Content mining deals with discovery of useful information from unstructured, semi structured or structured

contents of web documents. Text, images, audio, video comprised by unstructured document, semi structured data includes HTML documents and lists and tables represent structured documents. The main aim of web content mining is to act as tool to retrieve information easily and quickly. Web Content Mining works by organizing a group of documents into related categories which helps web search engine to extract information more quickly and efficiently. Web Structure Mining [9], [10], [11], [12] mines the information by utilizing the link structure of the web documents. It works on inter document level and discovers hyperlink structure. It helps in describing the similarities and relationships between sites. Web structure mining categorizes the pages into authorities and hubs. Authority pages are considered as high quality pages which are related to particular query and hub pages provide pointers to authority pages. Web structure mining is further divided into two categories hyperlinks and documents. Web Usage Mining [13] is a data mining technique that mines the information by analyzing the log files that contains the user access patterns. Web Usage Mining mines the secondary data which is present in log files and derived from the interactions of the users with the web. Web usage Mining techniques are applied on the data present in web server logs, browser logs, cookies, user profiles, bookmarks, mouse clicks etc. The main focus of WUM is on techniques that are able to predict user behavior while user interacts with the web. Practically three web mining techniques can be used in isolation or together in an application depending upon the requirements and helps to overcome the problem of information overload on the web.

2.2 WEB USAGE MINING

Web Usage Mining concentrates on the techniques that could predict the navigational pattern of the user while the user interacts with the web. It is mainly divided into two categories, they are general access pattern tracking and customized usage tracking. In general access pattern tracking information is discovered by using the history of web page visited by user while in customized usage tracking mining is targeted on specific user. Mainly there are four types of data sources present in which usage data is recorded at different levels they are: client level collection, browser level collection, server level collection and proxy level collection as shown in Fig 1.

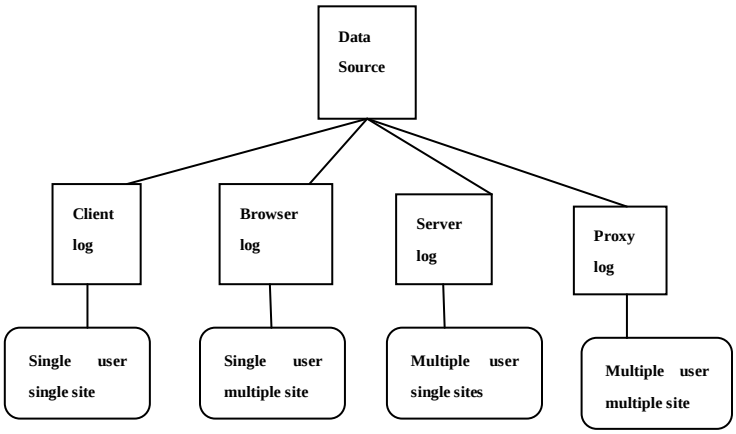

Fig 1: Various Data sources for Web Usage Mining

2.2.1 Client Level collection: At this level data is gathered together by means of java scripts or java applets. This data shows the behavior of a single user on single site. Client side data collection requires user participation for enabling java scripts or java applets. The advantage of data collection at client side is that it can capture all clicks including pressing of back or reload button [1].

2.2.2 Browser Level Collection: Second method of data collection is by modifying the browser. It shows the behavior of single user over multiple sites. The data collection capabilities are enhanced by modifying the source code of existing browser. They provide much more versatile data as they consider the behavior of single user on multiple sites [1].

2.2.3 Server Level Collection: Web server log stores the behavior of multiple users over single site. These log files can be stored in common log format or extended log format. Server logs are not able to store cached page views. Another technique used for usage data collection at server level is TCP/IP packet sniffing. Packet sniffers works by monitoring the network traffic and retrieve usage data directly.

2.2.4 Proxy Level Collection: Proxy servers are used by internet service provider to provide World Wide Web access to customers. These server stores the behavior of multiple user at multiple site. These server functions like cache server and they are able to produce cached page views.By predicting the usage pattern of the visitor Web Usage Mining improves the quality of e- commerce services, personalizes the web [15] or enhances the performance of web structure and web server.

2.3 Web Usage Mining Procedure and Techniques

The overall procedure of Web Usage Minin

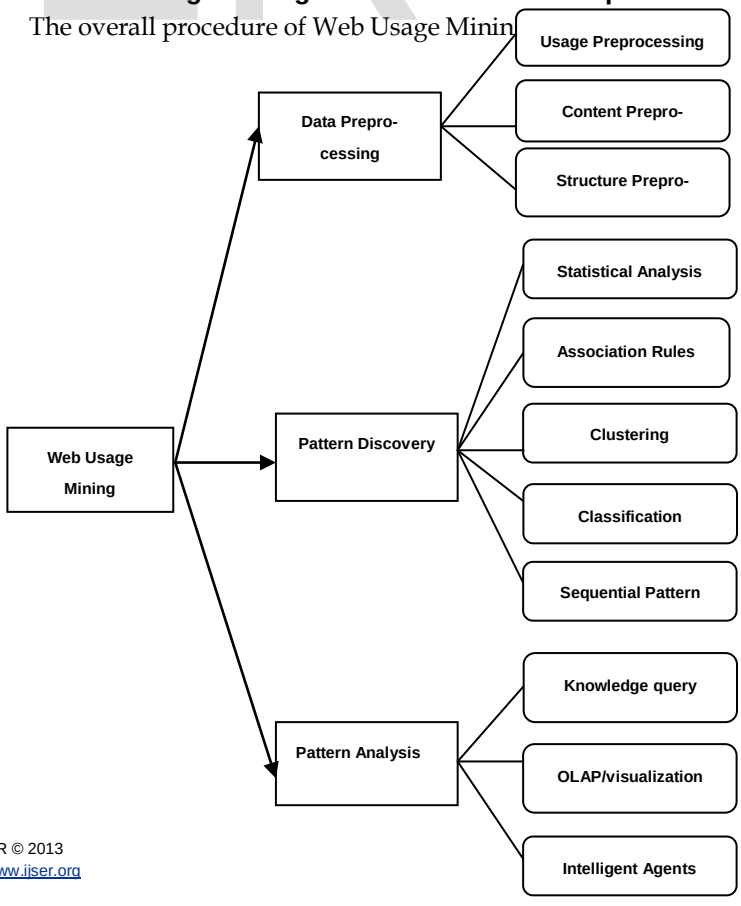

Fig 2: Web Usage Mining Procedures and techniques

three main steps: Data preprocessing, pattern discovery and pattern analysis. This section presents an overview of each step and techniques used in them as shown in Fig. 2.

2.3.1 Data Preprocessing

The first step of Web Usage Mining is preprocessing of data stored in web logs as it is noisy in nature. The process of converting usage, content and structure information into data abstraction is described in preprocessing. The processing of preprocessing consists of four phases: data cleaning, session reconstruction, content and structure information retrieval and data abstraction. Raw data is cleaned by removing all noisy, redundant and useless data in first step. Reference to style files, graphics or sounds is also removed. Second step deals with identification of user sessions from cleaned data and then reconstructing the user sessions. This step suffers from the problem of caching performed on proxy server which disallows the IP address to be used as user identification. This problem is rectified by making use of cookies [22]. Third step deals with the use of content based and structure based information to enrich the web log data. Final step of data preprocessing is concerned with the development of various abstractions which can be in the form of sessions, episodes, click streams and page views. Pattern discovery techniques are applied on these abstractions. Data Preprocessing is categorized into three types: usage preprocessing, content preprocessing and structure preprocessing.

Usage Preprocessing

This is considered as most difficult task of web usage mining because of presence of incomplete and inconsistent data in server log. Only IP address, agent and server side click stream are available to identify users and server sessions which faces many problems like single IP address/multiple server sessions, multiple IP address/single server session, multiple IP address/single user and multiple agent/single user. Usage preprocessing also encountered the problem of inferring cached page references.

Content Preprocessing

Content preprocessing concerned with transforming unstructured and semi structured documents into the forms that are suitable for web usage mining. It is used for limiting the discovered patterns for web usage mining. Vector space model [16] is applied on page views in order to convert them into suitable format. After proper formatting content mining algorithm are applied. Preprocessing of the content of each page view is performed either by HTTP request or from a combination of template, script and database access [1].

Structure Preprocessing

Structure of page views includes the hyper links. Structure preprocessing is performed similarly as content preprocessing is performed. In case of dynamic content a different site structure is developed for each session.

2.3.2 Pattern Discovery

Pattern discovery focuses on to uncover patterns from the abstractions produced as a result of preprocessing phase. It focuses on applying various methods and techniques developed from several fields such as data mining, machine learning, statistics and pattern recognition. Discovery of desired patterns and to extract understandable knowledge from them is a challenging task. This section explains some of algorithms suitable for pattern discovery.

Statistical Analysis

Statistical analysis focuses on analyzing the data which is aggregated by predetermined units such as days, sessions, visitors in order to gain knowledge about user behavior. Mainly three types of statistical analysis(frequency, mean, median) is performed on sessions and result of analysis shows most frequently accessed pages, average view time or length of the page. These results become helpful in improving system performance or enhancing the security [1].

Association Rules

Association rules are used for finding the correlations among web pages that frequently appear together in a user browsing session. Apriori algorithm [17] is the most popular algorithm that expresses the frequent co-occurrence of web pages together. These rules help in providing the recommendations to user as they visit the sites and web designers to restructure web sites. These rules show that if page M is accessed by a user then it is very likely that he visits page N also in same session. Algorithms used for association rules include maximal forward references, Markov Chains, FP growth and prefix span.

Clustering

The main purpose of clustering in web usage mining is to aggregate the similar sessions together [18], [19]. Self organized maps, graph partitioning, ant based technique, K-means with genetic algorithms, EM-CFuzzy means algorithm are the algorithms used for clustering the sessions. Mainly clustering is of two types usage clustering and page clustering. Usage clusters discover user that have same browsing pattern where page cluster gathers the content related pages together. This is performed by transforming the session into vectors of n elements where n may be no of pages or page views. Afterwards distance measure is used to find out the similarity between sessions. Clustering helps in personalizing the web site as it identifies the users with similar behavior.

Classification

Classification is supervised way of learning which maps the data items to one of many predefined classes [20].

Various supervised learning algorithms used for doing classification are decision trees, naïve Bayesian classifiers, k-nearest neighbor classifiers and support vector machine. Classification

mainly performs the automatic categorization of documents. In web usage mining, application of classification algorithm on server logs may lead to detection of interesting patterns such as 40% of users who visits news site are in the age group of 30-35 years.

Sequential Patterns

In web usage mining, sequential pattern is used to discover the sessions that are found in a sequence. They include the sequence of items that frequently occur in a particular order. MIDAS (Mining Internet Data for Association Sequences) algorithm is most commonly used for finding sequential patterns which provide marketing intelligent behavior for e-commerce scenario [21].

2.3.3 Pattern Analysis

The last step of web usage mining process is pattern analysis. This phase separates the interesting and uninteresting patterns from the overall patterns discovered during pattern discovery phase. Result of pattern analysis phase is used in various applications such as system performance improvement, site modification, personalization, e-commerce etc. Patterns can be analyzed by using following techniques described below:

Knowledge Query Mechanism

Structured Query Language (SQL) is most commonly used language for knowledge query mechanism. This language is applied in order to extract the useful patterns from discovered patterns.

OLAP/Visualization tools

Patterns are also analyzed by using OLAP tools in which discovered facts are placed on to data cubes for performing various OLAP operations such as roll up and drill down and interesting facts are retrieved. OLAP provides an integrated framework for analysis which allows changes in aggregate levels. Output of OLAP queries acts as an input to data mining or data visualization tools. Graphing patterns or assigning colors to different values are used as visualization technique for same purpose.

Intelligent Agents

Various agents are also devised that helps in examining the patterns in web usage mining. These agents perform the work of analyzing the discovered patterns.

3. LITERATURE WORK

Web usage mining system is divided into two tiers of tracking and analysis in order to find out the user access pattern. The resultant of web usage mining which is knowledge about user access pattern is useful in various applications. Due to information overload on the web its full economic potential has not been realized completely. Web personalization is introduced as a solution to this problem in [23]. A complete framework for web usage mining and FM model is introduced in order to analyze the access pattern of user. The framework has been capable in doing implicit tracking which handles sparsity problem in web personalization and capturing short term changes in user behavior. Web usage mining is being used in various spheres. In [24] web usage mining is used for improving the scalability and answer time of search engines. A simple storage model is proposed in which main memory is assigned with the function of doing dynamic caching of the answers and of inverted lists that are available in secondary memory.

The possibility of web usage mining on proxy servers is discussed in [25]. Collection of data from server log does not include the page views generated as a result of cache. Because of this, behavior of multiple users can be analyzed only on a single site. In order to analyze the access pattern of multiple users on multiple site proxy server log data has been used. Data from several different sectors such as TV, radio, newspaper and finance has been analyzed and compared by making use of e-metrics. The result shows that proxy servers contain a large amount of interesting information which has been used for finding the navigational pattern of users. In [26] pattern recognition strategy has been used for developing web personalization based on web usage mining. This strategy helps in analyzing both static and dynamic features. This strategy comprises of two phases: in first phase unsupervised clustering algorithms has been used for performing pattern analysis and classification. In second phase reclassification is performed to overcome the inaccuracy of registration information.

A framework for mining of client side web usage data called personal web usage mining is proposed in [31]. Author considers client side data because it will provide individualism and personalization of data whether server side considers a group of data for performing web usage technique. Another problem was that it is also possible that a group of user do not have same interest so server side mining proves to be inefficient in this respect. The framework includes four modules: logging, data warehousing, data mining and tool application. Logging stores the activities of user in activity recorder and cache pages. The logs are cleansed, transformed and stored in the data warehouse. In data mining useful information is extracted which is used in tool application.

In [32] author proposes a new technique of web usage mining called self organized map(SOM), a type of artificial neural network for determining the access patterns of users. The problems that are faced in performing web usage mining include processing and cleaning of logs, user sessions identification and user habit identification. The solution proposed is SOM that helps in identification of user habits. The input to SOM is 0 or 1 vector that shows presence or absence of a page and output is a map of N*N matrix. The SOM is also compared with another system called K-means for two different websites and result shows that SOM is better in terms of collection of users and pages.

Web usage mining proves to be useful in the area of enhancing the usage of web based learning environment [29]. Typical web based learning system like virtual-u and web-ct not fully successful as they did not allow educators to thoroughly know about all activities performed by learner and in learners case it is desirable that online recommendations are provided for giving direction to search but that does not happen. Two types

of web usage mining activities exist with respect to learning system: offline web usage mining which will provide patterns to educator for determining the access behavior of user and integrated web usage mining that provides automatic recommendations and personalization of activities. Web usage mining was used in the discovery of frequent patterns stored in web server to obtain user navigational pattern [30]. Frequent pattern mining includes three types of pattern: frequent item set, sequences and tree patterns which are utilized in pattern analysis phase of web usage mining. Three algorithms proposed by user: item code set algorithm finds frequent set of pages by using Apriori hypothesis. Second algorithm is SM-tree that finds the order in which pages are arranged and is based on deterministic finite machine and third algorithm PD-tree uses pushdown automata for determining whether a tree includes another tree or not. The experiment is performed and result shows hat frequent pattern mining support pattern analysis phase of web usage mining which discovers the user navigational pattern on the web.

Relationship between web usage mining and web structure mining was discussed in [28]. Integration of both the techniques provides the advantage of faster web access to user, saving server memory space and better bandwidth utilization. Preprocessing, pattern analysis and knowledge discovery are three phases in web usage mining. Main task of preprocessing is to remove dirty, redundant and inconsistent data from server log files. Pattern analyses discovers interesting patterns by making use of web tools like webviz and categorize the pages into excellent, medium and weak. Then knowledge is discovered which helps in decision making based on hit counts. After the three phases restructuring of website is done so that popular pages can be moved closer to home page.

A review has been done on recent developments in web usage mining research and is discussed in [27]. The process, techniques and applications of web usage mining is explained.

4. Privacy Concerns

With the increase in the growth of e-commerce on the web, the need for maintaining the privacy of user while using any site also increases. Global and self regulatory nature of web raises the need for improving the privacy bar of user. For enhancing the site structure, administrators are always in requirement of demographics of user as well as their usage pattern, with which the user may not be comfortable with. Proper guidelines and rules need to be introduced such that site administrator can do their work without compromising the identity of user. To deal with this issue W3C introduced Platform for Privacy Preference (P3P) [28]. A protocol is provided by P3P which allows site administrator to publish privacy policies, and when user visit site he has to agree with these policies in order to use site.

5. Web usage Mining Tools

Different types of tools used in all the three stages of web usage mining are described in table 1.

TABLE 1

Tools used in various stages of Web Usage Mining

Tools	Features
Data Preprocessing Tools	
Data Preparator	Performs cleaning, extraction and transformation of data before pattern discovery.
Sumatra TT	Platform independent data transformation tool. Based on Sumatra script and support Rapid application Development.
Lisp Miner	Performs data preprocessing by analyzing the click stream and data collected.
SpeedTracer	Mines web server logs and reconstruct the user navigational path for session identification
Pattern Discovery Tools	
SEWEBAR-CMS	Provides interaction between data analyst and domain expert to perform discovery of patterns. Helps in selection of rules among various rules in association rule mining [34].
i-Miner	Discover data cluster by using fuzzy clustering algorithm and fuzzy inference system for pattern discovery and analysis [33].
Argunaut	Develop the patterns of useful data by using sequence of various rules.
MiDas(Mining Internet Data for Associative Sequences)	Discover marketing based navigational pattern from log files. It applies more features to traditional sequential method.
Pattern Analysis Tools	
Webalizer	GNU GPL license based and produces web pages after analyzing patterns.
Naviz	Visualization tool that combines 2-D graph of visitor access and grouping of related pages. It describes the pattern of user navigation on the web.
WebViz	Analyze the patterns and provides them in the form of graphical patterns
WebMiner	Mines the useful patterns and provides the user specific information
Stratdyn	Enhances WUM and provides visualization of patterns

6. Conclusion

This paper has attempted to provide a review of rapidly growing area of web usage mining. Web usage mining is used in various areas such as e-business, e-CRM, e-education, bioinformatics, and digital libraries and so on. The only information left behind by user when they visit any site is about their ac-

cess pattern. Web usage mining makes use of this information in order to mine the desired information and make it available to user efficiently and efficaciously. Content and structure preprocessing allows raw data to be preprocessed along these dimensions also. Patterns are discovered by making use of various techniques like statistical analysis, clustering, association rules and so on. The involvement of intelligent agents and knowledge query mechanisms improves the efficiency of pattern analysis. Privacy issues related with web usage mining is also discussed in last section of this paper.

7. Future Directions

The future work requires the need for development of more specialized logs from the application side. They will eliminate the problem of insufficient data record on web server. This will help in much better analyses of the data. A better prototype system and intelligent agent needs to be developing for providing the useful recommendations on the site and improving the efficiency of personal web usage mining. With multimedia content growing rapidly on the web, multimedia data mining is another research direction of web usage mining. Various machine learning techniques such as pattern recognition, image analysis have been used for this purpose but still there is a lot of requirement of improvement. Semantic web is another future vision in web usage mining. It is still in its infancy and provides considerable prospects for web usage mining research. The major weakness is it depends upon web authors for its success, this needs to be eliminated.

8. References

[1] J. Srivastava, R. Cooley, M. Deshpande and P. Tan., "Web Usage Mining: Discovery and Applications of Usage Patterns from Web data", Department of Computer Science and Engineering, University of Minnesota. SIGKDD Explorations, 1(2):12, January 1999.

[2] Kosla, R. and Blockeel, H.2000, "Web Mining Research: A Survey", SIGKDD Explorations Vol. 2, 1-15.

[3] J. Srivastava, P. Desikan and V. Kumar, "Web Mining- Accomplishments and Future Directions", University of Minnesota, Minneapolis, USA.

[4] Ajoudanian, S. and jazi, M.D.2009, "Deep Web Content Mining", World Academy of Science, Engineering and Technology 49.

[5] Van.C.J. Information Retrieval Butterworths *unpublished(unpublished manuscript)*

[6] M.T. Pazienza," Information Extraction: A multidisciplinary Approach to an Emerging Information Technology", Volume 1299 of lecture notes in Computer Science. International Summer School, SCIE-97 Frascati (Rome), Springer 1997.

[7] I. Muslea,"Extraction Patterns for Information Extraction tasks: A Survey", In *AAAI-99 Workshop on Machine Learning for Information Extraction, 1999*

[8] Mitchell, T.1997. Machine Learning McGraw Hill.

[9] S. Chakrabarti, B. Dom and P. Indyk,"Enhanced Hypertext Categorization using Hyperlinks", In SIGMOD 98: Proceddings of the 1998 ACM SIGMOD International Conference on Management of Data. New York, NY, USA: ACM Press, 1998 pp 307-318.

[10] J.M. Kleinberg, R.Kumar, P.Raghvan, S.RAjagopalan and A.S. Tomkins," The Web as a graph: Measurements, Models and Methods", Lecture notes in Computer Science, vol. 1627, pp 1-18, 1999.

[11] J. Hou and Y. Zhang," Effectively finding relevant Web Pages from Linkage Information", IEEE Trans. Know. Data Eng. Vol. 15, No.4, pp-940-951, 2003

[12] H. Han and R. Elmasri.," Learning Rules for Conceptual Structure on the Web", J.Intell. Inf. Syst. Vol.22, No3 pp 237-256, 2004.

[13] O. Etzioni.,"The World Wide Web: Quagmire or Gold Mine", Communications of the ACM, 39(11):65-68, 1996

[14] M. Eirinaki and M. Vazirgiannis.,"Web Mining for web Personalization.", ACM Trans. Inter. Tech. Vol. 3, No.1, pp 1-27, 2003.

[15] G. Salton and M.J. McGil.,"Introduction to modern Information Retrieval.", McGraw-Hill, New York, 1983.

[16] J. Han and M. Kamber.,"Data Mining Concepts and Techniques", Morgan Kaufmann, 2001.

[17] A. Banerjeee and J. Ghosh,"Click stream Clustering using Weighted Longest Common Subsequences", In Proceedings of the Web Mining Workshop at the 1st SIAM Conference on Data Mining, 2001.

[18] J. Heer and Ed H. Chi.,"Mining the Structure of User Activity Using Cluster stability", In Proceedings of the Workshop on Web Analysis, Second SIAM Conference on Data Mining ACM Press, 2002.

[19] U. Fayyad G. Piatetsky-Shapiro and P. Smyth.,"From Data Mining to Knowledge Discovery: An Overview", In Proc. ACM KDD 1994.

[20] A.G.Buchner, M. Baumgarten, S.S.Anand MD.Mulvenna and J.G. Hughes. "Navigation Pattern Discovery from Internet Data", In WEBKDD, San Diego, CA 1999.

[21] R. Cooley,"Web Usage Mining: discovery and Applications of Interesting Patterns from Web Data", PhD thesis, University of Minnesota, 2000.

[22] C. Shahabi and F. b. Kashani.,"Efficient and Anonymous Web Usage Mining for Web Personalization", Department of Computer Science Integrated Media Systems Center, University of Southern California, Los Angeles, California 90089-2561, USA.

[23] R. Yates.,"Web Usage Mining in search engines", Center for web research Department of Computer science .Universidad di Chile.

[24] J. Kerkhofs, K. Vanhoof and D. Pannemans,"Web Usage Mining on Proxy servers: A Case Study", Limburg University Center, July 30, 2001

[25] M. Albanese, A. Picariello and C.L.Sansane.,"A Web Personalization System based on Web Usage mining Techniques", University di Napoli Federico II, Italy.

[26] F.M. Facca, P.L. Lanzi. ,"Recent Developments in web Usage mining Research", Artificial Intelligence and Robotics Laboratory.

[27] G.R.B., S.Totad, P.PVGD.,"Amalgamation of Web Usage Mining and Web Structure Mining.", International Journal of Recent trends in Engineering Vol.1 No2 May 2009

[28] R. Ivancsy, I. Vajk.,"Frequent Pattern Mining in Web Log Data", Acta Polytechnica Hungarica vol3 No1 2006.

[29] O. Zaiane.,"Web Usage Mining for a Better Web Based learning Environment." Unpublished (unpublished manuscript)

[30] Y.Fu, M. Shih. ,"A Framework for Personal Web Usage Mining." Unpublished (unpublished manuscript)

[31] P. Britos, D. Martinelli, H. merlino, R. Martinez.,"Web Usage Mining using self Organized Maps.",IJCSNS International Journal of Computer Science *and Network Security Vol7 No 6 2007.*

[32] Abraham, A. "i-Miner: A Web Usage Mining Framework Using Hierarchial *Intelligent systems", IEEE International Conference on fuzzy systems Fuzz*-IEEE" 03, IEEE Press, pp. 1129-1134.

[33] T. Kliegr, D. Chudan, A. Hazucha and J. Rauch," SEWBAR_CMS: A System for postprocessing Association rule Models", Faculty of information and statistics VSE Praha., http://www.joomla.org.