

A Study of Web Usage Mining Research Tools

Chhavi Rana

Department of Computer Science Engineering, University Institute of Engineering and Technology,
MDUniversity, Rohtak, Haryana,124001, India.
Email: chhavi1jan@yahoo.com

ABSTRACT

Web usage mining focuses on techniques that could predict user behavior while the user interacts with the Web. It tries to make sense of the data generated by the Web surfer's sessions or behaviors. There is an attempt to provide an overview of the state of the art in the research of web usage mining, while discussing the most relevant tools available in the sphere as well as the niche requirements that the current variety of tools lack. It will give an outlook on the existing tools, their specialized focus with respect to an applicative objectives and the need for a more comprehensive new entrant in this sphere in the light of the current scenario. In the end, the paper will be concluded by listing some challenges and future trends in this research area. Overall the focus of the paper will be to present a survey of the recent developments in this area which is getting too much attention from web development arena.

Keywords - Artificial Intelligence, Data Mining, Information overload, Navigation Patterns, Web Log, Web Mining, Web Personalization, Web site structure, Web usage mining,

Date of Submission: January 23, 2012

Date of Acceptance: March 24, 2012

I. INTRODUCTION

Web mining is a very interesting research topic which combines two of the activated research areas: Data Mining and World Wide Web. With the huge amount of information available online, the World Wide Web is a fertile area for data mining research. The Web mining research relates to several research communities, such as database, information retrieval, and AI. The World Wide Web (Web) is a popular and interactive medium to disseminate information today. The Web is huge, diverse, and dynamic and thus raises the scalability, multimedia data, and temporal issues respectively.

It was Oren Etzioni who first coined the term Web mining in his paper in 1996. Etzioni starts by making a hypothesis that the information on the Web is sufficiently structured and outlines the subtasks of Web mining [1]. His paper describes the Web mining processes. Web data mining can be defined as the discovery and analysis of useful information from the WWW data. Since then, there have been several works around the survey of data mining on the Web. Although Web mining puts down the roots deeply in data mining, it is not equivalent to data mining. The unstructured feature of Web data triggers more complexity in the process of Web mining. An exponential growth in on-line information combined with the almost unstructured web data necessitates the development of powerful yet computationally efficient web data mining tools [2].

Web mining is the use of data mining techniques to automatically discover and extract information from Web documents and services [1]. This area of research is so huge today partly due to the interests of various research communities, the tremendous growth of information

sources available on the Web and the recent interest in e-commerce. Subsequently, Madria, et al. [2] and Borges and Levene [3], categorized Web mining into three areas of interest based on which part of the Web to mine: Web content mining, Web structure mining, and Web usage mining. In practice, the three Web mining tasks above could be used in isolation or combined in an application, especially in Web content and structure mining since the Web documents might also contain links. For example, Chakrabarti et al. [4] uses as Web content the terms in a document's link neighborhood and as Web structure the links from its neighbors, to classify Web pages. Joachims, Freitag, and Mitchell [5] use Web content and usage to build a software tour agent for assisting users browsing a Web site.

II. WEB USAGE MINING

Web usage mining focuses on techniques that could predict user behavior while the user interacts with the Web. As mentioned before, the mined data in this category are the secondary data on the Web as the result of interactions. These data could range very widely but generally we could classify them into the usage data that reside in the Web clients, proxy servers and servers [6]. The Web usage mining process could be classified into two commonly used approaches [3]. The Web usage mining process can be regarded as a three-phase process, consisting of the data preparation, pattern discovery and pattern analysis phases (See figure 1, Mobasher et al. [7]). In the first phase, Web log data are preprocessed in order to identify users, sessions, page views, and so on. In the second phase, statistical methods, as well as data mining methods (such as association rules, sequential pattern discovery, clustering, and classification) are applied in order to detect interesting patterns. These patterns are

stored so that they can be further analyzed in the third phase of the Web usage mining process [8].

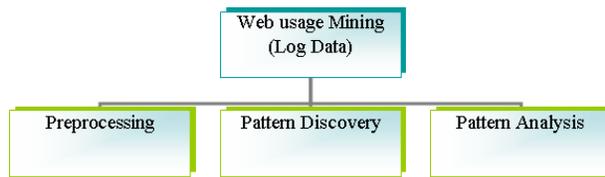


Figure 1

The first approach maps the usage data of the Web server into relational tables before an adapted data mining technique is performed. The second approach uses the log data directly by utilizing special pre-processing techniques. As is true for any typical data mining applications, the issues of data quality and pre-processing are also very important here. The typical problem is distinguishing among unique users, server sessions, episodes, etc. in the presence of caching and proxy servers [7,9]. In general, typical data mining methods could be used to mine the usage data after the data have been pre-processed to the desired form [6]. However, modifications of the typical data mining methods are also used such as composite association rules [3], an extension of a traditional sequence discovery algorithm MIDAS [10] and hypertext probabilistic grammars. The Web usage data could also be represented with graphs [10]. Often Web usage mining methods use some background or domain knowledge such as navigation templates, Web content, site topology, concept hierarchies, and syntactic constraints [10,11].

The applications of Web usage mining could be classified into two main categories: On one hand, learning a user profile or user modeling in adaptive interfaces (personalized) and learning user navigation patterns (impersonalized) [12]. Web users would be interested in, among others, techniques that could learn their information needs and preferences, which is user modeling possibly combined with Web content mining. On the other hand, information providers would be interested in, among others, techniques that could improve the effectiveness of the information on their Web sites by adapting the Web site design or by biasing the user's behavior towards satisfying the goals of the site. In other words, they are interested in learning user navigation patterns. Then the learned knowledge could be used for applications such as personalization (at a Web site level), system improvement, site modification, business intelligence, and usage characterization (Srivastava et al., 2000).

III. RELATED RESEARCH

There have been some works around content mining, and structure mining, based on the research of Data mining and Information Retrieval, Information Extraction, and Artificial Intelligence. But, in the web usage mining research area, several groups did distinguished work. From the business and applications point of view, knowledge

obtained from the Web usage patterns could be directly applied to efficiently manage activities related to e-business, e-services, e-education and so on [13]. Accurate Web usage information could help to attract new customers, retain current customers, improve cross marketing/sales, effectiveness of promotional campaigns, track leaving customers and find the most effective logical structure for their Web space [14]. User profiles could be built by combining users' navigation paths with other data features, such as page viewing time, hyperlink structure, and page content.

The last comprehensive survey on web usage mining has been done by Koutri, Avouris, and Daskalaki [15]. Pierrakos et al. [16] earlier explored the area of web usage mining as a tool for the personalization process. Eirinaki, and Vazirgiannis [8] also presented an excellent survey upon a overall area of web mining for personalization. Kosala and Blockeel [17] explored the terms of web mining and the related research area earlier in their work. Since then web mining research has been somewhat traced in the annual WebKDD workshop. WebKDD 2008 was the tenth of a successful series of Workshops. Launched in 1999 by Brij Masand and Myra Spiliopoulou, the first workshop of the series WebKDD 1999 invited contributions on "Web usage mining". In the 10 years that followed, the scope of WebKDD was broadened to cover the emerging KDD topics on the web. Together with this, historical study has been conducted by several researchers that are specialized in web mining techniques and several frameworks have already been explored. The result of these researches has lead to the development of lot of applications in the area of web mining and they are successfully applied in business and e-commerce domain areas.

Few landmark researches have been followed here. Cooley et al. [18] proposed a framework for web mining using various web mining task and implemented a prototype namely Webminer. It is implemented by applying a framework that perform cluster analysis on association rules and sequential pattern discovery. Zaiane [19] proposed the idea of how to implement the OLAP technique on the Web mining. Their works on the multimedia data also provided a valuable solution for content mining. Spiliopoulou [12] focused on the applications of the usage mining. Cooley [20] in University of Minnesota did in-depth research to all the procedure of usage mining. They proposed a mining prototype Web Miner and derived a system Web SIFT to perform the usage mining, which is relatively practical.

Lee and Liu [21] proposed an intelligent multi-agent based environment known as intelligent Java Development Environment (iJADE) to provide an integrated and intelligent agent based platform in the e-commerce environment on Internet shopping. The application of intelligent agent for helping users is applied in various applications and not only in e-commerce environment. Mobasher et al. [22] proposed an effective and scalable techniques for Web personalization based on association

rule discovery from usage data. Toolan and Kushmerick [23] proposed techniques based on web usage mining to deliver personalized Site Maps that are specialized to the interest of each individual visitor. Applying the agent technology has improved the performance of web mining compared to traditional approach such as database approach. On the other hand, Eirinaki and Vazirgiannis [8] developed a module that comprises a web personalization system concerning the web usage mining module. Also, Lu, Dunham, and Meng [24] later proposed a technique to generate Significant Usage Patterns (SUP) and used them to acquire significant “user preferred navigational trails”. Falkowski et al. [25] proposed two approaches to analyze the evolution of two different types of online communities on the level of subgroups.

Labroche, Lesot, and Yaffi [26] introduces a new tool for web usage mining and visualization that relies on the bio-mimetic relational clustering algorithm based on typicality computation to produce an efficient visualization of the activity of users on a website. Khalil, Li and Wang [27] recently endeavors to provide an improved Web page prediction accuracy by using a novel approach that involves integrating clustering, association rules and Markov models according to some constraints. Experimental results prove that this integration provides better prediction accuracy than using each technique individually. Tao, Hong and Su [28] explore a new data source called intentional browsing data (IBD) for potentially improving the effectiveness of WUM applications. Nasraoui et al. [29] present an approach for discovering and tracking evolving user profiles and enrich it with explicit information gathered from web log data. Together with this profiles are also provides other domain-specific information and a validation strategy is implied to assess the quality of the mined profiles. Masegla et al. [30] propose to perform a specific data mining process to extract frequent behaviours by discovering the densest periods. Such periods are the one having at least one frequent sequential pattern for the set of users connected to the Web site in that period. Khiribi, Jemni, & Nasraoui [31] build a personalized recommendation engine which aim to compute on-line automatic recommendations to an active learner based on his/her recent navigation history. This is done by exploiting similarities and dissimilarities among user choices and among the contents of the resources.

David et al. [32] proposed a probabilistic model for a web site that uses the entropy of a Markov chain in order to compute the user navigation patterns from the log data. Most of the research in the area of web usage mining focus on the algorithm while disregarding the type of data on which the algorithm will be applied. Hasan, Mudur and Shiri [33] proposed a simple yet effective technique called generalization of web sessions that replaces actual page-clicks with their general concepts. This approach is very effective in overcoming the problem of scalability with respect to web usage mining. Dai and Mobasher [34] emphasized the need to associate Web usage and content knowledge, by enhancing the information in the Web

usage logs with semantics derived from the content of the Web site’s pages. Rao, Kumari, and Raju, [35] developed an algorithm based on association rule mining with incremental technique to suit the dynamically changing log scenario which is more efficient than running a number of scans of database. Kumar and Rukmani [36] compared how Apriori algorithm and Frequent Pattern Growth algorithm differ in terms of memory usage and time usage while discovering the web usage patterns of websites from the server log files. Thakare and Gawali [37] emphasis the importance of an the effective and complete preprocessing of access stream before actual mining process can be performed. This could significantly improve the automatic discovery of meaningful pattern and relationships from access stream of user. Senkul and Salin [38] worked upon investigating the effect of semantic information on the patterns generated for Web usage mining in the form of frequent sequences. The frequent navigational patterns are composed of ontology instances instead of Web page addresses.

IV. APPLICATION AND TOOLS

Web mining for usage pattern is the key to discover marketing intelligence in e-commerce. It helps tracking of general access pattern, personalization of web link or web content and customizing adaptive sites. It can disclose the properties and inter-relationship between potential customers, users and markets, so as to improve Web performance, on-line promotion and personalization activities [39]. There are many popular programs for usage pattern mining (see Table 1). Web Log Mining [6] uses KDD techniques to understand general access patterns and trends to shed light on better structure and grouping of resource providers. For e.g., Web miner [9] discovers association rules and sequential patterns automatically from server access logs. Commercial software Web Analyst by Megaputer learns the interests of the visitors, based on their interaction with the website. Clementine and DB2 Intelligent Miner for Data are two general-purpose data mining tools, which can be used for web usage mining with suitable data preprocessing.

Table 1. Major applications of web usage mining and related projects:

Sr.No.	Application	Tools/Projects
1.	<u>Usage Characterization</u>	1. Manley 2. Pitkow 3. Almeida
2.	<u>System Improvement</u>	1. Rexford 2. Schecter 3 Agarwall
3.	<u>Business Intelligence</u>	1. SurfAid 2. Buchner 3. Tuzilin 4. Abraham
4.	<u>Personalization</u>	1. Site Helper 2. Webshifter 3. Mobasher

		4. Letizia
5.	<u>Site Modification</u>	1. Etzioni 2. Perkowitz

There are several commercial software tools (see Table 2) that could provide Web usage statistics. These stats could be useful for Web administrators to get a sense of the actual load on the server. However, the statistical data available from the normal Web log data files or even the information provided by Web trackers could only provide the information explicitly because of the nature and limitations of the methodology itself. Generally, one could say that the analysis relies on three general sets of information given a current focus of attention: (1) past usage patterns; (2) degree of shared content; and (3) inter-memory associative link structures. After browsing through some of the features of the best trackers available it is easy to conclude that rather than generating statistical data and texts they really do not help to and much meaningful information. For small web servers, the usage statistics provided by conventional Web site trackers may be adequate to analyze the usage pattern and trends. However as the size and complexity of the data increases, the statistics provided by existing Web log file analysis tools may prove inadequate and more intelligent knowledge mining techniques will be necessary. Maier and Reinartz [40] conducted a comprehensive evaluation of Web Usage Analysis Tools in their research.

V. CHALLENGES AND FUTURE TRENDS

The Web presents new challenges to the traditional data mining algorithms that work on flat Data. We have seen that some of the traditional data mining algorithms have been extended or new algorithms have been used to work on the Web data. With explosive growth of the information sources available on the World Wide Web, it has become increasingly necessary for users to utilize automated tool in order to find the required information resources, and to track and analyze their usage patterns. These factors give rise to the necessity of creating server side and client side intelligent systems that can effectively mine for knowledge. The analysis of large web log files is a complex task not fully addressed by existing web access analyzers. However, it is hard to find appropriate tools for analyzing raw web log data to retrieve significant and useful information. There are several commercially available web log analysis tools, but most of them are disliked by their users and considered too slow, inflexible, expensive, difficult to maintain or very limited in the results they can provide.

While some tools using data mining techniques to help web log analyses are being developed, the research is still

in its infancy. The existing techniques for analyzing web usage have different drawbacks, i.e., either huge storage requirements, excessive I/O cost, or scalability problems when additional information is introduced into the analysis. Most of the currently available Web server analysis tools provide only explicitly and statistical information without real useful knowledge for Web managers. The task of mining useful information becomes more challenging when the Web traffic volume is enormous and keeps on growing.

The potential of using a website as a data collection tool for web based information systems is enormous. This because of its interactivenss, simplicity and unobtrusiveness. The results of the data mining would ideally be integrated into the dynamic website to provide an automated, end-to-end functional system for target marketing and customer relationship management. Most of the web mining tools are evolving and the present web mining techniques still have rooms for improvement to make them prevail in the web based information systems. Some problems like the need for greater integration, scalability issue, and the need for better mining tools are frequently mentioned by many researchers. The sharpening on the mining tools in many different aspects is important for the future development in this area:

- Web usage mining must handle the integration of offline data with e-business analytic tools, RDBMS, catalogs of products and services and other applications.
 - Some new variables or logs should be sought that can be used for finding more natural, meaningful and useful patterns.
 - New tools are needed which will not use up too much resources or process time during the web mining process.
 - There will always be a need to have benchmark tests to improve the performance of mining algorithms, as the efficiency and effectiveness of a mining algorithm can be measured and a better tool for web data mining can be derived.
 - It is important to improve visualization, as much of the data is unorganized and difficult for the user to understand.
- Web mining is a new and rapidly developing research and application area. With more collaborative research across different disciplines like database, artificial intelligence, statistics and marketing, we will be able to development web mining applications that are very useful to the web based information systems.

Table 2: Some web usage mining tools

Sr. No.	Tool	Researcher/Company	Feature	Function
1	ArchCollect	Esmin et al., 2008.	It uses fast semantic interaction acquisition algorithms and form a dimensional cube of information directly which serve as input to an IR approach [41].	To monitor users' interactions in web media.
2	i-Miner	Abraham et al., 2003.	To optimize the concurrent architecture of a fuzzy clustering algorithm (to discover data clusters) and a fuzzy inference system to analyze the trends [42].	Pattern Discovery and trend analysis from web usage data.
3	AWUSA	Tiedtke et al., 2002.	A framework based on combination of information architecture, automated usability evaluation and web mining techniques for data-gathering and analysis [43].	Automated website usability evaluation.
4	i-JADE Web-Miner	Lee and Liu, 2001.	To use Agent technology, together with Web mining technology, to automate a series of product search and selection activities. It is based on multiagent development platform iJADE [19].	E-commerce applications.
5	Web Quilt	Hong et al., 2001.	Web logging and visualization system that helps web design teams capture usage traces which can be aggregated and visualized in a zooming interface that shows the web pages people viewed [14].	To run usability tests and analyze the collected data from web logs.
6	KOINOTITES	Pierrakos et al., 2000.	A system which uses data mining techniques for the construction of user communities on the Web [44].	Personalisation
7	INSITE	Shahabi et al., 2000.	To generate user profiles in real time through the use of a unique Connectivity Matrix Model (CM-model), And show the efficacy and scalability[45].	INSITE acquisition Extracts and stores the essence of the captured information in real time and visualize the result.
8	STRATDYN	Berendt,B. 2000.	Developed as add on module to the WUM, extends its capability by exploiting site semantics [46].	Visualization of navigation patterns.
9	SEWeP	Eirinaki M. et al. 2003	Developed as a system that makes use of both the usage logs and the semantics of a Web site's content in order to personalize it [47].	Personalisation
9	WebTool	Masseglia et al., 2000.	It uses sequential pattern mining which relies on PSP an algorithm developed by the authors [48].	Usage profiling.
10	Web SIFT (Based on WEBMINER prototype 1997)	Cooley et al., 1999.	System uses the content and structure information from a website in order to identify interesting results from mining usage data [49].	To mine interesting frequent item sets automatically from real web data.
11	MiDAS (Mining Internet Data for Associative Sequences)	Büchner et al., 2000.	It introduces a new algorithm called MiDAS that extends traditional sequence discovery with a wide range of web-specific features [10].	Pattern discovery.
12	Web usage miner (WUM)	Spiliopoulou and Faulstich 1998.	It exploits an innovative aggregated storage representation for the information in the web server log. It discovers patterns comprised of not necessarily adjacent events [50].	Mining interesting navigation patterns in the form of graphs.
13	Web Mate	Chen & Sycara, 1998.	The user profile is inferred from training examples [13].	As Proxy agent provides effective browsing and searching Help.
14	SpeedTracer	Wu, Yu and Ballman, IBM, 1998.	Reconstructs the user transversal paths for session identification by using the referrer page and the URL of the requested page as a traversal step [51].	Mining web server log files.
15	WebLogMiner	Zaiane, 1998	Use data mining and OLAP on treated and transformed web access files [17].	Mining web server log files.

16	DB2 Intelligent Miner for Data	IBM cradle	Provides a single framework for database mining using proven, parallel mining techniques.	User database miner
17	Poly Analyst version 6.0	Megaputer, (1997-2007)	Integrates the data and text mining capabilities of analytical software directly.	Profiles the website resources and dynamically identifies the most appropriate resources to serve each visitor.
18	Clementine	SPSS, Apache Software Foundation, 1997.	To browse data using interactive graphics to find important features and relationships.	CRM
19	WEBMINER	Cooley, Srivastava and Mobasher., 1997.	A general and flexible framework for Web usage mining, the application of data mining techniques, such as the discovery of association rules and sequential patterns, to extract relationships from data collected in large Web data repositories [16].	Restructure a Web site, and in analyzing user access patterns to dynamically present information tailored to specific groups of users.
20	WEBVIZ	Pitkow & Bharat, 1994	By incorporating Web-Path paradigm into interactive software, users can see not only the documents but also the hyperlinks traveled and have lot more options [47].	Provides graphical view of their local database and access patterns.

VI. Conclusion

Designing and maintaining web based information systems, such as Web sites, is a real challenge. On the Web, it is much easier to find inconsistent pieces of information than a well structured site. The study of web usage mining and its research could help a lot in building tools that can support the design, development and maintenance of complex but coherent sites. The approach is multi-disciplinary, involving Software Engineering and Artificial Intelligence techniques. There is a strong relation between structured documents (such as Web sites) and a program; the Web is a good candidate to experiment with some of the technologies that have been developed in software engineering. Web Mining has been an important topic in data mining research in recent years from the standpoint of supporting human-centered discovery of knowledge. The present day model of web mining suffers from a number of shortcomings as listed earlier. As services over the web continue to grow (Katz 2002), there will be a continuing need to make them robust, scalable and efficient. Web usage mining can be applied to better understand the behavior of these services, and the knowledge extracted can be useful for various indices of optimizations. There is need to study the loopholes in the analysis of web usage patterns through existing tools and to design efficient, scalable and powerful analysis tools. The development of these new tools will deal with highly structured content such as XML that can only be processed with more sensitive tools than raw text mining.

Web usage mining (WUM) can be used to determine if the information architecture of a web site is structured correctly. Existing WUM tools however, do not indicate which data mining algorithms are being used or provide effective graphical visualizations of the results obtained. There are many commercial tools which perform analysis on log data collected from Web servers. With respect to Web Mining commercial tools, it is worth noting that since the review made in [6], the number of existing products

almost doubled. Most of these tools are based on statistical analysis techniques, while only a few products actually exploit Data Mining techniques. The majority of the public and shareware tools for the analysis of Web application usage are traffic analyzers. Their functionality is limited to producing reports about site traffic, (e.g., number of visits, number of hits, page view time, etc.), diagnostic statistics, (such as server errors and page not found), referrer statistics, (such as search engines accessing the application), user and client statistics (such as user geographical region, Web browser and operating systems, etc). Only few of them also track user sessions and present specific statistics about individual users' accesses.

In future, web usage mining research promises lot of space for advancements in the techniques and tools that can make some improvements in web sites specifically by focusing on the visualization of user navigation pattern by using combination technology of knowledge-based system and web-mining method. The purpose of web usage mining is to make contributions in improving the overall quality of Information Systems, to support designers during the design process and to ensure ease of use to end users.

REFERENCES

- [1] O. Etzioni, The World-Wide Web: quagmire or gold mine?, *Communications of the ACM*, 39(11), 1996, 65-68 .
- [2] S. K. Madria, S. S. Bhowmick, W. K. Ng, and E. Lim, Research Issues in Web Data Mining, *Data Warehousing and Knowledge Discovery*, 1999, 303-312.
- [3] J. Borges, and M. Levene, Data Mining of User Navigation Patterns', *Web Usage Analysis and User Profiling*, San Diego, CA, USA, 2000, 31-39.
- [4] S. Chakrabarti, B. E. Dom, S. Ravi Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, D. Gibson, and J. Kleinberg, Mining the Web's Link Structure', *Computer*, 32(8), 1999, 60-67.

- [5] T. Joachims, D. Freitag, T. Mitchell, WebWatcher: A Tour Guide for the World Wide Web, *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*, Morgan Kaufmann, 15, 1997, 770-777.
- [6] J. Srivastava, R. Cooley, M. Deshpande, P. Tan, Web usage mining: Discovery and applications of usage patterns from web data' *SIGKDD Explorations newsletter*, 1(2), 2000, 12-23.
- [7] B. Mobasher, R. Cooley, and J. Srivastava, J. Automatic personalization based on Web usage mining, *Communications of the ACM*, 43(8), 2000, 142-151.
- [8] M. Eirinaki and M. Vazirgiannis, Web mining for web Personalization, *ACM Transactions on Internet Technology*, 3(1), 2000, 1-27.
- [9] B. Masand, M. Spiliopoulou, J. Srivastava, and O. R. Zaiane, Web Mining for Usage Patterns & Profiles, *WEBKDD02*, *SIGKDD Explorations*, 4(2), 2002, 125-127.
- [10] A. G. Buchner, M. Baumgarten, S. S Anand, M. D. Mulvenna, and J. G. Hughes, Navigation Pattern Discovery from Internet Data, *Proceedings of Web Usage Analysis and User Profiling at the International WEBKDD99 Workshop*, 2000, 74-91.
- [11] S. Baron, and M. Spiliopoulou, Monitoring the Evolution of Web Usage Patterns, *EWMF 2003*, 181-183.
- [12] M. Spiliopoulou, Data mining for the Web, *Proceedings of the Third European conference, PKDD'99*, 1999, 588-589.
- [13] L. Chen, and K. Sycara, WebMate: A Personal Agent for Browsing and Searching, *Proceedings of the 2nd International Conference on Autonomous Agents*, Minneapolis MN, USA, 1999, 132-139.
- [14] J. I. Hong, J. Heer, S. Waterson, and J. A. Landay, WebQuilt: A proxy-based approach to remote web usability testing, *ACM Transactions on Information Systems*, 19(3), 2001, 263-285.
- [15] M. Koutri, N. Avouris, and S. Daskalaki, A survey on web usage mining techniques for web-based adaptive hypermedia systems, *Adaptable and Adaptive Hypermedia Systems Idea*, 2004, 1-23.
- [16] D. Pierrakos, G. Paliouras, C. Papatheodorou, and C. D. Spyropoulos, Web usage mining as a tool for personalization: A survey, *User Modeling and UserAdapted Interaction*, 13(4), 2003, 311-372. Kluwer Academic Publishers.
- [17] R. Kosala, and H. Blockeel, Web Mining Research: A Survey', *Machine Learning*, 2(1), 2000, 1-15.
- [18] R. Cooley, B. Mobasher, and J. Srivastava, Web mining: information and pattern discovery on the World Wide Web', *Proceedings Ninth IEEE International Conference on Tools with Artificial Intelligence*, 97(2.1), 1997, 558-567.
- [19] O. R. Zaiane, Discovering Web access patterns and trends by applying OLAP and data mining technology on Web logs, *Proceedings IEEE International Forum on Research and Technology Advances in Digital Libraries ADL98*, IEEE Computer Society, Santa Barbara, CA, 1998, 19-29.
- [20] R. Cooley, Web Usage Mining: Discovery and Application of Interesting Patterns from Web data, PhD thesis, Dept. of Computer Science, University of Minnesota, USA, 2000.
- [21] R. S. T. Lee, and J. N. K. Liu, iJADE Web-Miner: An Intelligent Agent Framework for Internet Shopping, *IEEE Transactions on Knowledge and Data Engineering*, 16(4), 2004, 461- 473.
- [22] B. Mobasher, H. Dai, T., Luo, and M. Nakagawa, Effective personalization based on association rule discovery from web usage data', *Proceeding of the third international workshop on Web information and data management WIDM 01*, 9, USA, 2001, 9-15
- [23] F. Toolan, and N. Kusmerick, Mining Web Logs for Personalized Site Maps, *Proceedings of the Third International Conference on Web Information Systems Engineering (Workshops) - (WISEw'02) (WISEW '02)*. IEEE Computer Society, Washington, DC, USA, 2002, 232-237.
- [24] L. Lu, M. H. Dunham, and Y. Meng, Discovery of Significant Usage Patterns from Clusters of Clickstream Data, *Proceedings of the ACM SIGKDD Workshop on Knowledge Discovery in Web WebKDD05*, 2005, Chicago, IL, USA.
- [25] T. Falkowski, J. Bartelheimer, and M. Spiliopoulou, Mining and Visualizing the Evolution of Subgroups in Social Networks, *Proceedings of the 2006 IEEE/WICACM International Conference on Web Intelligence*, 2006, 52-58.
- [26] N. Labroche, M. J. Lesot, and L. Yaffi, A New Web Usage Mining and Visualization Tool', *19th IEEE International Conference on Tools with Artificial Intelligence ICTAI 2007*, 1, 2007, 321-328.
- [27] F. Khalil, J. Li, and H. Wang, Integrating Recommendation Models for Improved Web Page Prediction Accuracy', *Reproduction*, 74(Acsc), Australian Computer Society, Inc. ACM International Conference Proceeding Series, Wollongong, Australia, Vol. 312, 2008, 91-100.
- [28] Y. Tao, T. Hong, and Y. Su, Web usage mining with intentional browsing data, *Expert Systems with Applications*, 34(3), 2008, 1893-1904.
- [29] O. Nasraoui, M. Soliman, E. Saka, A. Badia, and R. Germain, A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites', *IEEE Transactions on Knowledge and Data Engineering*, 20(2), 2008, 202-215.
- [30] F. Masseglia, P. Poncelet, M. Teisseire, and A. Marascu, Web usage mining: extracting unexpected periods from web logs, *Data Mining and Knowledge Discovery*, 16(1), 2008, 39-65.
- [31] M. K. Khribi, M. Jemni, and O. Nasraoui, Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval, *Eighth IEEE International*

- Conference on Advanced Learning Technologies, 12(4), 2008, 241-245.
- [32]N. David, L. Patrascu, A. Sasu, & D. Damian, A probabilistic model for web usage mining, *Proceedings of the 8th Wseas international conference on Telecommunications and informatics*, World Scientific and Engineering Academy and Society (WSEAS), 2009, 129–133.
- [33]T. Hasan, S. P. Mudur, and N. Shiri, A session generalization technique for improved web usage mining', *WIDM 09 Proceeding of the eleventh international workshop on Web information and data management*, New York, NY, USA, 2009, 23-30.
- [34]H. Dai, and B. Mobasher, Integrating Semantic Knowledge with Web Usage Mining for Personalization', *Information Systems Journal*, 2009, 1-28.
- [35]M. Rao, M. Kumari, and K. Raju, Understanding User Behavior using Web Usage Mining', *International Journal of Computer Applications*, 1(7), 2010, 55–61.
- [36]B. S. Kumar, and K. V. Rukmani, Implementation of Web Usage Mining Using APRIORI and FP Growth Algorithms, *International Journal of Advanced Networking and applicatons*, 1(06), 2010, 400-404.
- [37]S. B. Thakare, and S. Z. Gawali, A Effective And Complete Preprocessing For Web Usage Mining, *International Journal On Computer Science And Engineering*, 2(3), 2010, 848-851.
- [38]P. Senkul, and S. Salin, Improving Pattern Quality in Web Usage Mining by Using Semantic Information, *Knowledge and Information Systems*, 1(6), 2011, 400-404.
- [39]I. Cingil, A. Dogac, and A. Azgin, A broader approach to personalization', *Communication of the ACM*, 43(8), 2000, 136-141.
- [40] T. Maier, and T. Reinartz, Evaluation of Web Usage Analysis Tools', *Künstliche Intelligenz*, 18(1), 2004, 65-67.
- [41]Esmin, A., Lima, J., Yano, Tiago, E. T., Carneiro, G. S. (2008) 'ArchCollect - A Tool for WEB Usage Knowledge Acquisition from User's Interactions', *Proceedings of the Tenth International Conference on Enterprise Information Systems*, Barcelona, Spain, pp. 375-380
- [42]Abraham, A. (2003) 'i-Miner: A Web Usage Mining Framework Using Hierarchical Intelligent Systems', *IEEE International Conference on Fuzzy Systems FUZZ-IEEE'03*, IEEE Press , pp. 1129-1134 .
- [43]Tiedtke, T. Martin, C. and Gerth, N. (2002) 'AWUSA – A Tool for Automated Website Usability Analysis', *PreProceedings of the 9th International Workshop on the Design, Specification and Verification of Interactive Systems*.
- [44]Pierrakos, D. Paliouras, G. Papatheodorou, C. and Spyropoulos, C. D. (2000) 'KOINOTITES: A Web Usage Mining Tool for Personalization', *Proceedings of Panhellenic Conference on Human Computer Interaction*, Greece, Patras, pp. 231-236.
- [45]Shahabi, C. Faisal, A. Kashani, F. B. and Faruque, J. (2000) 'INSITE: A Tool for Real-Time Knowledge Discovery from Users Web Navigation', *Proceedings of the 26th International Conference on Very Large Databases (VLDB)*, Cairo, Egypt, pp. 635-638.
- [46]Berendt, B. (2000) 'Web usage mining, site semantics, and the support of navigation', *KDD Workshop on Web Mining for ECommerce Challenges and Opportunities*" pp. 83–93.
- [47] Eirinaki, M., Vazirgiannis, M. and Varlamis, I. (2003) 'SEWeP: using site semantics and a taxonomy to enhance the Web personalization process', *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 99-108.
- [48]Masseglia, F. Poncelet, P. and and Cicchetti. R. (1999) 'WebTool: An Integrated Framework for Data Mining', *Proceedings of the 10th International Conference on Database and Expert Systems Applications (DEXA '99)*, Trevor J. M. Bench-Capon, Giovanni Soda, and A. Min Tjoa (Eds.). Springer-Verlag, London, UK, pp. 892-901.
- [49]Cooley, R., Tan, P-N. and Srivastava, J. (1999) 'WebSIFT: The Web Site Information Filter System', *Proceedings of Workshop on Web Usage Analysis and User Profiling WEBKDD in conjunction with ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 1999*, San Diego, California, USA.
- [50]Spiliopoulou, M. and Faulstich, L. C. (1998) 'WUM : A Web Utilization Miner', *EDBT Workshop on Web Databases*, pp.1-7, Valencia, Spain.
- [51]Wu, K. L., Yu, P. S. and Ballman, A. (1998) 'SpeedTracer: A Web usage mining and analysis tool', *IBM Systems Journal on Internet Computing*, Vol. 37, pp. 89-105.
- [52]Pitkow, J. and Bharat, K. (1994) 'WebViz: A Tool for World Wide Web Access Log Analysis', *Advance Proceedings First International World-Wide Web Conference*. pp. 271-277.

Authors Biography

Ms. Chhavi Rana has an experience of over 5 years teaching data mining and web development subjects at various engineering Institutes. Currently she is working as Assistant Professor in Computer Science Engineering, University Institute of Engineering and Technology, Maharshi Dayanand University, Rohtak, Haryana. She is also pursuing her PhD. in Web Mining from National Institute of Technology, Kurukshetra, Haryana, India. She have been interested in area of Data mining and web development research from the past 3 years attending conferences and presenting papers related to this field. Besides Data mining, her research interests also include information management, Information Retrieval and ICT.