A Comprehensive Review of Web Mining Techniques

Gurpreet Singh[a], Ajay Kumar[b]

[a]Research Scholar(gps.srm@gmail.com), [b]Assoc. Professor (ajay.s@srmuniversity.ac.in)

[a,b]Deptt. of Computer Science Engg, SRM University, Sonipat, Haryana India

Abstract: Web Mining approaches are aimed at assisting web users in a number of application domains such as Web-semantics, E-Commerce, information retrieval, user customizable information and entertainment. The web-based data through internet, provides the basic computing resource to the web miners that can be mined for generating useful information with the aid of a number of data mining tools. Due to the elevated volume and very large allotmentof the data on the web, it is exigentjob to provide reliable and timely information to all the users. To address this issue, a number of web mining schemes has been used in various existing proposals in literature. Web-mining is the method to automatically determine and dig out information from Web documents. A comprehensive analysis of all presentedproposal in literature is provided to their contribution for improving web Mining. This comprehensive review of various presented proposals allows diverse users working in its domain to choose the best amongst them.

1. Introduction

The amount of information on the web is so large and in large number of formats that sometimes it becomes very difficult to find relevant and correct information. So,there is a requirement of some kind of methods and techniques to find exact and useful information [1-3]. The solution for this issue is provided through Web Mining. Web Mining is one of the applications of data mining, with the difference that instead of a database or warehouse we have to use the web as our data source [4-8]. Web Mining focuses on disparate types of content such as web-based documents, information related to how users access some resource, different types of links between web sites and other forms of Internet resources to achieve internal properties of data objects by incorporating different computer technologies such as statistical analysis, soft computing and machine learning to discover patterns that may be of use to users.

The four steps involved in web-mining which are classified: resource finding from web, data selection, pre-processing for removing outliers, generalization and web analysis [51]. Web Mining is classified into

three different categories, as follows: a) Web contentMining, b) Web Structure mining and c) Web usage mining [9-11]. These categories along with their important fields are illustrated in Figure 1.
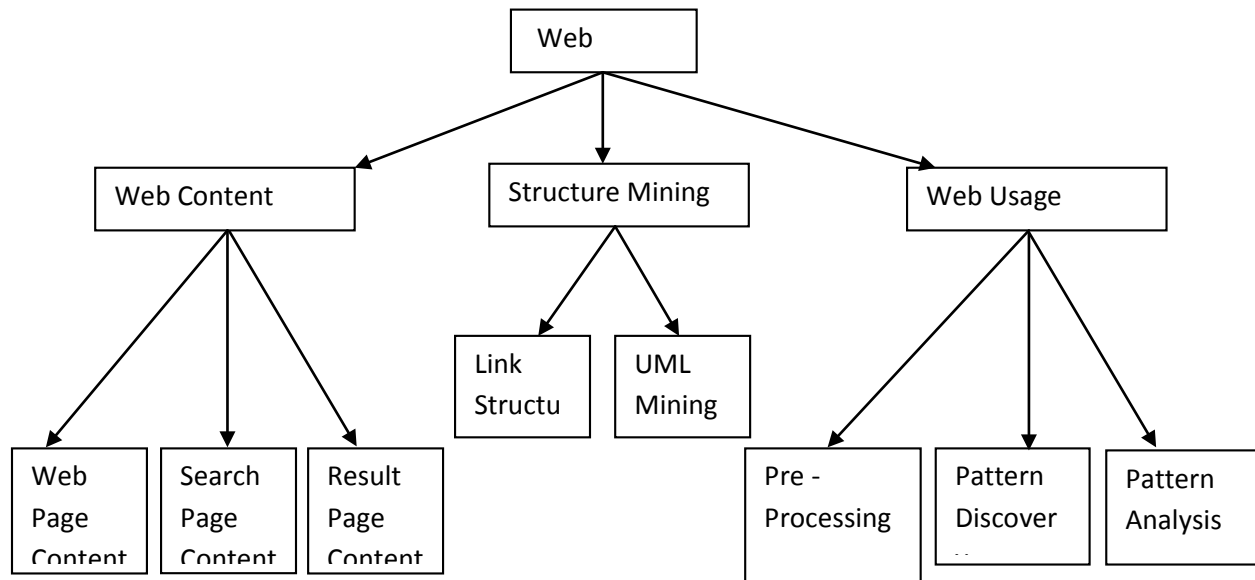

Figure 1: Categorization of Web Mining Techniques

2. Review of Web Mining Techniques

In this section we present a comprehensive review of existing web mining schemes. Most of these schemes have proposed generic solutions for overcoming some of the common research issues in the domain of Web Mining. The section also investigates the key contribution of these proposals in this area.

Yadav *et al.* [1] described web mining as a most important application of data mining. They discussed and applied various data mining techniques and tools for mining the huge data over the web. The main aim of web-mining is to discover and extract information over web; this knowledge can be very useful in the area of e-commerce and electronic industry. They have classified web mining into four categories as (i) web content mining, (ii) web structure mining, (iii) web usage mining and (iv) user profiles.

Mobasher*et al.* [3] have explained the main computer science aidended by web mining, they mentioned various applications, and highlighted future research directions. They provided taxonomy of web mining and summarize some of the key concepts related to the field. To start with the first concept is Page

Gurpreet Singh, Ajay Kumar

Ranking metric, to find out the quality of pages and relevant pages to query, authors mentioned two ranking matrices, i) PageRank, the key plan is that a page will have high rank if it is pointed to by many highly ranked pages, ii) Hubs and authorities, uses a score computed by an iterative procedure called HITS. Robot Detection and filtering is used to identify between robot behavior and human behavior, as it is necessary to understand usage patterns of users for industry not the robotic patterns. Interestingness is the measure of the user interests computed from the knowledge extracted from usage data and structure data. Preprocessing is a step to make web data suitable for mining, in this step all the useless and redundant information is removed. And the last one is Maximum-Flow models which are used to identify Web communities. They have specified few prominent applications like Amazon's broadapproach to bespoke customers experience, Google's usage of link structure for returning relevant pages to a query rather than content, Double-click which serves advertisements according to the interests of the users, America Online (AOL) applied web mining to better understand web communities and their interests and provide them with good products accordingly etc. They outlined some future directions for researchers in the area of web metrics and process mining.

R. Cooley *et al.*[19, 20, 21] have done extensive work in web mining area. Their work can be considered as the base in this area. As in [19] they provided an early taxonomy of mining: web content mining and web usage mining. In this paper they described all the basic terminologies necessary to understand the web mining concept. They also surveyed various tools and techniques for web content mining, web usage mining and to analyze discovered patterns from web. They provided a general architecture of a web usage mining system. They highlighted the main issues and problems in this area for further research. In order to apply web mining algorithms over huge amount of web usage data some kind of preprocessing is required. In [20] they have presented several techniques for data preparation and algorithms for change raw web servers log into user sessions files that are suitable for performing Web usage mining on them. In [21] they proposed a transaction identification model, and compared it with already existing models. They discussed three modules in their paper, Reference Length Module, Maximal Forward Reference Module and Time Window Module. Finally, they used the tool Web Miner system to identify transactions.

Simranjeet*et al.* in [29] have reviewed web mining and compared web mining with data mining. In this paper they have discussed the three categories of web mining in every aspect like type of data required, suitable representation of data, methods in each category, scope application areas and challenges. Khosla et al. in [42] survey the research done in the era of web-mining. They tried to remove some confusion regarding the usage of the term web mining. They explored the connection between the three categories

Gurpreet Singh, Ajay Kumar

of web mining i.e. web structure mining, web content mining and web usage mining and the related agent paradigm.

Table 1 provides a basic comparative analysis b/w the different types of web mining. Comparison is based parameters such as type of data, representation, various methods and different application categories.

Table 1: Web mining categories

	Web Mining			
	Web Content Mining		Web Structure Mining	Web Usage Mining
	IR View	DB View		
View of Data	- Unstructured - Semi structured	- Semi structured - Web site as DB	- Links structure	- Interactivity
Main Data	- Text documents - Hypertext documents	- Hypertext documents	- Links structure	- Server logs - Browser logs
Representation	- Bag of words, n-grams - Terms, phrases - Concepts or ontology - Relational	- Edge-labeled graph (OEM) - Relational	- Graph	- Relational table - Graph
Method	- TFIDF and variants - Machine learning - Statistical (including NLP)	- Proprietary algorithms - ILP - (Modified) association rules	- Proprietary algorithms	- Machine Learning - Statistical - (Modified) association rules
Application Categories	- Categorization - Clustering - Finding extraction rules - Finding patterns in text - User modeling	- Finding frequent sub-structures - Web site schema discovery	- Categorization - Clustering	- Site construction, adaptation, and management - Marketing - User modeling

In the following sections, the various categories and existing literature in the domain of Web Mining has been explored. These sections are divided based on different domains where web mining can be applied. In each section a detailed review of existing techniques on web mining are discussed based on their usability in each domain.

2.1 Web Content Mining

A content of web site consists of different data types and data structures used for representing web based information. This content may be further classified into visual content such as text images, multimedia content such as audio and video, and other structured records such as lists and tables. Web Content Mining is a process which is used for analyzing the data from web documents and then extracting some useful information from the content. In the past there have been quite a lot of significant techniques that have been proposed to perform the function of extracting information from traditional domains such as signals and images.

Gurpreet Singh, Ajay Kumar

The web content mining model proposed by the authors of [58] [59] for performing information retrieval involves the domain of graph analysis, where nodes represent web pages and edges represent hyperlinks. This information retrieval model is based on object oriented paradigm. They have addressed an important issue of finding frequent item sets and improving the quality of ranking in a query. The proposed information retrieval model pertains to characteristics like modularity, flexibility, broadness and applicability. There are few steps in web content mining process as listed in the figure below:
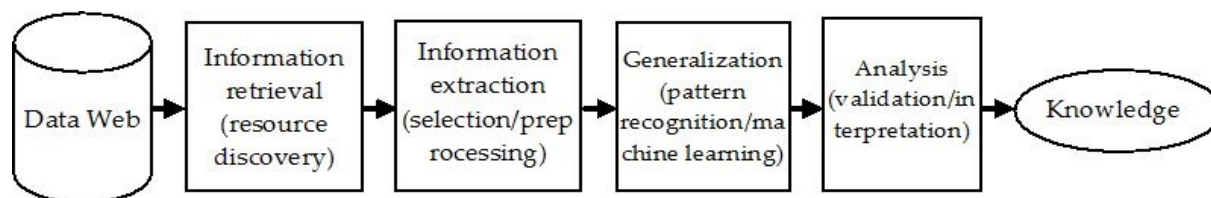

Figure2: The Step wise procedure of web content mining

In web content mining there are two approaches: 1) Agent Based Approach, 2) Database Approach. The diagram below shows the classification for these approaches.
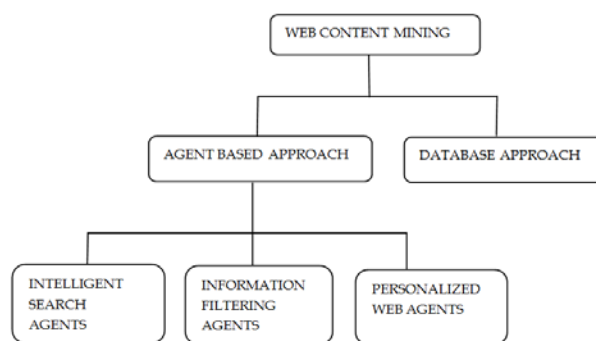

Figure 3:The various approaches of Web Content Mining.

2.1.1 Agent Based Approach

Its main focus is on searching the relevant information from web. It is further divided into three parts:

a. Intelligent search agents – It basically performs automatic search with a particular query.

b. Information filtering: / categorizing agents – It is used for filters data.

c. Personalized web agents: It discoverdocuments which are linked to the user Profiles.

2.1.2 Database Approach

It contains databases, which consist tables, attributes, and schemes. Its main focus is on technique for organizing the semi-structured data on web in-to collection of resources. The benefit of this is that we can use data mining techniques to analyze it, Example areweb querying system andmultilevel database.

Zheng *et al.* in [37] have discussed the structural semantic entropy. This technique used tolocate data on web page. It is an automated extraction algorithm that can extract useful pair of attributes from different web sites. The structural semantic entropy method is based on the extent of density occurrence of the interested information. The authors have experimented and the results were encouraging.

Rohilla*et al.* in [36] have implemented the web content mining in their project called INFUSION. This concept has been applied on managing the social websites. As HTML web pages consist of various elements, some are nested and some are not. So HTML document have a tree like structure. Any kind of changes in the web page is reflected in the structure. In their implementation they have used an algorithm called Tree Edit Distance algorithm to make sure that changes get reflected in the tree structure of the document. The purpose of their work is based on the fact that web content mining is useful in extracting useful information, new changes and updates in their social network etc from users' various social accounts and showing it at a common place. They have used technique called Web Scraping which is a process of extracting information from web in structured way. So data is fetched from various social web sites and showed to user in a well-structured way.

2.2 Web Usage Mining

The term web usage mining is the application of one or more computing techniques to perform the functionality of web mining so as to obtain some information from the web based data that depicts some interesting usage patterns [60]. This information may be utilized so as to understand the user behavior

Gurpreet Singh, Ajay Kumar

and improve the usability of any web-based applications [61]. The extraction process may reveal different types of data that includes some web logs, database transactions and the streams of click that are done at some web site or at some group of similar web links. Another issue that is related to web usage mining, the implementation of privacy concern of the website users this topic is also being extensively researched in this domain.

Web usage mining itself can be further classified based upon the kind of usage data considered:

  a.   Web Server Data: its user log, like IP address, page reference and access time etc.

  b.   Application level Data: This information can be used for personalization of the web sites.

  c.   Application server Data: these are commercial application servers used for e-commerce based applications.

In the process of web usage mining the figure below shows the steps involved, initially the raw data is processed, just to remove useless and redundant data, then to find the usage patterns some useful data mining techniques are applied. These steps in the form of steps are shown in the diagram below.
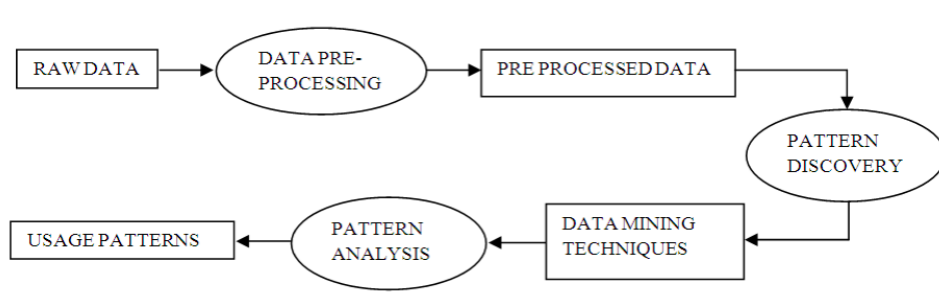

Figure5. The process of web usage mining

Raju *et al.* [4] highlighted the significance of Web personalization. The term web personalization means that it enables the dynamic insertion, customization or suggestions that can be useful for users based on user's implicit behaviour such as item purchased and page viewed, and explicit behavior like rating and preferences etc. As Web is an "ocean" of information, user faces the problem of finding relevant, needed and useful information.

Gurpreet Singh, Ajay Kumar

Sahu*et al.* [5] have discussed Data cleaning used in Data Preprocessing and Web Usage Mining. They also provided a summary of Log Cleaner that filters out plenty of some irrelevant, inconsistent data based on the common of their URL. Data cleaning is very important process because noisy data can cause difficulty during page ranking. They mentioned various web usage mining techniques like Two-Level clustering method, Noise Detector, Community Detection, Effective and scalable technique and EPLog Cleaner filtering method.

Naga *et.al.* [4] have discussed web server log files that are plain text files which store click stream data of users. The various log files comes in the following file formats:a) Common Log Format (CLF),   b) Microsoft Internet Information Services (IIS) Log file Format, c) World Wide Web (World Wide Web Consortium) Extended Log file Format and d) NCSA Common Log file Format.

Srivastava *et al.* in [11] have provided an advanced survey of speedily growing region of web usage mining. Web usage mining allow administrators to better understand web traffic behavior, and this information is in turn is very useful for developing policies, providing better content to user, e-commerce, resolving security issues and so on. Web usage mining have three phases, namely pattern analysis, pattern discovery and preprocessing. Due to increasing interest in this area a lot of effort has been put in research work and application development. Authors have provided taxonomy of major application areas of web usage mining and various commercial products available. They categorized web usage mining into five major application areas: Personalization, System improvement, Site Modification, Business Intelligence and Usage Characterization, and explained each one in detail. They also provided the commercial product according to the above categories, like Web SIFT, WUM [12], Speed Tracer [13], Weblog Miner [14] as web usage mining products, Site Helper [15], Letizia [16], Web Watcher [17], Mobasher [18], Analog and Krishnapuram as Personalization Products and so on. They also gave a brief overview of the Web SIFT system as an example.

Jaideep *et.al.* [9] have explained the process of web usage mining. They mentioned that after the data passes through the Preprocessing stage, algorithms from Data Mining viz. Classification, Clustering, Association Rules, Statistical Analysis etc. are to be applied.   The Pattern Mining phase to be applied depends upon the area in which it is to be applied. Often, Visualization techniques are utilized to show these trends in data. Web Log Data has been utilized across all domains.

2.3 Web Structure Mining

The term web structure mining can be defined as the process of discovering information related to the structure of a web page from the world wide web. To perform web structure mining the web can also be considered as a collection of semi structured documents. The graphical structural representation of any generic web site may typically consist of the nodes being the web pages and edges connecting two related pages may be represented by hyperlinks.
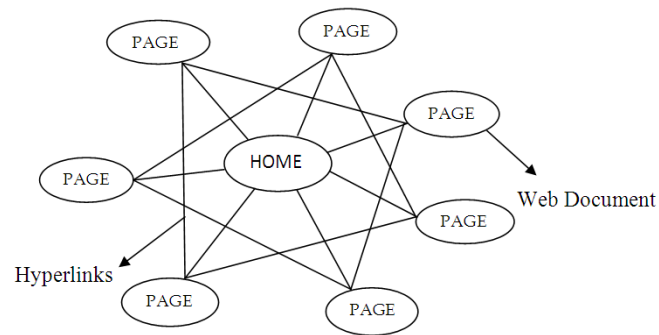
Figure 6: The structure of World Wide Web

Costa Da *et al.* in [23] have surveyed the research area of web mining and they mainly focused on the category web structure mining. They have discussed the concept of Link mining and how it can be applicable to web mining. They also discussed block-level link mining issues. As they mainly focused on web structure mining, they reviewed two well-liked algorithms applied in web str. mining: Page Rank,HITS.

Krishna Murthy *et al.* in [26] proposed an algorithm for the classification of XML URLs based on their semantic. As a huge amount of information is there on web, it is very important to know what information is to be considered and what information is to be avoided, so some kind of method should be there to classify the information. HTML based web pages are not well structured so XML based web pages are a better option for web mining. Authors categorized web page into four categories.

3. Conclusion

The use of web mining techniques can help to reduce the amount of information dissemination thereby providing correct and up to date information for the users. This work has provided a detailed review about existing web mining techniques along with their application areas. The reviewed proposals have been classified into different sub categories and a comprehensive classification of existing literature into these categories has also been done. The work also provides an overview about important issues that need

attention of research community. Finally, a brief description about some important tools being used in the web mining also presented. The presented work can be considered as a useful reference point for all researchers in the region of web mining.

References:

1. Yadav S., Ahmad K., Shekar J, "Analysis of Web Mining Applications and beneficial Areas", IIUM Engineering Journal. Vol. 12, No.2, 2011.

2. Vijiyarani S., Suganya E., "Research Issues in Web Mining", International Journal of Computer-Aided Technologies, Vol. 2, No. 3, 2015.

3. Mobasher, Bamshad, Robert Cooley, and Jaideep Srivastava. "Automatic personalization based on web usage mining." *Communications of the ACM* 43.8 (2000): 142-151.

4. Raju Y.,SureshBabu D., "A Novel Approach in Web mining Techniques in case of web Personalization", International Journal of Research in Computer Applications and Robotics, Vol. 3, Issue 2, 2015.

5. Sahu Shashi, Sahu L., "A Survey on Frequent Web Page Mining with Improving Data Quality of Log Cleaner ", International Journal of Advanced Research in Computer Engineering & Technology, Vol. 4, Issue 3, 2015.

6. Sakthipriya C., Srinaganya G., Sathiaseelan J. G. R., "An Analysis of Recent Trends and Challenges in Web Usage Mining Applications", International Journal of Computer Science and Mobile Computing, Vol. 4, Issue 4, 2015.

7. Ziang Li, Wei Xu, Likuan Zhang, Raymond Y.K. Lau, "An ontology-based Web mining method for unemployment rate prediction", Decision Support Systems, Elsevier, 2014.

8. Thorleuchter D., Van den Poel D., "Web Mining based extraction of problem solution ideas", Expert System with Applications, ScienceDirect, Pg No. 3969-3969, 2013.

9. Thorleuchter, D., Van den Poel, D., &Prinzie, A. (2010a). Mining ideas from textual information. Expert Systems with Applications, 37(10), 7182-7188.

10. Juan D. Velasquez, Luis E. Dujovne, Gaston L'Huillier, "Extracting Significant website Key Objects: A Semantic Web mining approach", Engineering Applications of Artificial Intelligence , Elsevier, 1532-1541, 2011.

11. Srivastava J., Cooley R., Deshpande M., Tan P. N., "Web Usage Mining: Discovery and Applications of usage Patterns from Web Data", SIGKDD, ACM, Vol. 1, Issue 2, 2000.

12. Myra Spiliopoulou and Lukas C Faulstich. WUM: A web utilize miner. In EDBT Workshop Web DB98, Velencia, Spain, 1998. Springer Verlag.

13. Kun-Lung Wu, Philip S Yu, and Allen Ballman. Speed-Tracer: A web usage mining Analysis tool. IBM System Journal, 37(1), 1998.

14. Robert Cooley, Pang-Ning Tan and Jaideep Srivastava. Discovery of interesting usage patterns from web data. Technical Report T R 99-022, University of Minnesota, 1999.

15. D. S. W. Ngu and X. Wu. Sitehelper: A localized agent that helps incremental exploration of the World Wide Web. In 6th International World Wide Web Conference, Santa Clara, CA, 1997.

16. H. Lieberman, "Letizia: An Agent that assists web browsing", In Proc. of the 1995 International Joint Conference on Artificial Intelligence, Montreal, Canada.1995.

17. T. Joachims, D. Freitag, and T. Mitchell, "Webwatcher: A tour guide for the world wide web", In the 15th International Conference on Artificial Intelligence, Nagoya, Japan, 1997.

18. BamshadMobasher, Robert Cooley and Jaideep Srivastava, "Creating adaptive web site through usage based clustering of urls", In Knowledge and Data Engineering Workshop, 1999.

19. Cooley R., Mobasher B., Srivastava J., "Web Mining: Information and pattern discovery on the world wide web", In International Conference on Tools with Artificial Intelligence, pages 558-567, Newport Beach. 1997. IEEE.

20. Cooley R., Mobasher B., Srivastava J., "Data preparation for mining world wide web browsing patterns", Knowledge and Information Systems, 1(1), 1999.

21. Cooley R., Mobasher B., Srivastava J., "Grouping Web Page References into Transactions for Mining World Wide Web Browsing Patterns".

22. Li Yuefeng, Zhong Ning, "Web Mining model and its applications for information gathering", Elsevier, 2004.

23. Da Costa, Miguel Gomes, and Zhiguo Gong. "Web structure mining: an introduction." *2005 IEEE International Conference on Information Acquisition*. IEEE, 2005.

24. MurlidharA ,Pattabiraman V, "An Efficient Association Rule Based Clustering of XML Documents", 2nd International Symposium on Big Data and Cloud Computing, 2015.

25. Lee Kun Chang, Lee Sangjae, "Interpreting the web-mining results by cognitive map and association rule approach", Elsevier, 2010.

26. Krishna Murthy .A ,Suresha, "XML URL Classifications based on their semantic structure orientation for web Mining Applications", International Conference on Information and Communication Technologies, 2014.

27. Z. Hui, Q. Shigang, L. Jinhua and C. Jianli, "Study on Website Search Engine Optimization", International Conference on Computer Science and Service System, 2012.

28. K. Sharma, G. Shrivastava and V. Kumar, "Web Mining: Today and Tomorrow", IEEE, 2011.

29. Simaranjeet Kaur, Kiranbir Kaur, "Web Mining and Data Mining: A comparative Approach", International Journal of Novel Research in Computer Science and Software Engineering, 2015.

30. R. Lokeshkumar1, R. Sindhuja2, Dr. P. Sengottuvelan, "A Survey on Pre-processing of Web Log File in Web Usage Mining to Improve the Quality of Data" International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 8,August 2014.

31. Mitali Srivastava, Rakhi Garg, P. K. Mishra, "Preprocessing Techniques in Web Usage Mining: A Survey" International Journal of Computer Applications (0975 – 8887) Volume 97– No.18, July 2014.

32. Ananthi.J, "A Survey Web Content Mining Methods and Applications for Information Extraction from Online Shopping Sites", International Journal of Computer Science and Information Technologies, Vol. 5 (3), 2014.

33. Ananthi.J , "A Survey Web Content Mining Methods and Applications for Information Extraction from online shopping sites" , International Journal of Computer Science and Information Technologies (IJCSIT), Voume 5 , Issue 3 , ISSN :0975- 9646.

34. R.Malarvizhi, K.Saraswathi , "Web Content Mining Techniques Tools & Algorithms – A Comprehensive Study" , International Journal of Computer Trends and Technology (IJCTT) , volume 4 ,Issue 8, August 2013, ISSN: 2231-2803.

35. E.Raju , K.Sravanthi , "Analysis of Social Networks Using the Techniques of Web Mining" , International Journal of Advanced Research in Computer Science and Software Engineering.

36. Pooja Rohilla, Ochin Sharma, "Web Content Mining: An Implementation on Social Websites", International Journal of Advanced Research in Computer and Communication Engineering, Vol 4, Issue 7, July 2015.

37. Xiaoqing Zheng,YilingGu,Yinsheng Li, "Data Extraction from Web Pages Based on Structural Semantic Entropy", International World Wide Web conference Committee (IW3C2),April 2012,pp.93-102.

38. Wei Liu, XiaofengMeng ,WeiyiMeng , "ViDE: A Vision-based Approach for Deep Web Data Extraction", IEEE Transactions on Knowledge and Data Engineering, Volume:22 , Issue: 3, March 2010,pp. 447 – 460.

39. Ali Ghobadi,MaseudRahgozar, "An ontology based Semantic Extraction Approach for B2C eCommerce",The International Arab Journal of Information Technology Vol.8, No. 2,April 2011,pp.163-170.

40. Jun Zhu, ZaiqingNie, Ji-Rong Wen, Bo Zhang, Wei-Ying Ma, "2D Conditional Random Fields for Web Information Extraction",Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 2005.

41. GengxinMiao,JunichiTatemura, Wang-pin Hsiung, ArsanySawires, Louise E.Moser, "Extracting Data Records from the Web Using Tag Path Clustering",International World Wide Web conference Committee (IW3C2),April,2009,pp.981-990.

42. Kosala, Raymond, and Hendrik Blockeel. "Web mining research: A survey.", ACM Sigkdd Explorations Newsletter 2.1 (2000): 1-15.

43. Zahra Hojati, RozitaJamiliOskouei, "A Comprehensive Comparison between Web Content Mining Tools: Usages, Capabilities and Limitations", International Congress of Electrical Engineering Computer Science & Information Technology, Iran, 2015.

44. S. Balan, P. Ponmuthuramalingam, (2013), "A Study of Various Technique of Web Content Mining Research Issues and Tools", International Journal of Innovation Research and Studies (IJIRS), Vol. 2 Issue 5, pp. 507-517.

45. M.Karpagam, R.Sasikala, (2013), " Analysis of Web Content Mining Tools" , International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 3, Issue 12, pp. 124-130.

46. Chhavi Rana, "A Study of Web Usage Mining Research Tools", *Int. J. Advanced Networking and Applications Volume:03 Issue:06 Pages:1422-1429 (2012) ISSN : 0975-0290.*

47. RoopRanjan, SameenaNaaz, Neeraj Kaushik, "Web Miner: A Tool for Discovery of Usage Patterns From Web Data", International Journal on Computer Science and Engineering (IJCSE), Vol. 5 No. 05 May 2013.

48. Naga Lakshmi, Raja Sekhara Rao and Sai Satyanarayana Reddy, "An Overview of Preprocessing on Web Log data for Web Usage Analysis", IJITEE, ISSN: 2278-3075, 328Vol.2, Issue-4, pp.274-279, March 2013.

49. JaideepSrivastava , Robert Cooley, Mukund Deshpande and Pang- Ning Tan, "Web Usage Mining : Discovery and Applications of Usage Patterns from Web Data", ACM SIGKDD, Vol1, Issue 2, pp.1- 12,Jan2000.

50. Neha Goel, Sonia Gupta, C. K. Jha, "ANALYZING WEB LOGS OF AN ASTROLOGICAL WEBSITE USING KEY INFLUENCERS", International Research Journal, Vol. 05 No. 01, 2015.

51. Mr. DushyantB.Rathod, Dr.Samrat Khanna, "A Review on Emerging Trends of Web Mining and its Applications" ISSN: 2321-9939.

52. Yuewen Liu, Kwok Kee Wei, Huaping Chen, "A meta-analysis on the effects of online auction design options: The moderating effect of value uncertainty", Elsevier, Electronic Commerce Research and Applications 9, 507–521, 2010.

53. Xenia Papadomichelaki, GregorisMentzas, "e-GovQual: A multiple-item scale for assessing e-government service quality", Elsevier, Government Information Quarterly 29, 98–109, 2012.

54. Berners-Lee, Tim, Robert Cailliau, Ari Luotonen, Henrik Frystyk Nielsen, and Arthur Secret. "The world-wide web.", In *Human-computer interaction*, pp. 907-912. Morgan Kaufmann Publishers Inc., 1995.

55. Ramakrishna, Mahesh Thylore, LathaKolalGowdar, MalateshSomashekarHavanur, and BanurPuttappaMallikarjunaSwamy. "Web mining: Key accomplishments, applications and future directions.'', In *Data Storage and Data Engineering (DSDE), 2010 International Conference on*, pp. 187-191. IEEE, 2010.

56. Hu Jiming, "Study on Web-based Model of E-business." , In *Journal of Research and Work of Library*, vol.2, pp 31-33, 2006.

Gurpreet Singh, Ajay Kumar

57. Meirong, Tian, and Chen Xuedong. "Application of Agent-based Web Mining in E-business.", In*Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2010 2nd International Conference on*, vol. 1, pp. 192-195. IEEE, 2010.

58. Salton, Gerard, and Christopher Buckley. "Term-weighting approaches in automatic text retrieval.", *Information processing & management* 24, no. 5 (1988): 513-523.

59. Cormen, Thomas H., Charles E. Leiserson, and Ronald L. Rivest. C. Stein "Introduction to Algorithms.", *MIT Press* 5, no. 3 (2001): 55.

60. Berendt, Bettina. "Using site semantics to analyze, visualize, and support navigation. *"Data Mining and Knowledge Discovery* 6, no. 1 (2002): 37-59.

61. Srivastava, Jaideep, Robert Cooley, Mukund Deshpande, and Pang-Ning Tan. "Web usage mining: Discovery and applications of usage patterns from web data."*ACM SIGKDD Explorations Newsletter* 1, no. 2 (2000): 12-23.