A Hand to Hand Taxonomical Survey on Web Mining

Neha Sharma
Amity University,
Noida(U.P.),201303,India

Sanjay Kumar Dubey
Amity University,
Noida(U.P.),201303,India

ABSTRACT

The data mining techniques exploitation in the field of web is referred as web mining. The enormous data is present at the websites and this need to be tackled well with the help of different data mining techniques. Searching, puling data together and analyzing the data are the main focus of web mining. The application of web mining is in the field of e-commerce and e-learning, web search, database, AI, information retrieval, system improvement etc. Information extraction from the web documents is a typical task and can be done efficiently after the through study of mining. This paper would facilitate to comprehend the concept of web mining by analyzing the facts retrieved from various sources. The paper presents the literature survey on web mining. It also explains the detailed view of three kinds of web mining techniques viz. web content mining, web structure mining and web usage mining. For the survey, different papers are analyzed and then presented as the study of web mining and its subtasks.

Keywords
Web mining,web content mining,web structure mining,web usage mining,information retrieval, information extraction.

1. INTRODUCTION
Today, each and everything is available on internet and it's widely used by people. The internet is wide, diverse and lively and hence it is being used in different fields also. A lot of information is present online and can be accessed by anyone at any point of time. So, the task of tackling this huge data is very much important and this is being done with the help of the web mining. Web mining refers to the data mining application and related techniques that are used to automatically discover and extract information from web documents and services [1].

Web mining can be broadly categorized into three types:

i. Web usage mining aims to automatically discover and analyse patterns in click stream and associated data collected or generated as a result of user interactions with web resources, on one or more web sites. Behavioural patterns and profiles of users interacting with a web site are captured, modelled and analysed in order to improve services [2].

ii. Web structure mining is the process of using graph theory to analyze the node and connection structure of a web site. According to the type of web structural data, web structure mining can be divided into two kinds:

- Extracting patterns from hyperlinks in the web: a hyperlink is a structural component that connects the web page to a different location.

- Mining the document structure: analysis of the tree-like structure of page structures to describe HTML or XML tag usage [3].

iii. Web content mining is the process of extracting useful information from the contents of the web documents. Content data corresponds to the collection of facts a web page designed to convey to the user [4].

The design of paper is very simple and is being divided into different sections. Section 2 concentrates on the primary research done on the web mining area. A lot of work has been published in this field and few researchers also focused on the survey of web mining. So, the primary goal of this section is to go through important works being done in this area.

Section 3 is research methodology used for the survey. It basically provides an idea that how papers are searched and relevant papers are collected out of them depending upon the three categories of web mining. The classification strategy is critical analysis and rigorous study of the research papers from different sources.

Section 4 totally focuses on the web mining, its overview, types, web mining subtasks and web mining taxonomy. The main motive of the section is thorough study of web mining in all aspects given till date.

In the section 5, advantages and limitations of web mining are explained. Web mining is a very big term and has lots of applications too but in all there are also some limitations of this technique with some advantages. This section focuses on the same.

Finally, overall conclusion of the survey on web mining, after the thorough study of so many research papers is being provided. Each and every research has their point of view and different way of analyzing the web mining. After studying all aspects of web mining, a brief conclusion about it is being provided.

2. LITERATURE SURVEY
Kosala et al surveyed the research in the area of web mining [5]. The paper explores the connection between web mining categories and the related agent paradigm. This paper focuses on representation issues, on the process, and on the learning algorithm, and the application of the recent works as the criteria.

Chang et al describes that Information extraction (IE) from semi-structured web documents is a critical issue for information integration systems on the Internet [6]. The discovery of repeated patterns is realized through a data structure call PAT tree. The paper also focuses that incomplete patterns are further revised by pattern alignment to comprehend all pattern instances.

Li et al proposed a step from data mining to its application, which is called the second data mining or knowledge presentation and management [7].

Yusifov et al proposes an intelligent model to discover and analyze useful knowledge from the available web log data [8]. They explain that discovering hidden and meaningful information about web user's usage patterns is critical to determine effective marketing strategies. The task of mining useful information becomes more challenging when the Web traffic volume is enormous and keeps on growing.

Punnin et al proposed two new XML applications, XGMML and LOGML [9]. XGMML is a graph description language and LOGML is a web-log report description language. They generated a web graph in XGMML format for a web site using the web robot of the WWWPal system. The web-log reports are also generated in LOGML format for a web site from web log files and the web graph.

Srivastava et al has attempted to provide an up-to-date survey of the rapidly growing area of web usage mining [10]. They explain that how web usage mining raises some hard scientific questions that must be answered before robust tools can be developed. This paper has described such challenges.

Seydim explained the agent paradigm along with the main applications and the use of this technology in data mining [11]. This paper explains that inherent parallelism and complexity of the classification and discovering patterns from large amounts of data can be delegated to intelligent software agents.

Cooley et al proposed a definition of web mining, and developed taxonomy of the various ongoing efforts related to it [12]. The paper also presented a survey of the research in this area and concentrated on Web usage mining.

Pater et al defined web mining and presented an overview of the various research issues, techniques and development efforts [13]. The paper briefly described the strategies for pattern discovery techniques in web mining. The paper say key component of web mining is the mining process itself.

Eirinaki et al presented a survey of the use of web mining for Web personalization [14]. This paper presented modules that comprise a web personalization system, emphasizing on the web usage mining module.

Fu et al proposed an approach for reorganizing web sites based on user access patterns [15]. The approach consists of three steps: preprocessing, page classification, and site reorganization. An algorithm for reorganizing Web sites has also been developed.

Zheng et al developed a framework for web mining, based on a general architecture that decouples input data, learning method, evaluation method, and visualization [16].

Srivastava et al provided a brief overview of the accomplishments of the field, both in terms of technologies and applications, and outlines key future research directions [17].

Alhawamdeh described the contribution of intelligent systems in designing the strategic website for small business to satisfy the customers need [18]. This paper concludes that data mining applied to the web has the potential to be quite beneficial.

Jain et al. explore the data mining in privacy preservation in which the knowledge discovery process of finding the useful information and patterns out of large database [19]. In recent times data mining has gained immense importance as it paves way for the management to obtain hidden information and use them in decision-making for exposing it for the betterment of web mining.

Júnior et al provided an introduction of Web mining as well as a review of the Web mining categories [20]. This paper also focuses on one of its categories: the Web structure mining. The paper introduces link mining and review two popular methods applied in Web structure mining: HITS and Page Rank.

3. RESEARCH METHODOLOGY

Web mining is an immense term and so many researchers have worked in this field since 1990. While searching for web mining, we found so many relevant papers on this topic. The relevant material is scattered across different journals and conferences. To provide the all inclusive bibliography of the web mining techniques, most of the papers are taken from the database of following journals and conferences database:

- ACM
- IEEE
- Springer
- Elsevier
- Scopus Database
- Google Scholar

The papers are analyzed and separated according to the one of the three categories of web mining viz. usage mining, content mining and structure mining. The methodology opted behind the classification of the web mining techniques is its basic analysis and its scrupulous study from different sources.

4. WEB MINING

Web mining is the use of data mining techniques to automatically discover and extract information from web documents and services [4]. Since the online information and online access of data is tremendously increasing day by day, as a result the interest of researchers in this field also increasing. This paper explores web mining is and what are its basic types.

4.1 Web Mining Subtasks

Pazzani [5][7][21] crumbled web mining into four subtasks, which are as follows:

i. Resource finding: it's mainly the task of retrieving intended web documents. It's the process of retrieving the data that is either online or offline from the web sources like text, relational data and semi structural data like XML.

ii. Information selection and pre-processing: It means automatically selecting and pre-processing specific information from retrieved web resources. It's basically any kind of transformation processes of the original data retrieved in the IR process. Transformations could be either a kind of pre-processing such as removing stop words, stemming, etc. or a pre-processing aimed at obtaining the desired representation such as finding phrases in the training corpus, transforming the representation to relational or first order logic form, etc

iii. Generalization: automatically discovers general patterns at individual web sites as well as across multiple sites. Machine learning or data mining techniques are used for generalization.

iv.   Analysis: It means validation and/or interpretation of the mined patterns. Humans play an important role in the information or knowledge discovery process on the web since the web is an interactive medium. This is especially important for validation and/ or interpretation.

4.2 Web Mining Taxonomy

Web mining refers to the overall process of discovering potentially useful and previously unknown information or knowledge from the web data. It implicitly covers the standards process of knowledge discovery in databases (KDD) [22]. Web mining is basically categorized into 3 types:

4.2.1. Web Structure Mining

Web structure mining aims to discover useful knowledge from hyperlinks, which represent the structure of the web. Hyperlink is a link that exists in a web page and refers to another region in the same web page or another web page. The most popular application of web structure mining is to calculate the importance of web pages. This kind of application is used in Google search engine to order its search results. A web structure mining algorithm, Page Rank, is invented by Google founders [59].

4.2.2 Web Content Mining

Web content mining extracts or mines useful information or knowledge's from web page contents. There are two categories of web content mining: structured data extraction and text mining. The idea of structured data extraction is that many web site display important information retrieved from their database using some fixed templates. We can identify those templates by finding repeated patterns in web pages. Apart from structured data, the Web also contains a huge amount of unstructured text, written in natural language. One of the common tasks in text mining is to extract people's opinions or sentiments expressed in product reviews, forum reviews, social networks and blogs[17].

4.2.3 Web Usage Mining

Web usage mining aims to capture and model behavioral patterns and profiles of users who interact with a web site. Such patterns can be used to better understand the behaviors of different user segments, to improve the organization and structure of the site, and to create personalized experiences for users by providing dynamic recommendations of products and services. Unlike two previous web mining tasks, the primary data source for web usage mining is web server access log, not the web pages[23].

4.3 Distribution of articles by year of publication

The distribution of articles by year of publication is shown in Fig. 1. The number of publications considered in the paper has increased a lot from the year 1995 to 2000 and there are comparatively less papers taken form the time period in the 2000 to 2005.
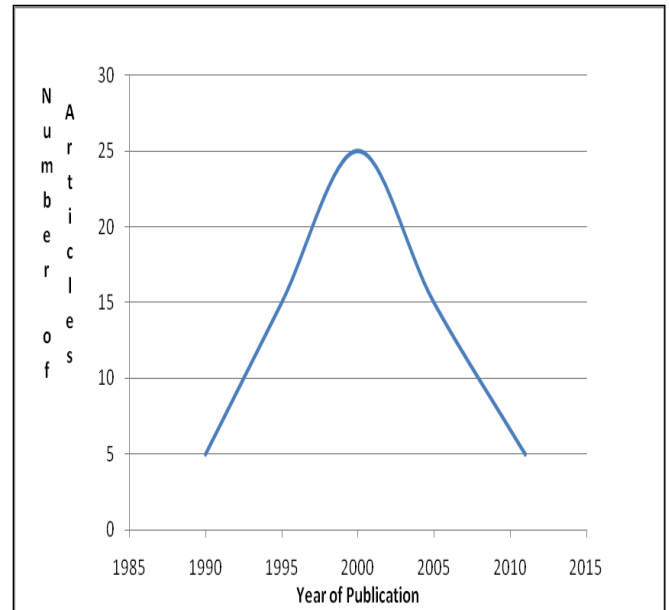

Figure 1: Distribution of Articles by Year of Publication

5. MERITS AND DEMERITS OF WEB MINING

Web mining is now being an old concept and after the use of web mining in real world, we have came to know about pros and cons of web mining technique. Merits of web mining are listed below:

- Web mining has enabled e-commerce to do personalized marketing, resulting in higher trade volumes.
- The government using web mining to classify threats and fight against terrorism.
- Predicting capability of the mining application can benefits the society by identifying criminal activities[15].
- The companies can establish better customer relationship by giving them exactly what they need. Companies can understand the needs of the customer better and they can react to customer needs faster [16].

Disadvantages of web mining are listed below:

- The most criticized ethical issue involving web usage mining is the invasion of privacy. Privacy is considered lost when information concerning an individual is obtained, used, or disseminated, especially if this occurs without their knowledge or consent [17].
- The obtained data will be analyzed, and clustered to form profiles; the data will be made anonymous before clustering so that there are no personal profiles [17]. Thus these applications de-individualize the users by judging them by their mouse clicks.

Table 1:Web Content Mining

S.No.	Method/Tool/Technique	Application	Year	Source
1.	Intelligent Search Agents (Harvest system)	building topic specific content Index	1994	[24] [25]
2.	Instance based information of source data(TSIMMIS)	Integrates Heterogeneous data	1994	[26]
3.	FAQ-FINDER	Extraction of Frequently asked questions(FAQ) from web files	1995	[27]
4.	ILA	Extracts information from unfamiliar resources	1995	[28]
5.	Shopbot	Extracts product information from web vendors	1996	[29]
6.	Web Query System(Lorel)	Lightweight object query language to query semistructured data	1996	[30]
7.	Schema based querying over websites(Weblog)	Retrieval of information from web document.	1997	[31]
8.	Multilevel Database Approach (ARANEUS System)	Unstructured HTML documents are analyzed to extract their structure	1997	[32]
9.	Visual content Description	Content based image retrieval	1999	[33]
10.	Relevance feedback Algorithm	Content based image Retrieval	2003	[34]
11.	Web structure together with summarization techniques	Semantic virtual document	2005	[35]
12.	Graph-based overlapping, Clustering Algorithm	Meta-search engine called WISE	2006	[36]
13.	Query-URL co-clustering	Categorize queries and URLs related to special web site	2007	[37]
14.	SPARSE technique	Localized CBIR System	2008	[38]
15.	Signed approach and full word matching	Retrieval of documents takes less time and less space	2009	[39]
16.	Agent Technology	Semantic web	2010	[22]
17.	PageRank and TrustRank algorithm	Security in e-commerce websites	2011	[58]

Table 2:Web structure Mining

S.No.	Method/Tool/Technique	Application	Year	Source
1.	Warehouse of Web Data (WHOWEDA project)	To design the tools and techniques for web data mining	1999	[40]
2.	HITS	Discovering authoritative sources in a Hyperlinked environment	1999	[41]
3.	Data mining and Machine learning	Exploiting the graph structure of the Web	2002	[42]
4.	Clustering, self-organized map	Mapping user navigation patterns	2003	[43]
5.	Heuristic Approach	Hyperlink selection for portal page	2004	[44]
6.	Sequence Alignment Method	Mining navigation Patterns	2004	[45]
7.	Design bookmark Structure	Bookmark	2005	[46]
8.	HITS and PageRank	Page blocks	2005	[20]
9.	Frequent access path Identification algorithm, Fuzzy set theory	Mining Web Browsing patterns for e-commerce	2006	[47]
10.	Various dimensionality Reduction techniques	To extract the implicit structures hidden in the web hyperlink connectivity	2007	[48]
11.	Graph clustering Algorithm	Mining the community structure of a graph	2009	[49]
12.	PageRank and TrustRank algorithm	Security in e-commerce websites	2011	[22]

- Another important concern is that the companies collecting the data for a specific purpose might use the data for a totally different purpose, and this essentially violates the user's interests.

- Some mining algorithms might use controversial attributes like sex, race, religion, or sexual orientation to categorize individuals. These practices might be against the anti-discrimination legislation.[18]

Table 3: Web Usage Mining

S.No.	Method/Tool/Technique	Application	Year	Source
1.	Web Mate	Provides effective browsing and searching help	1998	[73]
2.	SpeedTracer	Web server log	1998	[74]
3.	WebLogMiner	Web server log	1998	[7]
4.	Web usage miner (WUM)	Mining navigational patterns in the form of graphs	1998	[72]
5.	Web SIFT (Based on WEBMINER)	To mine interesting frequent item sets automatically from real web data	1999	[70]
6.	Mining Internet Data for Associative Sequences (M*i*DAS)	Pattern discovery	2000	[71]
7.	Statistical Analysis Association Rules	Personalization Site Modification	2000	[10]
8.	INSITE	To generate user profiles in real time	2000	[66]
9.	STRATDYN	Visualization of navigation patterns	2000	[67]
10.	KOINOTITES	Personalisation	2000	[65]
11.	WebTool (sequential pattern mining - PSP algorithm)	Usage profiling	2000	[69]
12.	i-JADE Web-Miner	E-commerce	2001	[63]
13.	Web Quilt	usability testing and collected data analysis	2001	[64]
14.	Clustering algorithm called Citation Cluster	Construct a conceptual hierarchy of the Web site	2002	[50]
15.	Sequential and non-Sequential patterns	Web Personalization	2002	[59]
16.	AWUSA	Automated website usability evaluation	2002	[62]
17.	i-Miner	Pattern Discovery and trend analysis	2003	[61]
18.	SEWeP	Web Personalization	2003	[68]
19.	Dynamic clustering based method	Representing a collection of user web Navigation sessions	2004	[51]
20.	Inductive Database Approach	Knowledge Discovery Ease	2004	[58]
21.	Improved WAP tree	Sequential pattern mining	2006	[52]
22.	Taxonomy of Browsing data	Decision Support	2007	[53]
23.	Web based recommender Systems	predict user's intention and their navigation behaviors	2008	[37]
24.	ArchCollect	monitor user's interactions in web media	2008	[59]
25.	Longest common Subsequences algorithm	Predict user near future movement.	2009	[54]
26.	WebPUM	Predict user near future Movement	2010	[55]
27.	Apriori-all algorithm	Exploring user Pattern habits	2010	[57]

6. CONCLUSION

In this paper, the literature survey in the area of web mining is being provided. The paper basically focuses on the methodologies, techniques and tools of the web mining. The basic emphasis is given on the three categories of the web mining and different techniques incorporated in web mining. The paper explains the web mining subtasks and web mining taxonomy as a base. Then after, three literature review tables are being provided on web content, web structure and web usage mining. The survey came up with the pros and cons of the web mining techniques. It clears the scope of the web mining and presents a better analysis and comparison of web mining and its types.

7. REFERENCES

[1] Vel L. , Royakkers L. , "Ethical Issues in Web Mining", *Ethics and Information Technology* **6:** 129–140, 2004.

[2] Dzitac I. and Moisil I. ,"Advanced AI Techniques for Web Mining", Proceeding of the 10th WSEAS international conference on Mathematical methods, computational techniques and intelligent systems,2008.

[3] Yadav S. , Ahmad K. and Shekar J.,"Analysis of web mining applications and beneficial areas",IIUM Engineering Journal,vol. 12,No. 2 , 2011 .

[4] Etzioni O., " The world wide web: Quagmire or gold mine", Communications of the ACM, Volume 39 Issue 11, Nov. 1996,Pages 65 - 68 .

[5] Kosala R. and Blockeel H.,"Web Mining Research: A Survey", *ACM SIGKDD Explorations Newsletter*, June 2000, Volume 2 Issue 1..

[6] Chang C., Lui S., Wu Y.,"Applying Pattern Mining to Web Information Extraction", Advances in Knowledge Discovery and Data,2001 – Springer.

[7] Li A., Zhang L." A Study of the Gap from Data Mining to its Application with Cases", International Conference on Business Intelligence and Financial Engineering", 2009.

[8] Yusifov F. F.,"Web Traffic Mining using Neural Networks", World Academy of Science, Engineering and Technology 21 2008

[9] Punin J, Krishnamoorthy M, Zaki M (2001)," Web usage mining: Languages and algorithms", Proceedings of Studies in classification, data analysis, and knowledge organization,Springer, Heidelberg.

[10] Srivastava J., Cooley R., Deshpande M.,Tan P.,"Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data", ACM SIGKDD Explorations Newsletter volume1,issue 2, 12-23,2000.

[11] Seydim Y. A., "intelligent agents: a data mining perspective" Techreport,1999.

[12] Cooley R., Mobasher B., Srivastava J. ,"Web Mining: Information and Pattern Discovery on the World Wide Web", Proceedings, Ninth IEEE International Conference on Tools with Artificial Intelligence, 1997.

[13] Pater M., Popescu E. D., Mastei D.," Pattern discovery techniques in Web mining", Journal of Computer Science and Control Systems,volume 1,issue 1,2008.

[14] Eirinaki M., Vazirgiannis M., "Web Mining for Web Personalization",Journal ACM Transactions on Internet Technology (TOIT), Volume 3 Issue 1, February 2003.

[15] Fu Y., Shih M.,Creado M.,Ju1 C.," Reorganizing Web Sites Based on User Access Patterns",Proceeding of the tenth international conference on Information and knowledge managemen, Pages 583 – 585,2001.

[16] Zheng T.,Niu Y., Goebel R.," WebFrame: In Pursuit of Computationally and Cognitively Efficient Web Mining", Proceedings of the 6th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining,2002.

[17] Srivastava J., Desikan P., Kumar V.,"Web Mining - Concepts,Applications & Research Directions", Foundations and Advances in Data Mining,Studies in Fuzziness and Soft Computing Volume 180, 2005, pp 275-307.

[18] Alhawamdeh A. A. M.," Web Mining:Strategic Web Site Design for Small Business",Proceedings of the World Congress on Engineering, WCE 2007, London, UK, 2-4 July, 2007.

[19] Jain D.,Sinhal A.,Gupta N.,Narwariya P., Sarawat D. and Pandey A.,"Hiding Sensitive Association Rules without Altering the Support of Sensitive Item(s)", International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.2, March 2012.

[20] Júnior C.G.M.,Gong Z.," Web Structure Mining: An Introduction",IEEE International Conference on Information Acquisition, 2005.

[21] Maes P.," Agents that reduce work and information overload",Communications of ACM,Volume 37 Issue 7, July 1994,Pages 30 – 40.

[22] Fayyad U.," From data mining to knowledge discovery: An overview",Advances in Knowledge Discovery and Data Mining, pp.1-34, AAAI Press, 1996.

[23] Gomes, M. and Gong, Z., 2005, "Web Structure Mining: An Introduction", Proceedings of the 2005 IEEE International Conference on Information Acquisition

[24] Brown, C.M., Danzig, P.B. Hardy, D., Manber, U., and Schwartz, M.F.,"The harvest information discovery and access system", In *Proceedings of the 2d International World Wide Web Conference*, 1994, pp. 763–771.

[25] Bowman, C. M., Danzig, P. B., Hardy, D. R., Manber, U., Schwartz, M. F., Wessels, D. P., "Harvest: A Scalable, Customizable Discovery and Access System", Technical Report, University of Colorado,Boulder, Colorado, USA, 1995.

[26] Chawathe S., Garcia-Molina H., Hammer J.,Ireland K.,Papakonstantinou Y., Ullman J., Widom J.," The TSIMMIS project: Integration of heterogenous information sources", Proceedings of IPSJ, Tokyo, Japan, October 1994.

[27] Hammond K., Burke R., Martin C., and Lytinen S.," FAQ finder: A case-based approach to knowledge navigation",In Working Notes of the AAAI Spring Symposium: Information gathering from Heterogeneous, Distributed Enviornments, 1995, AAAI Press, Stanford University, pp. 69–73.

[28] Perkowitz, M. and Etzioni, O.," Category translation: Learning to understand information on the internet", *Proceedings of the Fifteenth International Joint Conference on AI*, (Montreal, Can.), Aug. 1995, pp. 930– 936

[29] Doorenbos R.B., Etzioni O. and Weld D.S.," A scalable comparison- shopping agent for the world-wide web",Technical Report 96–01–03, University of Washington, Dept. of Computer Science and Engineering, January 1996.

[30] Abiteboul S., Quass D., McHugh J., Widom J. and Wiener J.," The Lorel query language for semistructured data",Journal of digital Libraries, November 1996.

[31] Baldi M.,Damiani E., and Insaccanebbia F.," Structuring and querying the Web through graph-oriented languages" , *Proceedings* of SEBD 97, SEBD Conferences, Verona, Italy, June 1997.

[32] Atzeni P., Masci A.,Mecca G.,Merialdo P., and Tabet E.," ULIXES: Building relational views over the Web",Proceedings of the 13th International Conference on Data Engineering (ICDE'97), April 1997. IEEE.

[33] Long F.,Zhang H. and Feng. D.D.,"Fundamentals of content based image retrieval",www.cse.iitd.ernet.in/~pkalra/siv864/Projects/c h01_Long_v40-proof.pdf

[34] Zhang H.,Chen Z.,Li M. and Su Z., "Relevance feedback and learning in content-based image search", World Wide Web 6(2) (2003) 131–155.

[35] Chen L.,Lian W. and Chue W., "Using web structure and summarization techniques for web content mining",International Journal on Information Process Management 41(5) (2005) 1225– 1242.

[36] Campos R.,Dias G.,Nunes C., "WISE: Hierarchical Soft Clustering of Web Page Search Results Based on Web Content Mining Techniques," wi, pp.301- 304, 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI'06), 2006.

[37] Hosseini M.,Hassani A. H. ,"Mining Search Engine Query Log for Evaluating Content and Structure of a Web Site" in Proceedings of the 2007 IEEE/WIC/ACM International Conference on Web Intelligence.

[38] Rahmani R.,Goldman S.A., Zhang H.,Cholleti S.R and Fritts J.E. , "Localized Content-Based Image Retrieval", IEEE Transactions on Pattern Analysis and Machine Intelligence , vol.30, no.11, pp.1902-1912,Nov.2008

[39] Poonkuzhali G.,Thiagarajan K. "Signed Approach for Mining Web Content Outliers", World Academy of Science, Engineering and Technology 56,2009

[40] Madria K.S. , Bhowmick S.S.,Ng K.W. and Lim E.," Research Issues in Web Data Mining", Proceedings of the First International Conference on Data Warehousing and Knowledge Discovery, p.303-312, September 01, 1999

[41] Kleinberg K.J.," Authoritative sources in a hyperlinked environment", Journal of the ACM, 46(5):604–632, 1999.

[42] Furnkranz J.," Web structure mining — Exploiting the graph structure of the worldwide web", OGAI-J. 21(2) (2002) 17–26

[43] Smith A.K. and Ng A.," Web page clustering using a self-organizing map of user navigation patterns", Decision Support Syst. 35(2) (2003) 245–256.

[44] Fang X. and Sheng O., "LinkSelector: A web mining approach to hyperlink selection for web portals", ACM Trans. Internet Tech. 4(2) (2004) 209–237

[45] Hay B., Wets G. and Vanhoof K., "Mining navigation patterns using a sequence alignment method", Knowledge Inform. Syst. 6(2) (2004) 150–163

[46] Guan S. and McMullen P., "Organizing information on the next generation web —design and implementation of a new bookmark structure", International Journal Inform Technol. Decision Making 4(1) (2005) 97–115

[47] Song Q. and Shepperd M.," Mining web browsing patterns for e-commerce", Computation Industry 57(7) (2006)622–630

[48] Chikhi F. N.,Rothenburger B. and Aussenac-Gilles N. "A Comparison of Dimensionality Reduction Techniques for Web Structure Mining", Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence,P.116-119 ,2007

[49] Moussiades L. and Vakali A., "Mining the Community Structure of a Web Site", bci, pp.239-244, 2009 Fourth Balkan Conference in Informatics, 2009

[50] Zhu J. and Hong J.," Using Markov Models for Web Site Link Prediction" ,College Park, Maryland, USA ACM June 11-15, 2002

[51] Borges and Levene M.,"A dynamic clustering-based markov model for web usage Mining", cs.IR/0406032,2004

[52] Xiaoqiu T. and Min Y.," Mining Maximal Frequent Access Sequences Based on Improved WAPtree" ,Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications(ISDA'06)

[53] Tao H.Y.,Hong P.T. and Su M.Y.," Web usage mining with intentional browsing data" , international journal of Expert Systems with Applications 34 (2007) 1893–1904

[54] Jalali M. and Mustapha N.," A Recommender System Approach for Classifying User Navigation Patterns Using Longest Common Subsequence Algorithm", American Journal of Scientific Research ISSN 1450-223X Issue 4 (2009), pp 17-27

[55] Jalali M. and Mustapha N. ," WebPUM: A Web-based recommendation system to predict user future movements", international journal Expert Systems with Applications 37 (2010) 6201–6212

[56] Yong-gui W. and Zohn J.,"Research on Semantic Web Mining", 2010 International Conference On Computer Design And Appliations (ICCDA 2010).

[57] Gaol L. F.," Exploring The Pattern of Habits of Users Using Web Log Squential Pattern", Second International Conference on Advances in Computing, Control, and Telecommunication Technologies,2010.

[58] Meo R.,Lanzi L.P.,Matera M. And Esposito R.," Integrating Web Conceptual Modeling and Web Usage Mining", Proceedings of the sixth WEBKDD **w**orkshop: Webmining and Web Usage Analysis (WEBKDD'04), in conjunction with the 10th ACM SIGKDD conference (KDD'04), Seattle, Washington, USA, 2004.

[59] Mobasher B., Dai H.,Luo T. And Nakagawa M., "Using Sequential and Non-Sequential Patterns in PredictiveWeb Usage Mining Tasks", Proceeding ICDM '02 Proceedings of the 2002 IEEE International Conference on Data Mining, IEEE Computer Society Washington, DC, USA ©2002.

[60] Esmin A., Lima J., Yano,Tiago E. T. And Carneiro G. S. ,"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. 375-380,2008.

[61] Abraham A. ,"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 .

[62] Tiedtke T. ,Märtin C. and Gerth N.,"AWUSA – A Tool for Automated Website Usability Analysis", PreProceedings of the 9th International Workshop on the Design, Specification and Verification of Interactive Systems,2009.

[63] Hong I.J., Heer J., Waterson S. and Landay A.J., "WebQuilt: A proxy-based approach to remote web usability testing", ACM Transactions on Information Systems, 19(3), 2001, 263-285.

[64] Zaiane R.O.," 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.

[65] Pierrakos D. ,Paliouras G., Papatheodorou C. and Spyropoulos C. D. ,"KOINOTITES: A Web Usage Mining Tool for Personalization", Proceedings of Panhellenic Conference on Human Computer Interaction, Greece, Patras, pp. 231-236,2000.

[66] Shahabi C., Faisal A., Kashani F. B. and Faruque J. ,"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. 635-638,2000.

[67] Berendt B.,"Web usage mining, site semantics, and the support of navigation", KDD Workshop on Web Mining for ECommerce Challenges and Opportunities" pp. 83–93,2000.

[68] Eirinaki M., Vazirgiannis M. and Varlamis I.,"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. 99-108,2003.

[69] Masseglia F., Poncelet P. and Cicchetti R. ,"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. 892-901,1999.

[70] Cooley R., Tan P. N. and Srivastava J. ,"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.

[71] Buchner G.A.,Baumgarten M., Anand S. S.,Mulvenna D. M., and Hughes H. J.," Navigation Pattern Discovery from Internet Data", Proceedings of Web Usage Analysis and User Profiling at the International WEBKDD99 Workshop, 2000, 74- 91.

[72] Spiliopoulou M. and Faulstich L. C. ,"WUM : A Web Utilization Miner", EDBT Workshop on Web Databases, pp.1-7, Valencia, Spain,1998.

[73] Chen L., and Sycara K., "WebMate: A Personal Agent for Browsing and Searching", Proceedings of the 2nd International Conference on Autonomous Agents, Minneapolis MN, USA, 1999, 132-139.

[74] Wu K. L., Yu P. S. and Ballman A.,"SpeedTracer: A Web usage mining and analysis tool", IBM Systems Journal on Internet Computing, Vol. 37, pp. 89-105,1998.