International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 12 (2019) pp. 2830-2836

Extracting Multimedia Information and Knowledge Discovery using Web Mining: Challenges and Research Directions

M. Ravi[1], Dr. M. Ekambaram Naidu[2] and Dr. G. Narsimha[3]

[1]*Research Scholar, CMR Institute of Technology, Hyderabad-501401, India.*
[2]*Principal, SRK Institute of Technology, Vijayawada-521108, India.*
[3]*Professor, Jawaharlal Nehru Tech University, Hyderabad-500085, India.*

ABSTRACT

World Wide Web is an arrangement of interlinked hypertext records that are pervaded through the Internet. Internet browsers facilitate simple access to various sources of text content and multimedia effectively. Today billions of pages are indexed via various search engines as well as aquisite the desired data isn't so straightforward. This has provoked the requirement for creating programmed mining procedures on the WWW, by the new term known as - "Web mining". In the web mining scenario, the information material can be accumulated at the server side, proxy servers, customer side or can be obtained through the association's database. This obtained data is generally unstructured and massive in nature. Presently, the greater part of the mining procedures utilized are content driven and the calculations are arranged towards content mining system. Generally, the web oriented multimedia data is in the form of images, text, animations, audios and videos etc. Extraction and retrieval of significantly useful information and specific knowledge exploration from these multimedia forms is a great challenge. In this paper, we extensively surveyed significant developments occured over past years. We have listed some significant tools for performing the multimedia data mining and analysis, the data which is obtained from various web sources. We also given an argument on the various open challenges involved in this domain along with the problem identification and our possible research directions to overcome these challenges.

Index terms: Web data mining, Knowledge processing, Information Retrieval, Multimedia databases.

1. INTRODUCTION

The web is colossal, divergent therefore raises the adaptability, mixed media information and fleeting issues separately. The need to see substantial, complex, data rich informational collections [2][3] is normal to for all intents and purposes in all fields of business, science, and building. *The whole procedure of bestowing a computer-based approach for finding and extracting useful knowledge matter from diverse and hetrogeneous kind of web records is known as web mining.* Web oriented media mining [25-27] is characteristically at the intersection of research from a few subject varieties like computer vision, interactive media recovery, knowledge mining, [4] AI, statistical data mining, database and computerized reasoning. Modern improvements in advanced media innovations has made transmitting and putting away a lot of multi/rich media information (for example textual content corpus, pictures, music, video and animations etc.) progressively much affordable and equitable than anytime before.

Though, the cutting edge strategies to mining and deal with those multimedia are in their earliest stages. Advances and improvements in interactive media acquisition and capacity innovation the quick advancement has prompted the quickly developing inconceivable measure of information being kept in the databases. Valuable information to clients can be revealed if these media records are granularily analysed. Mixed media [6][9] mining manages the extraction of implicit knowledge, interactive media information correlations, or different example patterns not peculiarly stored in the diverse multimedia files. Likewise, during retrieval, ordering and grouping of web multimedia information with proficient data combination of the distinctive modalities is fundamental for the framework's general execution with improved performance.

1.1. Categorization of Web Mining

Web mining is generally typecasted into three ways -

1.1.1. Content Oriented Web Mining

Web can contain any type of information like - text, images, audio, animation, video or in any combinational form. The extraction of useful and knowledge oriented information from any of these web sources can be thought of as content oriented web mining.

1.1.2. Strcture Oriented Web Mining

It is the strategy for finding and examining structural and auxiliary data from the web. In view of the structure data, it tends to be subcategorized as hyperlinks and archive structure. Hyperlinks are utilized to interface a Web page with different Webpages or different parts of a similar Web page. A hyperlink that interfaces with an equivalent piece of a similar page can thought of as an intra-archive hyperlink that associates two specific pages is called a between report hyperlink. Report structure alludes to the tree-organized course of action, dependent on HTML and XML labels inside the page. This sort of mining is centered around automatically extricating Document Object Model architecture out of archives.

1.1.3. Usage Oriented Web Mining

The web contains a colossal gathering of diverse example patterns. The appliance of data mining strategies so as to get valuable pattern instances from the Web is called as Web Usage Mining. Utilization information gathers the origin of web clients with their perusing conduct at a website. It is additionally partitioned in view of the usage data considered. In case of Web Server information, the particular client logs details are compiled by the associated web server and commonly assimilate IP address, page source along with the access time. Commercial application servers utilized for internet commerce oriented applications can be tracked to get different sorts of commercial events and log them in the operational server logs. This comes under the categorization of Application Server Data. To create chronicles upon the previously mentioned logging and applications require a combination of these systems, named as Application Level Data.

1.2. Motivation

The web is a colossal accumulation of uncontrolled heterogeneous archives which makes the web a fruitful zone of information mining research with the tremendous amount of data which is attainable on the web. Web mining is characterized as the revelation and investigation of helpful data from WWW. In the web mining, information can be gathered at server side, customer side, intermediary servers or acquired from the association's database. On the basis of the source and nature of data, the mining type may differ. The issue of creating automatized devices so as to discover, concentrate, channel and assess the clients wanted data from unlabeled, inconsistent and heterogeneous web information required human intervention which can generally be fused utilizing soft computing techniques.

1.3. Organization of the Paper

Rest of the paper is structured as - Section 2 through some light on the multimedia directories terminology and its types. Section 3 discusses about the general process of multimedia information retrieval and knowledge discovery. Literarure review in this domain along with various web multimedia mining tools information is presented in section 4. Section 5 targets on the various open Issues, challenges and our research directions. Finally, section 6 presents the conclusive summary of the work.

2. MULTIMEDIA DATABASES

The multimedia databases are generally classified as -

- **Text -** The content can be stored in variety of structures. In expansion to ASCII based documents, content is commonly put away in processor documents, spreadsheets, databases and mixed media objects etc. With the accessibility and bounty of GUIs that permit exceptional impacts, for example, shading and styles, the assignment of storing text become increasingly complex procedure.
- **Images -** There is an incredible variety in the quality as well as the size of capacity for static images. Digitalized pictures are grouping of pixels that speaks to a locale in the client's graphical presentation. The variables influencing the space extension overhead for static images are resolution, size, and complexity involved to store picture. The image formats usually are in extension form of .jpg, .png, .bmp and .gif.
- **Audio -** Audio is an inexorably famous datatype being coordinated into the greater part of the applications. It's very much space intensive. One moment of sound can cover up to 2 to 3 MBs of storage space. A few strategies are utilized to shrink it in a convenient format.
- **Video -** Digitalized video is considered as one of the massive space consuming data storage format. The digitalized recordings are reserved as an arrangement of image frames. Contingent on its size an splitted single frame may occupy upto 1 MB. Furthermore, to have practical video playback, compression and de-compression of this digitalized form require the ceaseless fetching rate.
- **Graphics/Animations -** Realistic Objects comprises of uncommon information structures utilized to characterize 2D and 3D contours through which interactive media objects can be characterized. These incorporate different formats utilized by picture, video editing functionalities, for example, Computer-aided design/Computer-aided manufacturing objects. Various web multimedia content types are shown as Fig 1.

Fig.1 - Web Multimedia Content Types

2.1. Types of Multimedia Databases Mining

Mining Multimedia information is utilized to recover unique kinds of data. The way towards applying multimedia mining comprises of various procedural steps. Data accumulation is the primary purpose of any learning framework, as the nature of crude information is thought of as factor which decides the generally attainable performance. The fundamental objective of pre-processing is to find the significant examples from raw formatted data, which incorporates the ideas of data sanitation,

standardization, transformation, variable extraction and feature selection etc. Different types and components are given in Fig 2.

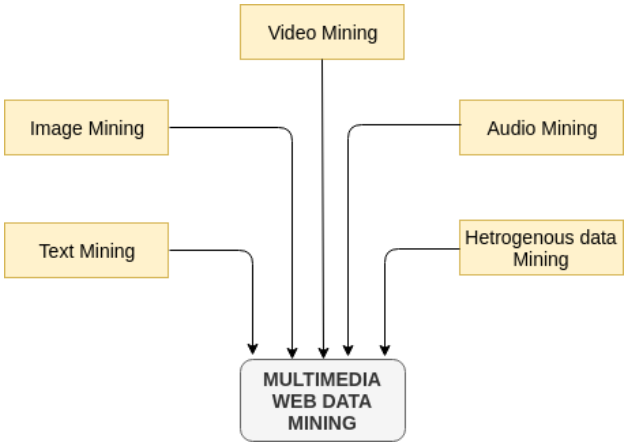

Fig.2 - Mining Types of Multimedia Data

3. GENERAL PROCESS OF MULTIMEDIA INFORMATION RETRIEVAL AND KNOWLEDGE DISCOVERY

The general process of multimedia retrieval and knowledge discovery is shown as Fig 3. In this process, first the data which is obtained from various homogeneous or hetrogeneous sources, is processed through data cleaning and integration module.

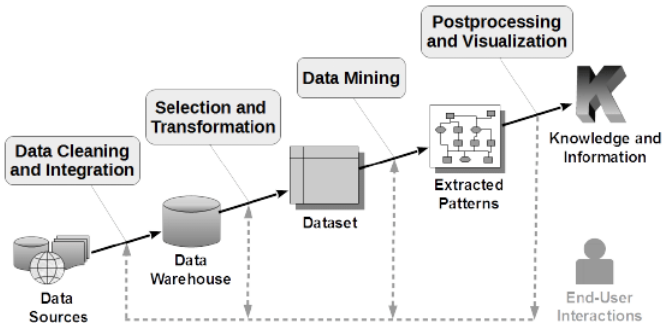

Fig.3 - General Process and Steps Module in Multimedia Information Retrieval

It further goes to data warehouse. There, various transformation functions are performed. On the obtained pre-processed dataset, the data mining sub-modules e.g. features dimension measurement, feature extraction, feature selection are performed. After extracting the patterns, the postprocessing can be performed on the data through the methods like - summarization and visualization. Finally, the predictive reasoning can be operated on the unseen data after model is trained on the training dataset.

3.1. Various Applications

Web applications being grown quickly in the industry depend on the utilization of web mining ideas, despite the fact that the associations that built up these applications would not consider it as such. Some portions of the remarkable applications are -

3.1.1. Web Search - Google:

Google is a standout amongst the most well known and generally utilized web indexes. It gives clients access to data from more than 2 billion site pages that it has listed on its own server. The quality and speed of the inquiry makes it the best internet search utility. Prior web indexes focused on web content alone to restore the important pages to a query. Google was the first to present the significance of the connection structure in mining data from the web. *PageRank*, which estimates the significance of a page, is the basic innovation in all Google inspection items, and utilizes auxiliary data of the web architecture to return significantly fantastic outcomes.
The Google toolbar is another administration facilitated by Google that tries to make seek simpler and useful by giving extra feature highlights, for example - featuring the interrogatory words on the returned site pages. Google's web index facilitates a quick and simple approach to seek inside a specific theme or related subjects. The publicizing program presented by Google targets clients by giving ads that are important to a hunt inquiry. One of the most recent and significant administrations offered by Google is Google News. It incorporates news from the online variants of all news sources and sorts out them completely to make it simpler for clients to go through "the most applicable news." It tries to deliver most recent data by continually recovering pages from news site worldwide which are being refreshed mostly all the time.

3.1.2. E - Government:

Electronic
government is the utilization of electronic specialized gadgets, for example, PCs and the Internet to facilitate open administrations to natives and different people in a nation or district. As indicated by Jeong, 2007 the term comprises of the computerized communications between a native and their administration (C2G), among governments and other government organizations (G2G), among government and residents (G2C), among government and workers (G2E), and among government and organizations/commerces (G2B). In various processes in it, web multimedia data mining plays a significant role. Some of the other most successful applications in this domain are as -

- Digital libraries
- Electronic commerce
- Security and Crime Investigation
- Business-to-consumer (B2C) e-Commerce

4. RELATED WORK REVIEW

This section extensively represents the research work and developments in this domain that took place over past years. Yoshitaka et al. [1] given an extensive survey on content based information retrieval from various multimedia web resources. B. Thuraisingham et al. [2], in their paper, discusses about the significant tools available for multimedia databases mining. Pravin M. Kamde et al. [3], Manda Jaya Sindhu et al. [4]

given a survey on web multimedia mining. Reshma P.K. et al. [5] given a soft computing framework for web mining of multimedia data.

R.Cooley et al. [6] given the state of the art review and method of information and pattern discovery on the WWW. M. Stamenovic et al. [7] given a visual classifier, useful for surmising a document's overall appearance, and a text classifier, for making content-informed decisions. Osmar R. et al. [8] a system prototype - *MultiMediaMiner*, for the multimedia data mining. Dianhui Wang et al. [9] discussed about learning based neural similarity metrics for multimedia data mining soft computing. L. A. Zadeh [10] given Fuzzy logic theory and discussed about its intersection with neural networks and soft computing. R. Yager [11] given a framework for linguistic as well as hierarchical queries for the document retrieval process. O. Etzioni et al. [12] presented a feasibility demonstration for web document clustering process.

A. Joshi et al. [13], in their paper, talks about various robust fuzzy clustering methods to support web mining. J. Shavlik et al. [14] proposed a system prototype for building the intelligent agents that learn retrieval and extraction of information. M. Boughanem et al. [15] given the genetic based approaches for the purpose of information retrieval. V. U. Maheswari et al. [16] proposed the rough set based approach for multimedia web usage mining. Sankar.K.Pal et al. [17], in their paper, given the state of the art, various challenges involved and future directions in web mining. D. Sridevi et al. [18] Kamika C. et al. [19] discuss about the various tools and techniques available for web multimedia data mining. V. Bharanipriya et al. [20] Preeti C. et al. [21] Zhang, Q. et al. [22] Darshna N. et al. [23], in their papers, given the web content mining tools with improved structural efficiency. Chidansh A. [24] given the state of the art and challenges in the domain of multimedia data mining.

Jaideep S. et al. [28][29], Kavita et al. [30] given the applications of usage patterns from web data and further research directions. Rim Rekik et al. [31] proposed a process of collecting and extracting data (criteria featuring web sites) from a list of studies. Text mining is applied for this SLR to construct a dataset. Image object's pattern analysis is performed by Imran K. et al. [32]. Y. Li [33] analyzed the realization of web content mining, their basic algorithm principles and their application areas.

4.1. Existing Works: Tabular Representation

Some significant existing works with their advantages, limitations and description of tools used information is compiled in Table 1. Authors used various tools in their works like - Autobox, NN, Self organizing map, OLAP, SVM, SAS, DataSift, Advanced NeuralNets, Soft Computing, Hierarchial Fuzzy model, Fuzzy Clustering, multi-layer NN, GA, Roughset, WEKA, i-Miner etc.

Table 1 - Significant Existing Works Summarization

Authors	Title	Advantages	Limitations	Tool used	Ref.
Manda Jaya Sindhu et al.	Multimedia Retrieval Using Web Mining	Efficiently functions on images and video web data	Suffers from overfitting problem	Autobox, NN	[4]
Reshma P.K. et al.	Web Mining for Multimedia Data-A Soft Computing Framework	Works better for hetrogenous data	text-centric approach	Self organizing map	[5]
R.Cooley et al.	Web Mining: Information and Pattern Discovery on the World Wide Web	Information Filtering is comparatively better	Computational complexity more	OLAP tools	[6]
M. Stamenovic et al.	Machine Identification of High Impact Research through Text and Image Analysis	Achieves attribute selection based dimension reduction	Attribute extraction takes more time	SVM	[7]
Osmar R. et al.	A System Prototype for MultiMedia Data Mining	Comparatively better web multimedia data categorization accuracy	Overfitting problem for some datasets	SAS, DataSift	[8]
Dianhui Wang et al.	Learning Based Neural Similarity Metrics for Multimedia Data Mining Soft Computing	Efficient backpropagation strategy	High computation complexity for slow learning rate	Advanced NeuralNets, Soft Computing	[9]
R. Yager et al.	A framework for linguistic and hierarchical queries for Document retrieval	Hierarchial and distributed framework	High complexity while error correction and weights updation	Hierarchial Fuzzy model	[11]
A. Joshi et al.	Robust fuzzy clustering methods to support web mining	Web data categorization accuracy improved	Sometimes suffers from underfitting	Fuzzy Clustering	[13]
M. Boughanem et al.	Connectionist and genetic approaches for information retrieval	Random mutation guarantees to some extent to get wide range of solutions	Hard to choose parameters like number of generations, population size etc	multi-layer NN, GA	[15]
V. U. Maheswari et al.	The variable precision roughset model for web usage mining	Generate in automatic way the sets of decision rules from data	Requirement of discrete format data	Roughset	[16]
Sankar.K.Pal et al.	Web mining in Soft Computing Framework	easy to solve nonlinear problem in data	Excessively large network are not able to respond correctly to new patterns	Soft Computing	[17]
J. Srivastava et al.	Web Usage Mining: Discovery and Applications of Usage Patterns from Web data	Optimal training strategy available, Performs granularization	Normally needs data, much more than a human normally does. This is because machine learning normally optimize over an artificial hypothesis space most of which we would consider ridiculous or highly unlikely to begin with	Fuzzy logic	[28]
Imran K. et al.	Object analysis in image web mining	The disappearance of a few pieces of information in one place does not prevent the network from functioning	Unexplained behavior of the learning network	WEKA, i-Miner	[32]

4.2. Web Multimedia Mining Tools

Different types of tools used in all categories of web multimedia mining. Some significant tools and their features are listed in Table 2.

Table 2 - Web Multimedia Mining Tools		
Sr. No.	**Tool**	**Features**
1	DataPreparator	Performs cleaning, extraction and transformation of data.
2	Lisp Miner	Achieves data pre-processing by scrutinizing the stream and the corpus collected.
3	Web Info Extractor	Its helpful in mining extract structure or unstructured data from web page, extracting web content.
4	RStudio	RStudio is a free and open-source integrated development environment for R, a programming language for statistical computing, multimedia data analysis and graphics.
5	Autobox	It is a tool for statistical analysis for web multimedia data.
6	GRETL	Its a cross-platform system software for econometric reasoning, written in the C programming script.
7	RATS	Regression Analysis of Time Series, is a analytical package tool for the time series analysis and geometric.
8	ORANGE	Performs data analysis and knowledge exploration with an attractive data visualization. It performs statistical assessments, box plots and scatter plots.
9	TANAGRA	It supports several standard web multimedia data mining tasks such as: Visualization, Descriptive statistics, Instance selection, feature selection, feature construction, regression, factor analysis, clustering, classification and association rule learning etc.
10	Sewebar-Cms	This tool helps while opting rules among diversified rules in association rule mining.
11	i-Miner	Identify data cluster by employing fuzzy clustering procedure and fuzzy inference system for the pattern exploration.
12	WebViz	Analyze the patterns and caters them in the style of graphical patterns.

5. OPEN ISSUES, CHALLENGES AND RESEARCH DIRECTIONS

This section discusses the existing open issues, various challenges in this research domain and finally our possible research directions to tackle them efficiently. The term - "Multimedia data mining" can be defined as - the procedure of exploring interesting structural patterns from the several media information such as audio, video, text and image which are not generally attainable by the primitive queries and its correlated results. The significant open issues and challenges in this domain are as follows:-

- The advancements in the multimedia data procuration as well as storage technology have drove to the tremendous growth of massive multimedia databases. Scrutinizing this huge amount of multimedia data to identify useful knowledge is a challenging issue. This challenge has opened significant opportunity for carrying out research in the domain of Multimedia Web mining.
- The speed at which multimedia is generated is also an open challenge, as due to high speed streaming, the data might be updated in real scenarios before the previous data is under processing and under analysis itself. This situation creates a problem.
- The web multimedia data might be available in hetrogenous and unstructured format. Processing this kind of data is also a substantial challenge in this domain.

5.1. Future Research Directions

Most of the web sources are in the form of text or images. Knowledge extraction from such homogenous or hetrogenous web multimedia databases implies the method of discovering helpful information from knowledge, and data processing concerns to a selected step during this means. Data processing is the employment of explicit algorithms for getting patterns from information. Unplanned utilization of information-mining tech- niques (rightly criticized as data dredging within the applied math literature) are often a risky activity, simply heading to the invention of trivial and invalid patterns. This domain has been already evolved upto some extent, and continuing to mature, from diverse research domains such as - pattern recognition, machine learning, directories, statistics, knowledge exploration and management for multimedia expert systems, high-performance computing, data summarization and visualization.
Machine learning can be considered as a significant branch of statistical computing which can be utilized in the process of various multimedia data analysis and generally involves two stages i.e. training and testing. Training aims to be learn from illustrious properties by exploitation learning algorithms and testing refers to form predictions on the unidentified properties by exploitation the information learned within the training phase. Machine learning is generally classified into two types - supervised learning and unsupervised learning, in congruence to the shape of learning. Supervised learning means that learning with an instructor as a result of all occurrences from a training set are labeled. The intention of this kind of learning scenario is to make a model by learning from labeled knowledge and so to form predictions on alternative unlabeled occurrences with reference to the worth of a predicted variable (attribute).

1. Today, to process the huge sized, inconsistent, incomplete and vague multimedia data generated by web sources and computing machines is a challenging task. In today's real

world problems, variable selection is an essential aspect due to presence of irrelevant variables in the data. As our future research work, we will propose soft computing based efficient approaches to deal with this problem.

$\mathcal{S}$oft $\mathcal{C}$omputing - *Soft coputing is a collection of tools including fuzzy sets, layered neural networks, genetic algorithms and rough set theory. These are used for modeling complex functions, dealing with uncertainty, provide learning and generalization capabilities.*
It will provide advantages in our work which is compiled as below:-

(a) Soft computing with provide flexible information processing capability for handling real life ambiguous situations related to multimedia web mining. We will exploit it to the tolerance for imprecision, uncertainty, approximate reasoning and partial truth to achieve tractability, robustness, low-cost solutions and close resemblance to obtain comparatively better decision making. Rough Sets will use in granular computation and knowledge discovery from various web multimedia databases.

(b) In proposed work we will use the tools - mainly *RStudio* and *R language environment*, its various associated libraries for experimental simulation purposes.

2. We also come up with new statistical models which can tackle the problem of unstructured and hetrogenous natured multimedia data in various forms like - images, text etc.

3. By utilizing our proposed approaches, we will perform experiments on standard datasets related to multimedia web sources such as images, text corpus. To prove the novelty of proposed procedures, we compare with the significant existing approaches from literature.

6. CONCLUSION

Web mining is developing at a great pace since its improvement and new strategies are being created both utilizing traditional and delicate processing approaches simultaneously. In the present situation, the web is picking up multimedia character with pages containing images, videos, text, and so forth. The aim of carrying out Multimedia data mining is to utilize the identified patterns to enhance decision making. This paper surveys significant state of the art developments. We also given a discussion on the various open challenges involved in this domain along with the problem identification and our possible research directions to overcome these challenges.

REFERENCES

[1] Yoshitaka A. and Ichikawa T., A Survey on Content-based Retrieval for Multimedia Databases. IEEE Transaction on Knowledge & Data Engineering, Vol 11, 1999, pp. 81-93.

[2] Bhavani Thuraisingham, Managing and Mining Multimedia Databases, CRC Press, 352 Pages, ISBN 9780849300370, June 28, 2001.

[3] Pravin M. Kamde, Dr. Siddu. P. Algur, A survey on web multimedia mining, The International Journal of Multimedia & Its Applications (IJMA) Vol.3, No.3, August 2011.

[4] Manda Jaya Sindhu, Y. Madhavi Latha, V. Samson Deva Kumar, Suresh Angadi, Multimedia Retrieval Using Web Mining, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-2, Issue-1, March 2013.

[5] Reshma P.K., Lajish V.L., Web Mining for Multimedia Data-A Soft Computing Framework, International Journal of Scientific & Engineering Research, Volume 5, Issue 9, September-2014.

[6] R.Cooley, B. Mobasher, J.Srivastava, Web Mining: Information and Pattern Discovery on the World Wide Web. IEEE Transactions, 1997.

[7] Marko Stamenovic, Sam Schick, Jiebo Luo, Machine Identification of High Impact Research through Text and Image Analysis, 2017 IEEE Third International Conference on Multimedia Big Data, 19-21 April 2017, IEEE, DOI: 10.1109/BigMM.2017.63.

[8] Osmar R., Zaane Jiawei, Han Ze-Nian, Li Sonny H., Chee Jenny, Y. Chiang, MultiMediaMiner: A System Prototype for MultiMedia Data Mining, Intelligent Database Systems Research Laboratory and Vision and Media Laboratory report, 2009.

[9] Dianhui Wang, Yong-Soo Kim, Seok Cheon Park, Chul Soo Lee and Yoon Kyung Han, Learning Based Neural Similarity Metrics for Multimedia Data Mining Soft Computing, Volume 11, Number 4, pp. 335- 340, 2007.

[10] L. A. Zadeh, Fuzzy logic, neural networks, and soft computing, Commun. AGM, vol. 37, pp. 77-84, 1994.

[11] R. Yager, A framework for linguistic and hierarchical queries for Document retrieval, in Soft Computing in Information Retrieval: Tech niques and Applications, F. Crestani and G. Pasi, Eds, Heidel berg: PhysicaVerlag,, vol. 50, pp. 3-20,2000.

[12] O. Etzioni and O. Zamir, Web document clustering: A feasibility demonstration, in Proc. 21st Annu. Int. ACM SIGIR Conf., pp. 46-54, 1998.

[13] A. Joshi and R. Krishnapuram, Robust fuzzy clustering methods to support web mining, in Proc. Workshop in Data Mining and Knowledge Discovery, SIGMOD, pp. 15-1-15-8, 1998.

[14] J. Shavlik and T. Eliassi, A system for building intelligent agents that learn to retrieve and extract information, Int. J. User Modeling User Adapted Interaction (Special Issue on User Modeling and Intelligent Agents), Apr. 2001.

[15] M. Boughanem, C. Chrisment, J. Mothe, C. S. Dupuy, and L. Tamine, Connectionist and genetic approaches for information retrieval, in Soft Computing in Information Retrieval: Techniques and Applications, F. Crestani and G. Pasi, Eds. Heidelberg, Germany: Physica-Verlag, vol. 50, pp. 102-121,2000.

[16] V. U. Maheswari, A. Siromoney, and K. M. Mehata, The variable precision rough set model for web usage mining, presented at the Proc. 1st Asia-Pacific Conf. Web Intell. (WI-2001, Maebashi, Japan, Oct. 2001.

[17] Sankar.K.Pal, Varun Talwar and Pabitra Mitra, Web mining in Soft Computing Framework: Relevance, State of the Art and Future Directions, IEEE Transactions on Neural Networks, September 2002.

[18] D. Sridevi, Dr. A. Pandurangan, Dr. S. Gunasekaran, Survey on Latest Trends in Web Mining, International Journal of Research in Advent Technology, Vol. 2, No.3, March 2014.

[19] Kamika Chaudhary, Santosh Kumar Gupta, Web Usage Mining Tools & Techniques: A Survey in International Journal of Scientific & Engineering Research, Volume 4, Issue 6,June-2013 1762 ISSN 2229-5518.

[20] V. Bharanipriya & V. Kamakshi Prasad, Web Content Mining tools: A Comparative Study in International Journal of Information Technology and Knowledge Management January-June 2011, Volume 4, No. 1, pp. 211-215.

[21] Preeti Chopra, Md. Ataullah, a Survey on Improving the Efficiency of Different Web Structure Mining Algorithms in International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-2, Issue-3, February 2013.

[22] Zhang, Q., Segall, R.S., Web Mining: A Survey of Current Research, Techniques, and Software, International Journal of Information Technology & Decision Making. Vol.7, No. 4, pp. 683-720. World Scientific Publishing Company (2008).

[23] Darshna Navadiya, Roshni Patel, Web Content Mining Techniques-A Comprehensive Survey, International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 10, December- 2012 ISSN: 2278-0181.

[24] Chidansh Amit kumar Bhatt, Mohan S. Kankanhalli, Multimedia data mining: state of the art and challenge. Journal Multimedia Tools and Applications archive, Volume 51 Issue 1, January 2011.

[25] http://en.wikipedia.org/wiki/Web_mining

[26] Mozenda, at: http://www.mozenda.com/web-mining -software Viewed 18 February 2013.

[27] Web Mining https://www. techopedia.com/definition/ 15634/web-mining

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

[29] Jaideep Srivastava, Prasanna Desikan, Vipin Kumar, Web Mining - Concepts, Applications, and Research Directions, AHPCRC technical report 2003-110, July 2003.

[30] Kavita, Priyanka Mahani, Neelam Ruhil, Web data mining: A perspective of research issues and challenges, 2016 3rd International Conference on Computing for Sustainable Global Development, 16-18 March 2016, IEEE.

[31] Rim Rekik, Ilhem Kallel, Jorge Casillas, Adel M.Alimi, Assessing web sites quality: A systematic literature review by text and association rules mining, International Journal of Information Management, Volume 38, Issue 1, February 2018, Pages 201-216.

[32] Imran Khan, Asif Khan, Riaz Ahmed Shaikh, Object analysis in image mining, 2015 2^{nd} International Conference on Computing for Sustainable Global Development (INDIACom), 11-13 March 2015, IEEE.

[33] Yeqing Li, Research on Technology, Algorithm and Application of Web Mining, 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), 21-24 July 2017, IEEE, DOI: 10.1109/CSE-EUC.2017.152.