

## International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

*IJCSMC, Vol. 3, Issue. 5, May 2014, pg.331 – 341*

### **RESEARCH ARTICLE**

# A STUDY ON ISSUES AND TECHNIQUES OF WEB MINING

**R. Munilatha<sup>1</sup>, K.Venkataramana<sup>2</sup>**

<sup>1</sup>Student, M.Tech, KMM Institute of Technology and Science, Tirupati, Andhra Pradesh, India

<sup>2</sup>Associate professor, KMM Institute of Technology and Science, Tirupati, Andhra Pradesh, India

<sup>1</sup> Email: [latha\\_ramineni2000@yahoo.com](mailto:latha_ramineni2000@yahoo.com), <sup>2</sup> Email: [ramanakv4@gmail.com](mailto:ramanakv4@gmail.com)

---

*Abstract: The hasty growth of the web is causing the stable growth of information, leading to several problems such as an increased difficulty of extracting potentially useful knowledge. The huge amount of information available online, the World Wide Web is a fertile area for data mining research. The research in web mining aims to develop new techniques to effectively extract and mine useful knowledge or information from these web pages. Due to the heterogeneity and lack of structure of Web data, automated discovery of targeted or unexpected knowledge/information is a challenging task. In this paper, we survey the research in the area of Web mining, point out the categories of Web mining and variety of techniques used in those categories. In this paper we elicit research scope in the areas of web usage mining, web content mining, web structure mining and concluded this study with a brief discussion on data managing, querying, representation issues.*

*Keywords: Web, data mining, sequential pattern, page rank, hits, hyper link analysis, database view*

---

## I. INTRODUCTION

The World Wide Web (WWW) is continuously growing with rapid increase of the information transaction volume and number of requests from Web users around the world. For web administrator's and managers, discovering the hidden information about the users' access or usage patterns has become a necessity to improve the quality of the Web information service performances. From the business point of view, knowledge obtained from the usage or access patterns of Web users could be applied directly for marketing and management of E-business, E-services, E-searching, and E-education and so on. The following problems will be encountered during interacting with the web.

**a. Finding relevant information:** People either browse or use the search service when they want to find specific information on the Web. When a user uses search service he or she usually inputs a simple keyword query and the query response is the list of pages ranked based on their

similarity to the query. However today's search tools have the following problems [3]. The first problem is low precision, which is due to the irrelevance of many of the search results. This results in a difficulty finding the relevant information. The second problem is low recall, which is due to the inability to index all the information available on the Web. This results in a difficulty finding the un-indexed information that is relevant.

**b. Creating new knowledge out of the information available on the Web:** Actually this problem could be regarded as a sub-problem of the problem above. While the problem above is usually a query-triggered process (retrieval oriented), this problem is a data-triggered process that presumes that we already have a collection of Web data and we want to extract potentially useful knowledge out of it (data mining oriented). Past research [4; 5; 6] focuses on utilizing the Web as a knowledge base for decision making.

**c. Personalization of the information:** This problem is often associated with the type and presentation of information, since it is likely that people differ in the contents and presentations they prefer while interacting with the Web. On the other hand, the information providers could encounter these problems, among others, when trying to achieve their goals on the Web:

**d. Learning about consumers or individual users:** This is a problem that specifically deals with the problem c above, which is about knowing what the customers do and want. Inside this problem, there are sub-problems such as mass customizing the information to the intended consumers or even to personalize it to individual user, problems related to effective Web site design and management, problems related to marketing, etc.

Web mining techniques could be used to solve the information overload problems above directly or indirectly. However, we do not claim that Web mining techniques are the only tools to solve those problems. Other techniques and works from different research areas, such as database (DB), information retrieval (IR), natural language processing (NLP), and the Web document community, could also be used. By the direct approach we mean that the application of the Web mining techniques directly addresses the above problems. For example, a Newsgroup agent that classifies whether the news is relevant to the user. By the indirect approach we mean that the Web mining techniques are used as a part of a bigger application that addresses the above problems. For example, Web mining techniques could be used to create index terms for the Web search services.

## II. WEB MINING

Web mining - is the application of data mining techniques to discover patterns from the Web. According to analysis targets, web mining can be divided into three different types, which are Web usage mining, Web content mining and Web structure mining.

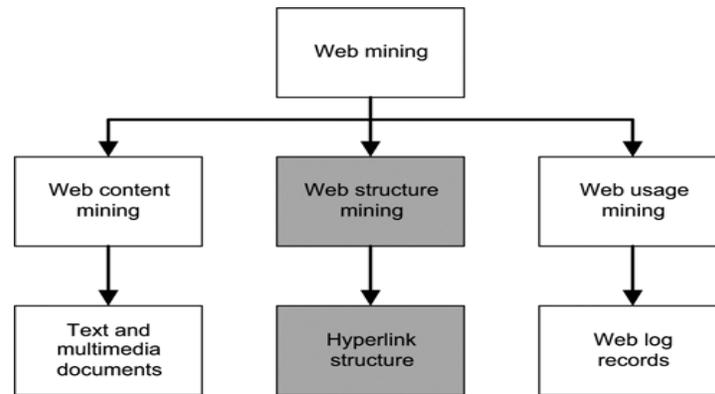


Fig-1

The web content mining mainly relates to the text and multimedia documents and web structure mining relates to the hyperlink structure and web usage mining relates to web log records.

### III. WEB USAGE MINING

Web usage mining is the process of extracting useful information from server logs e.g. use Web usage mining is the process of finding out what users are looking for on the Internet. Some users might be looking at only textual data, whereas some others might be interested in multimedia data. Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web based applications. Web usage mining can also be referred to as automatic discovery and analysis of 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. The goal is to capture, model, and analyze the behavioral patterns and profiles of users interacting with a Web site. The discovered patterns are usually represented as collections of pages, objects, or re-sources that are frequently accessed by groups of users with common needs or interests.

Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site. Web usage mining itself can be classified further depending on the kind of usage data considered:

1. **Web Server Data:** The user logs are collected by the Web server. Typical data includes IP address, page reference and access time.
2. **Application Server Data:** Commercial application servers have significant features to enable e-commerce applications to be built on top of them with little effort. A key feature is the ability to track various kinds of business events and log them in application server logs.
3. **Application Level Data:** New kinds of events can be defined in an application, and logging can be turned on for them thus generating histories of these specially defined events. It must be noted, however, that many end applications require a combination of one or more of the techniques applied in the categories above [2].

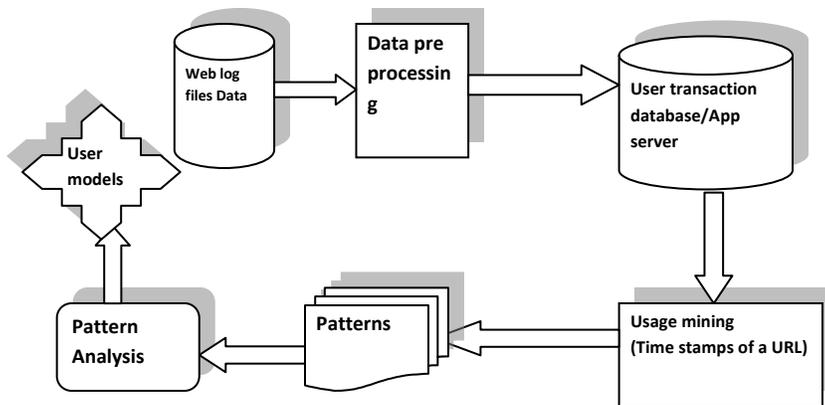


Figure-2 Web Usage Mining

The above diagram (Figure-2) represents the overall Web usage mining process can be divided into three inter-dependent stages: data collection and pre-processing, pattern discovery, and pattern analysis. In the pre-processing stage, the clickstream data is cleaned and partitioned into a set of user transactions representing the activities of each user during different visits to the site. Other sources of knowledge such as the site content or structure, as well as semantic domain knowledge from site ontology's (such as product catalogs or concept hierarchies), may also be used in preprocessing or to enhance user transaction data. In the pattern discovery stage, statistical, database, and machine learning operations are performed to obtain hidden patterns reflecting the typical behavior of users, as well as summary statistics on Web resources, sessions, and users. In the final stage of the process, the discovered patterns and statistics are further processed, filtered, possibly resulting in aggregate user models that can be used as input to applications such as recommendation engines, visualization tools, and Web analytics and report generation tools.

From the initial web usage log, pre-processing yields the required fields of <ip, date\_time, URL>. This refined log is used to identify user sessions. A session is defined as the sequence of URLs visited by the user. The time spent at a URL is determined by the difference in the timestamps of that URL visit and the next. Some of these values were found to be too 'large' to depict the duration of a URL visit. This large duration can be explained by scenarios like idling of the user or ending of a session. The term 'large' can be defined by a maximum limiting variable that can be specified. The time spent at the last URL in a session is required to be estimated as there is no URL succeeding it.

In web usage mining, pattern discovery is difficult because only bits of information like IP addresses and site clicks are available. But analysis of this usage data will yield the information needed for organizations to provide an effective presence to their customers. The most effective way to retrieve useful information from a database is application-dependent.

There are four main mining techniques that can be applied to Web access logs to extract knowledge, but we will focus on algorithms based on association rule mining (ARM) and

sequential mining because of their complexity, applicability and popularity. Here are the four techniques:

1. **Sequential-pattern-mining-based:** Allows the discovery of temporally ordered Web access patterns
2. **Association-rule-mining-based:** Finds correlations among Web pages.
3. **Clustering-based:** Groups users with similar characteristics.
4. **Classification-based:** Groups users into predefined classes based on their characteristics

The above techniques help in the design of better Web sites as well as in the development of effective marketing strategies. However, generally, users are not willing to disclose personal information and may tend to give false information (Cooley, Mobasher, & Srivastava, 1997b). Hence, it is more practical to assume the anonymity of users for WUM in general, especially for non-commercial sites where user registration is not required, and disregard clustering and classification.

#### **A. Sequential pattern mining**

Step1: Access log files analysis with sequential patterns

The general idea is similar to the principle proposed in [7]. It relies on three main steps. First of all, starting from a rough data file, a pre-processing step is necessary to clean "useless" information. The second step starts from this pre-processed data and applies data mining algorithms to find frequent item sets or frequent sequential patterns. Finally, the third step aims at helping the user to analyze the results by providing a visualization and request tool.

Step 1.1: Raw data is collected in access log files by Web servers. Each input in the log file illustrates a request from a client machine to the server (http daemon). Access log files format can differ, depending on the system hosting the Web site. For the rest of this presentation we will focus on three fields: client address, the URL asked for by the user and the time and date for that request. We illustrate

these concepts with the access log file format given by the CERN and the NCSA [4], where a log input contains records made of 7 fields, separated by spaces [17]:host user authuser [date: time] "request" status bytes

Step 1.3: The access log file is then processed in two steps. First of all, the access log file is sorted by address and by transaction. Then each "uninteresting" data is pruned out from the file. During the sorting process, in order to allow the knowledge discovery process to be more efficient, URLs and clients are mapped into integers. Each time and date is also translated into relative time, compared to the earliest time in the log file.

#### **IV. WEB STRUCTURE MINING**

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:

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

2. Mining the document structure: analysis of the tree-like structure of page structures to describe HTML or XML tag usage [2].

Web structure mining helps the users to retrieve the relevant documents by analyzing the link structure of the Web. The challenge for Web structure mining is to deal with the structure of the hyperlinks within the Web itself. Link analysis is an old area of research. However, with the growing interest in Web mining, the research of structure analysis had increased and these efforts had resulted in a newly emerging research area called Link Mining [8], which is located at the intersection of the work in link analysis, hypertext and web mining, relational learning and inductive logic programming, and graph mining. There is a potentially wide range of application areas for this new area of research, including Internet.

The Web contains a variety of objects with almost no unifying structure, with differences in the authoring style and content much greater than in traditional collections of text documents. The objects in the WWW are web pages, and links are in-, out- and co-citation (two pages that are both linked to by the same page). Attributes include HTML tags, word appearances and anchor texts [8]. This diversity of objects creates new problems and challenges since it is not possible to directly make use of existing techniques such as from database management or information retrieval.

#### **A. Hyperlink analysis.**

A Link Analysis Ranking algorithm starts with a collection of Web pages to be ranked. The algorithm then proceeds to extracting the hyperlinks between the pages, and constructing the underlying hyperlink graph. The hyperlink graph is constructed by creating a node for every Web page, and a directed edge for every hyperlink between two pages. The graph is given as input to the Link Analysis Ranking algorithm. The algorithm operates on the graph, and produces a weight for each Web page. This weight captures the authoritativeness of the page, and it is used to rank the pages. Our task is to devise Link Analysis Ranking algorithms that best discover the authoritative nodes in the graph.

Many web pages do not include words that are descriptive of their basic purpose (for example rarely a search engine portal includes the word “search” in its home page) and there exist Web pages which contain very little text (such as image, music, video resources), making a text-based search techniques difficult. However, how others exemplify this page may be useful. This type of “characterization” is included in the text that surrounds the hyperlink pointing to the page. Many researches ([11]; [13]; [15]) have done and solutions have suggested to the problem of searching, indexing or querying the Web, taking into account its structure as well as the meta-information included in the hyperlinks and the text surrounding them. There are a number of algorithms proposed based on the link analysis. Using citation analysis, co-citation algorithm [16] and extended co-citation algorithm [17] are proposed. These algorithms are simple and deeper relationships among the pages cannot be discovered.

Three important algorithms Page-Rank [19], Weighted Page-Rank (WPR) [18] and Hypertext Induced Topic Search (HITS) [20] are discussed below.

### **A. Page rank algorithm**

Page rank algorithm is link analysis algorithm [8] that was discovered by Larry page. This algorithm is used by Google internet search engine. In this algorithm numerical weight is assigned to each element of hyperlink set of document such as World Wide Web, with the purpose of measuring the relative importance of that particular set in that hyperlink. Page rank is a probability distribution algorithm used to represent the person's randomly clicking on links will arrive at any particular page.

A probability is expressed as a numeric value between 0 and 1. That numerical value is defined as damping factor. It is represented as  $d$  and usually its value set to be 0.85. Also  $C(A)$  is the number of link going out of that particular page and is known as back link. Page rank of any page is

Calculated by:  $PR(A) = (1-d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))$  (1)

Where  $PR(A)$  is page rank of particular web page  $A$

$D$  is damping factor.

$PR(T1)$  is page link with main page  $PR(A)$

$C$  is out-link

Note that the rank of a page is divided evenly among its out-links to contribute to the ranks of the pages they point to. Page Rank can be calculated using a simple iterative algorithm, and corresponds to the principal Eigen vector of the normalized link matrix of the web. Page Rank algorithm needs a few hours to calculate the rank of millions of pages. However this algorithm is not so efficient its efficiency is quite less than because it uses only one parameter i.e. back link.

### **B. Hyperlink Induced Topic Search Algorithm (HITS)**

Hyperlink-Induced Topic Search (HITS) is a link analysis algorithm which helps in rating Web pages also known as Hubs and authorities and is developed by Jon Kleinberg. It was a precursor to Page Rank.

The idea behind Hubs and Authorities stemmed from a particular insight into the creation of web pages when the Internet was originally forming; that is, certain web pages, known as hubs, served as large directories that were not actually authoritative in the information that it held, but were used as compilations of a broad catalog of information that led users directly to other authoritative pages. In other words, a good hub represented a page that pointed to many other pages, and a good authority represented a page that was linked by many different hubs

It concludes two main values for a page:

1. Page authority, which estimates the value of the content of the page.
2. Page hub value, which estimates the value of its links to other pages.

First it retrieves the set of results to the search query so that the computation is performed only on this result set and not across all Web pages.

The algorithm performs a series of iterations, each consisting of two basic steps:

**Authority Update:** Update every node's Authority score to be equal to the sum of the Hub Score's of every node that points to it. That is, a node is given a high authority score by being linked to by pages that are recognized as Hubs for information.

**Hub Update:** Update every node's Hub Score to be equal to the sum of the Authority Score's of every node that it points to. That is, a node is given a high hub score by linking to nodes that are considered to be authorities on the subject.

The Hub score and Authority score for a node are defined with the following algorithm:

1. Start with every node having a hub score and authority score of 1.
2. Run the Authority Update Rule
3. Run the Hub Update Rule
4. Normalize the values by dividing every Hub score by the sum of the squares of all Hub scores, and dividing each Authority score by the sum of the squares of all Authority scores.
5. Repeat from the second step as necessary.

## V. WEB CONTENT MINING

Web content mining is the mining, extraction and integration of useful data, information and knowledge from Web page content. The heterogeneity and the lack of structure that permits much of the ever-expanding information sources on the World Wide Web, such as hypertext documents, makes automated discovery, organization, and search and indexing tools of the Internet and the World Wide Web such as Lycos, Alta Vista, WebCrawler, ALIWEB [6], MetaCrawler, and others provide some comfort to users, but they do not generally provide structural information nor categorize, filter, or interpret documents. In recent years these factors have prompted researchers to develop more intelligent tools for information retrieval, such as intelligent web agents, as well as to extend database and data mining techniques to provide a higher level of organization for semi-structured data available on the web. The agent-based approach to web mining involves the development of sophisticated AI systems that can act autonomously or semi-autonomously on behalf of a particular user, to discover and organize web-based information.

Web content mining is differentiated from two different points of view: Information Retrieval View and Database View. Summarized the research works done for unstructured data and semi-structured data from information retrieval view. It shows that most of the researches use bag of words, which is based on the statistics about single words in isolation, to represent unstructured text and take single word found in the training corpus as features. For the semi-structured data, all the works utilize the HTML structures inside the documents and some utilized the hyperlink structure between the documents for document representation. As for the database view, in order to have the better information management and querying on the web, the mining always tries to infer the structure of the web site to transform a web site to become a database.

There are several ways to represent documents; vector space model is typically used. The documents constitute the whole vector space. If a term  $t$  occurs  $n(D, t)$  in document  $D$ , the  $t$ -th

coordinate of  $D$  is  $n(D, t)$ . When the length of the words in a document goes to [corrupted text]. This representation does not realize the importance of words in a document. To resolve this, tf-idf (Term Frequency Times Inverse Document Frequency) is introduced.

By multi-scanning the document, we can implement feature selection. Under the condition that the category result is rarely affected, the extraction of feature subset is needed. The general algorithm is to construct an evaluating function to evaluate the features. As feature set, Information Gain, Cross Entropy, Mutual Information, and Odds Ratio are usually used. The classifier and pattern analysis methods of text data mining are very similar to traditional data mining techniques. The usual evaluative merits are Classification Accuracy, Precision, Recall and Information Score [2].

#### **A. Information Retrieval view**

IR is the automatic retrieval of all relevant documents while at the same time retrieving as few of the non-relevant as possible. Some have claimed that resource or document discovery (IR) on the Web is an instance of Web content mining and the others associate web mining with intelligent IR. Actually IR has the primary goals of indexing text and searching for useful documents in a collection and nowadays research in IR includes modeling, document classification and categorization, user interfaces, data visualization, filtering, etc. The task that can be considered to be an instance of Web mining is Web document classification or categorization, which could be used for indexing. Viewed in this respect, Web mining is part of the (Web) IR process [1].

#### **B. Information Extraction:**

IE has the goal of transforming a collection of documents, usually with the help of an IR system, into information that is more readily digested and analyzed [28]. IE aims to extract relevant facts from the documents while IR aims to select relevant documents [27]. While IE is interested in the structure or representation of a document, IR views the text in a document just as a bag of unordered words [26]. Thus, in general IE works at a finer granularity level than IR does on the documents. Building IE systems manually is not feasible and scalable for such a dynamic and diverse medium such as web contents [25]. Due to this nature of the Web, most IE systems focus on specific web sites to extract. Others use machine learning or data mining techniques to learn the extraction patterns or rules for Web documents semi-automatically or automatically [24]. Within this view, Web mining is used to improve Web IE (Web mining is part of IE) [1].

An example of IE without Web mining is what done by (El-Beltagy, Rafea&Abdelhamid) for building a model for automatically augmenting segments documents with metadata using dynamically acquired background domain knowledge in order to assist users in easily locating information within these documents through a structured front end.

#### **C. Database View**

As mentioned in [21], the database techniques on the Web are related to the problems of managing and querying the information on the Web. According to [21] mentions that there are three classes of tasks related to those problems: modeling and querying the Web, information

extraction and integration, and Web site construction and restructuring. Although the first two tasks are related to Web content mining applications, not all the works there are inside the scope of Web content mining. This is due to the absence of the machine learning or data mining techniques in the process. Basically the DB view tries to infer the structure of the Web site or to transform a Web site to become a database so that better information management and querying on the Web become possible. In which the main objects are traditionally collection of text documents such as audio, video, images etc are embedded in or linked to the web pages as shown in the fig-3.

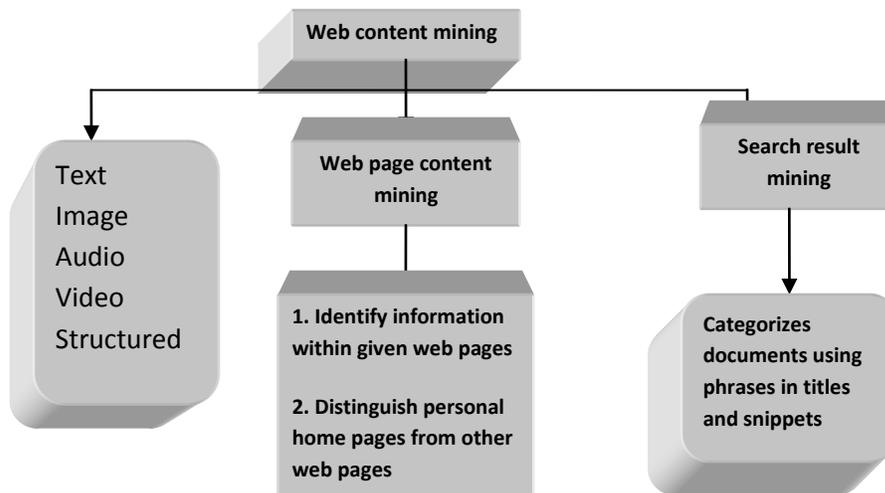


Figure -3 Web Content Mining using DB view

As mentioned previously, the DB view of Web content mining mainly tries to model the data on the Web and to integrate them so that more sophisticated queries other than the keywords based search could be performed. These could be achieved by finding the schema of Web documents, building a Web warehouse or a Web knowledge base or a virtual database. The research done in this area mainly deals with semi-structured data. Semi-structured data from database view often refers to data that has some structure but no rigid schema [22; 23].

#### Conclusion

In this paper, a study on Web mining has given with research point of view. Misperceptions regarding the usage of the term Web mining is elucidated and discussed briefly about web mining categories and various approaches. In this survey, we focus on representation issues, various techniques of web usage mining and web structure mining and information retrieval and extraction issues in web content mining, and connection between the web content mining and web structure mining.

#### References

- [1] Raymond Kosala and Hendrik Blockeel: Web Mining Research: A Survey. ACM SIGKDD, July, 2000
- [2] [http://en.wikipedia.org/wiki/Web\\_mining](http://en.wikipedia.org/wiki/Web_mining).
- [3] S. Chakrabarti. Data mining for hypertext: A tutorial survey. ACM SIGKDD Explorations, 1(2):1–11, 2000.
- [4] W. W. Cohen. What can we learn from the web? In Proceedings of the Sixteenth International Conference on Machine Learning (ICML'99), pages 515–521, 1999.

- [5] M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam, and S. Slattery. Learning to extract symbolic knowledge from the World Wide Web.
- [6] T. M. Mitchell. Machine learning and data mining. *Communications of the ACM*, 42(11):30–36, 1999.
- [7] Prakash S Raghavendra et al. Web Usage Mining using Statistical Classifiers and Fuzzy Artificial Neural Networks at Infonomics Society 2011.
- [8] web usage mining by Bamshad Mobasher. Page No 449-483.
- [9] Horowitz, E., S. Sahni and S. Rajasekaran, 2008. *Fundamentals of Computer Algorithms*. Galgotia Publications Pvt. Ltd., ISBN: 81-7515-257-5, pp: 112-118.
- [10] Broder, A., R. Kumar, F. Maghoul, P. Raghavan and S. Rajagopalan et al., 2000. Graph structure in the web Computing.
- [11] Chakrabarti, S., B. Dom, D. Gibson, J. Kleinberg and R. Kumar et al., 1999. Mining the link structure of the World Wide Web. *IEEE Computer.*, 32: 60-67.
- [12] Haveliwala, T.H., A. Gionis, D. Klein and P. Indyk, 2002. Evaluating strategies for similarity search on the web.
- [13] Varlamis, I., M. Vazirgiannis, M. Halkidi, B. Nguyen and Thesus, 2004. A closer view on web content management enhanced with link semantics. *IEEE Trans. Knowl.*
- [14] Gibson, D., J. Kleinberg and P. Raghavan, 1998. Inferring web communities from link topology. *Proceeding of the 9th ACM Conference on Hypertext and Hypermedia*, June 20-24, ACM Press, PA., USA., pp:225-234. DOI:10.1145/276627.276652
- [15] Kumar, R., P. Raghavan, S. Rajagopalan and A. Tomkins, 1999. Trawling the web for emerging cyber-communities.
- [16] Dean, J. and M. Henzinger, 1999. Finding related pages in the world wide web.
- [17] Hou, J. and Y. Zhang, 2003. Effectively finding relevant web pages from linkage information. *IEEE Trans. Knowl. Data Eng.*, 15: 940-951. DOI: 10.1109/TKDE.2003.1209010
- [18] Xing, W. and A. Ghorbani, 2004. Weighted PageRank algorithm. *Proceeding of the 2nd Annual Conference on Communication Networks and Services Research*, May 19-21, IEEE Computer Society, Washington DC., USA., pp: 305-314. DOI: 10.1109/DNSR.2004.1344743
- [19] S. Brin and L. Page. The anatomy of a large-scale hyper textual Web search engine. In *Seventh International World Wide Web Conference*, 1998. [20] Kleinberg, J., 1999a. Authoritative sources in a hyper-linked environment. *J. ACM*, 46: 604-632. DOI: 10.1145/324133.324140
- [21] D. Florescu, A. Y. Levy, and A. O. Mendelzon. Database techniques for the world-wide web: A survey. *SIGMOD Record*, 27(3):59–74, 1998.
- [22] P. Buneman. Semi structured data. In *Proceedings of the Sixteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pages 117–121. ACM Press, 1997
- [23] S. Abiteboul. Querying semi-structured data. In *Proceedings of Database Theory - ICDT '97*, 6th International Conference, volume 1186 of *Lecture Notes in Computer Science*, pages 1–18. Springer, 1997.
- [24] N. Kushmerick. Gleaning the web. *IEEE Intelligent Systems*, 14(2):20–22, 1999.
- [25] Muslea, I., Minton, S. & Knoblock, C (1998). Wrapper induction for semi structured, web based information sources. In *proceedings of the conference on automatic learning and discovery conald-98*.
- [26] Wilks, Y (1997). Information extraction as a core language technology. Volume 1299 of *lecture notes in computer science*.
- [27] Paziienza, M. T. (1997). Information Extraction: A multidisciplinary approach to an emerging information technology. Volume 1299 of *lecture notes in computer science*.
- [28] J. Cowie and W. Lehnert. Information extraction. *Communications of the ACM*, 39(1):80-91, 1996.