

AN INNOVATIVE APPROACH FOR USER PROFILE BASED ONLINE SEARCHING ON WEB CONTENT

Mahendra Pratap Singh Dohare¹ & Dr. R.S. Jadon²

Abstract:-Web Optimization has the vital role to capitalize a business. It is at the core of search engine optimization (SEO). Dynamic websites are commonly used for e-commerce because they are easier to update and expand but they are subjected to indexing related problems. The objectives of this research work is to explore dynamic websites indexing issues, investigate the SEO tools used major search engines like yahoo, bing and google and develop a prototype to experiment with SEO techniques. We propose to use web structure, web usage mining and user behavior for optimizing the search engine. Web personalisation or user profile base is viewed as an application of information mining supported their profiles, preferences, or expected interests of information on the market to personalization systems. Web structure will give URL encoding, key words, URL of various media types, metadata of the web-site and many more features. The user behavior is analyzed using the click log and user cache. We propose to integrate this feature set to classify the web-page. Further we plan to investigate the use of soft computing based learning techniques for classifying the web page. Obtained results as these web page and its features are further used by SEO.

Keywords: Dynamic Websites, Search Engine Optimization Technique, Search Engine, and User Profile.

I. INTRODUCTION

The World Wide Web (Web) is a popular and interactive medium to disseminate information today. The Web is huge, diverse, and dynamic and thus raises the scalability, multimedia data, and temporal issues respectively. Due to those situations, we are currently drowning in information and facing information overload.

1. Web content mining- Web content mining is the application of data mining techniques to content published on the Internet, usually as HTML (semi structured), plaintext (unstructured), or XML (structured) documents.

2 Web structure mining- Web structure mining operates on the Web's hyperlink structure. This graph structure can Provide information about a page's ranking or authoritativeness and enhance search results through filtering.

3 Web usage mining - Web usage mining analyzes results of user interactions with a Web server, including Web Logs, click streams, and database transactions at a Web site or a group of related sites. Web usage mining introduces privacy concerns and is currently the topic of extensive debate. Information users could encounter, among others, the following problems when interacting with the Web.

4 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 morality to the query. However today's a search tool have the following problems [13]. 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,

5 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). Recent research focuses on utilizing the Web as a knowledge base for decision making

6 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

¹ Dept. of Computer Application, Madhav Institute of Technology & Science, Gwalior, (M.P), India

² Dept. of Computer Application, Madhav Institute of Technology & Science, Gwalior, (M.P), India

7 Learning about client or individual users: This is a problem that specifically deals with the problem c above, which about knows 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.

8 User Behaviors Approach- Here we consider the problem of joint inference of relevance and quality of document snippets by using only search engine click logs. Click logs record the user behavior in response to the retrieved ranked list of documents for every query placed. For each user session, this corresponds to an ordered list of documents on which the user has made a click. Thus, the logs provide implicit feedback about both the summary quality and the relevance of the documents. In this work, our goal is to use click through data to, both, detect quality of document summaries and to infer the relevance of the document.

9 Web Crawler Architecture- A Web crawler is an Internet Robot is short name stand as a bot which systematically browses the World Wide Web, typically for the purpose of Web indexing. Web search engines and some other sites use Web crawling or spidering software to update their web content or indexes of others sites' web content. Web crawlers can copy all the pages they visit for later processing by a search engine which indexes the downloaded pages so the users can search much more efficiently. Crawlers consume resources on the systems they visit and often visit sites without tacit approval. Issues of schedule, load, and "politeness" come into play when large collections of pages are accessed. Mechanisms exist for public sites not wishing to be crawled to make this known to the crawling agent. As the number of pages on the internet is extremely large, even the largest crawler's fall short of making a complete index. Crawlers can validate hyperlinks and HTML code.

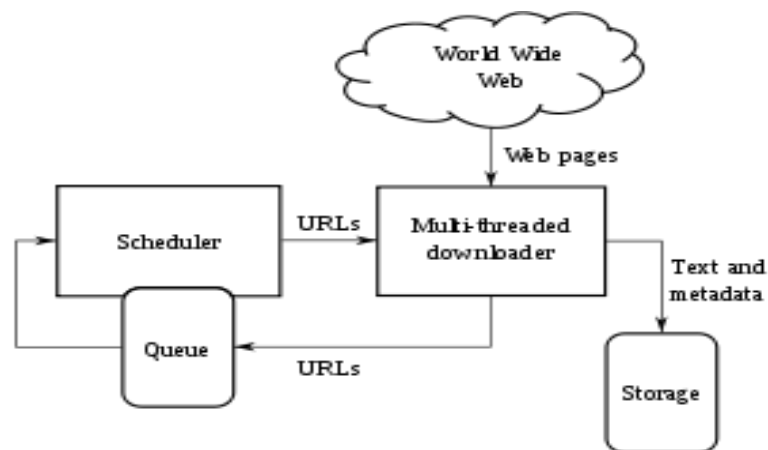


Fig -1: Web Crawler Architecture

II. LITERATURE SURVEY

2.1 WEB MINING

Web mining is the use of data mining techniques to automatically discover and extract information from Web documents and services. This area of research is so huge today partly due to the interests of various research communities, the tremendous growth of information sources available on the Web and the recent interest in e-commerce. This phenomenon partly creates confusion when we ask what constitutes Web mining and when comparing research in this area. We suggest decomposing Web mining into this process working as:

1. **Resource finding:** the task of retrieving intended Web documents.
2. **Information selection and pre-processing:** automatically selecting and pre-processing specific information from retrieved Web resources.
3. **Generalization:** automatically discovers general patterns at individual Web sites as well as across multiple sites.
4. **Analysis:** validation and/or interpretation of the mined patterns.

2.2 WEB CONTENT MINING- Web content mining is the application of data mining techniques to content published on the Internet, usually as HTML (semi structured), plaintext (unstructured), or XML (structured) documents. In this section we list some of the research in the respective Categories in separate tables. We should note that the lists are by no means complete. We just intend to give a taste on the variety of some representations, processes, methods, and applications that have been used

2.3 WEB STRUCTURE MINING

If in the database view of Web content mining we are interested in the structure within Web documents (intra-document structure), in Web structure mining we are interested in the structure of the hyperlinks within the Web itself (inter-document structure). This line of research is inspired by the study of social networks and citation analysis [12]. With social network analysis we could discover specific types of pages (such as hubs, authorities, etc.) based on the incoming and outgoing links. Web structure mining utilizes the hyperlinks structure of the Web to apply social network analysis to model the underlying links structure of the Web itself.

2.4 WEB USAGE MINING

Web usage mining focuses on techniques that could predict user behavior while the user interacts with the Web. As mentioned before, the mined data in this category are the secondary data on the Web as the result of interactions. These data could range very widely but generally we could

Classify them into the usage data that reside in the Web clients, proxy servers and servers [11]. The Web usage mining process could be classified into two commonly used approaches [12]. The first approach maps the usage data of the Web server into relational tables before an adapted data mining technique is performed. The second approach uses the log data directly by utilizing special pre-processing techniques. As is true for typical data mining applications, the Issues of data quality and pre-processing are also very important here. The typical problem is distinguishing among unique users, server sessions, episodes, etc. in the presence of caching and proxy servers [13].

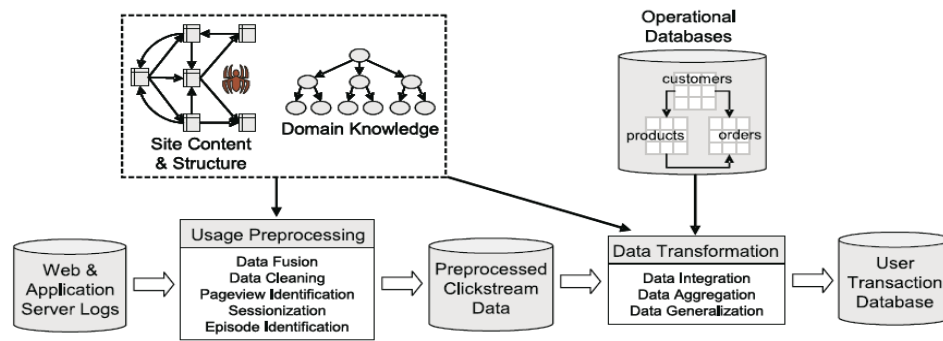


Fig-2: Primary Tasks in Data Pre-processing for Web Usage Mining

2.5 Encoding URLs for Dynamic Websites

URLs are entering door to webpage for search engines and website's users. Therefore, SEO experts/web developers are convinced to optimize URLs by making them short, simple and readable for both website's users and search engine point of view. There are some techniques for URL rewriting and redirecting these important to consider for both static and dynamic websites under some circumstances. The subsection below is provided with the brief details of some technique for URL encoding.

2.6 Search Engine for Friendly websites

The following section discusses some of the elements which become a need of website when it is needed to make easy navigation as well as easy interaction with website. On the other hand, these elements can cause obstacles in the attempt towards usability and their search engine friendliness of website. For those reasons, they need attention to optimize them if it is obliging to use elements in website design.

Optimizing Frames

Frames in website facilitate website's users to see/find contents of several webpage(s) on a single (current) webpage; this feature reduces trouble of following links; moreover, a deep hierarchy to read target page might risk that website's users will come back to the current page. In contrast, use of frame in webpage gives tough time to search engines spiders; as, frames do not correspond to the web's conceptual model [5] and [6] and [7].

Redirecting

The use of redirects indicates both web browsers and search engine spiders that content will be fetched from new URL. Whenever, old webpage (old-page.html) be followed, it will be redirected to new webpage (new-page.html). Without redirects both website's user and search engine crawler will get 404 error (Page not Found Error). The page not found error will result the loss of traffic of website. Moreover, this problem restricts search engines to find your webpage contents. In SEO campaign this problem can be quite paying in both cases [8] and [9] and [10].

III. PROPOSED WORK

Problem Statement - How to Understanding or Identify the behavior of User/client/customer/consumers is one of the big challenges for online searching marketers? When I was scrolling around on the Internet .this kind of problem occur .these problem is major problem to search all relevant and irrelevant data. There are plenty of books, techniques and models that aim to help us gain this kind of insight. In this is research mainly We Proposed a user Behavior model on based of their Log data. Further we focus on categories of user based on their cache Management data in searching style behavior. Their behaviors have to understand by database server/search engine/websites.

Proposed Work - The objective of proposed research work, mainly focus on perspective, better understanding User Behavior, Improve better performance in the criteria of time, space and more accurate relevant data or information provide to all online searching users and also filling research gap in exciting knowledge with meaningful impact upon the field of web mining study.

User Behavior Approach:

We propose a graphical model that describes the underlying user behaviour in a typical query session. Our probabilistic model directly models user clicks, thereby disentangling summary and relevance; Note that summary is document-specific while a document's relevance to a query depends on other documents. By considering the effects of presence of one document in the ranking with the others, we can obtain information that disentangles summary from relevance.

Web Usage Mining-It is the process by which we identify the browsing patterns by analyzing the navigational behavior of user. It focuses on techniques that can be used to predict the user behavior while the user interacts with the web. It uses the secondary data on the web. This activity involves the automatic discovery of user access patterns from one or more web servers.. Table 1 gives an overview of above mining categories

Web Search Engine - A web search engine is a software system that is designed to search for information on the World Wide Web. The search results are generally presented in a line of results often referred to as search engine results pages (SERPs). The information may be a mix of web pages, images, and other types of files. Some search engines also mine data available in databases or open directories. Unlike web directories, which are maintained only by human editors, search engines also maintain real-time information by running an algorithm on a web crawler.

User Log Data- This is the procedure where the information stored in Web server logs is processed by applying data mining techniques in order to (a) extract statistical information and discover interesting usage patterns, (b) cluster the users into groups according to their navigational behavior, and (c) discover potential correlations between Web pages and user groups. This process of extracting information concerning the browsing behavior of the users.

User Profile-User profile data provide information about the users of a Web site. A user profile contains demographic information (such as name, age, country, marital status, education, interests etc.) for each user of a Web site, as well as information about users' interests and preferences. Such information is acquired through registration forms or questionnaires, or can be inferred by analyzing Web usage logs.

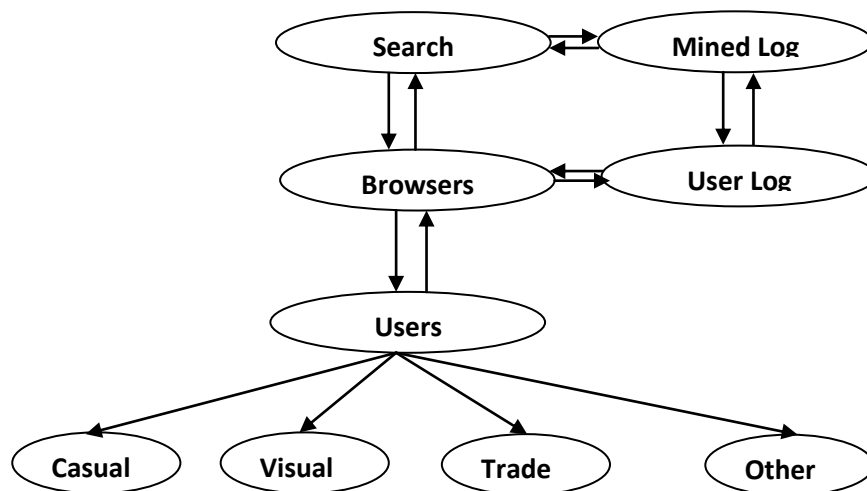


Fig- 3 User Behavior with log data Model

Web Personalization -Web site personalization can be defined as the process of customizing the content and structure of a Web site to the specific and individual needs of each user taking advantage of the user's navigational behavior. The steps of a Web personalization process include: (a) the collection of Web data, (b) the modeling and categorization of these data (preprocessing phase), (c) the analysis of the collected data, and (d) the determination of the actions that should be performed. The ways that are employed in order to analyze the collected data include content-based filtering, collaborative filtering, rule-based filtering, and Web usage mining.. This is based on the assumption that users with similar behavior (e.g. users that rate similar objects) have analogous interests. In rule-based filtering the users are asked to answer a set of questions. These questions are derived from a decision tree, so as the user proceeds to answer them, what he finally receives as a result is tailored to his needs. Content-based, rule-based, and collaborative filtering may also be used in combination, for deducing more accurate conclusions. In this work we focus on Web usage mining. This process relies on the application of statistical and data mining methods to the Web log data, resulting in a set of useful patterns that indicate users' navigational behavior. The data mining methods that are employed are: association rule mining, sequential pattern discovery, clustering, and classification. This knowledge is then used from the system in order to personalize the site according to each user's behavior.

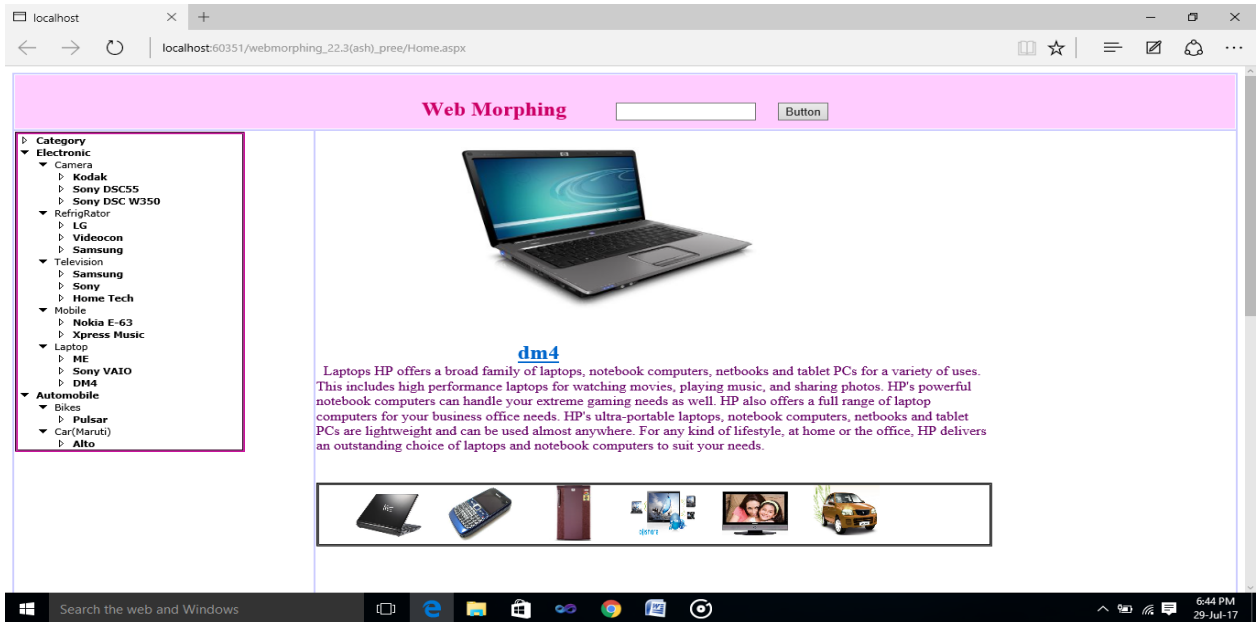


Fig: 4 User categories based online Searching on web content

IV. ANALYSIS OF PROPOSED OPTIMIZATION TECHNIQUES

link analysis algorithms-Web Mining technique provides the additional information through hyperlinks where different documents are connected [2]. We can view the web as a directed labeled graph whose nodes are the documents or pages and edges are the hyperlinks between them. This directed graph structure is known as web graph. There are number of algorithms proposed based on link analysis. Three important algorithms Page Rank , Weighted Page Rank and HITS (Hyper-link Induced Topic Search)[15] are discussed below.

4.1 Page Rank-This algorithm was developed by Brin and Page at Stanford University which extends the idea of citation analysis [5]. In citation analysis the incoming links are treated as citations but this technique could not provide fruitful results because this gives some approximation of importance of page. So Page Rank provides a better approach that can compute the importance of web page by simply counting the number of pages that are linking to it. These links are called as back links.

4.2 Weighted Page Rank-This algorithm was proposed by Wenpu Xing and Ali Ghorbani which is an extension of Page Rank algorithm [7]. This Algorithm assigns rank values to pages according to their importance rather than dividing it evenly. The importance is assigned in terms of weight values to incoming and outgoing links.

4.3 HITS (Hyper-link Induced Topic Search)

Kleinberg gives two forms of web pages called as hubs and authorities. Hubs are the pages that act as resource lists. Authorities are pages having important contents. A good hub page is a page which is pointing to many authoritative pages on that content and a good authority page is a page which is pointed by many good hub pages on the same content. A page may be a good hub and a good authority at the same time [8, 9].

The HITS algorithm treats WWW as directed graph $G(V, E)$; where V is a set of vertices representing pages and E is set of edges corresponds to link. Figure 1 shows the hubs and authorities in web [2].

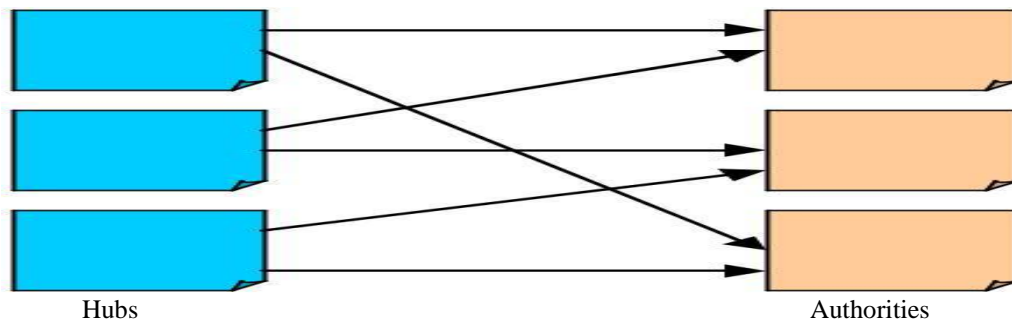


Figure 5: Hubs and Authorities

It has two steps:

1. Sampling Step: - In this step a set of relevant pages for the given query are collected.
2. Iterative Step: - In this step Hubs and Authorities are found using the output of sampling step.

Following expressions (a,b) are used to calculate the weight of Hub (H_p) and the weight of Authority (A_p).

$$H_p = \sum A_q \text{ where } q \text{ is belongs to } I(p) \dots\dots\dots (a)$$

$$A_p = \sum H_q \text{ where } q \text{ is belongs to } B(p) \dots\dots\dots (b)$$

here H_q is Hub Score of a page, A_q is authority score of a page, $I(p)$ is set of reference pages of page p and $B(p)$ is set of referrer pages of page p, the authority weight of a page is proportional to the sum of hub weights of pages that link to it. Similarly a hub of a page is proportional to the sum of authority weights of pages that it links to.

4.3.1 Constraints with HITS algorithm

Following are some constraints of HITS algorithm [10]

-Hubs and authorities: It is not easy to distinguish between hubs and authorities because many sites are hubs as well as authorities.

- Hubs and authorities: It is not easy to distinguish between hubs and authorities because many sites are hubs as well as authorities
- Topic drift: Sometime HITS may not produce the most relevant documents to the user queries because of equivalent weights
- Automatically generated links: HITS gives equal importance for automatically generated links which may not have relevant topics for the user query
- Efficiency: HITS algorithm is not efficient in real time.

HITS was used in a prototype search engine called Clever for an IBM research project. Because of the above constraints HITS could not be implemented in a real time search engine.

V. COMPARASION STUDY OF PROPOSED OPTIMIZATION TECHNIQUES

Page Rank VS Weighted Page Rank

In order to compare the WPR with the Page Rank, the resultant pages of a query are categorized into four categories based on their relevancy to the given query. [8] They are

1. Very Relevant Pages (VR): These are the pages that contain very important information related to a given query.
2. Relevant Pages (R): These Pages are relevant but not having important information about a given query.
3. Weakly Relevant Pages (WR): These Pages may have the query keywords but they do not have the relevant information.
4. Irrelevant Pages (IR): These Pages are not having any relevant information and query keywords.

$$K = \sum (N_i * w_i) \text{ where } I \text{ is belongs to } R(p) \dots\dots\dots (1)$$

The Page Rank and WPR algorithms both provide ranked pages in the sorting order to users based on the given query. So, in the resultant list, the number of relevant pages and their order are very important for users. Relevance Rule is used to calculate the relevancy value of each page in the list of pages. That makes WPR different from Page Rank. Relevancy Rule: The

Relevancy Rule is as given in (1). The Relevancy of a page to a given query depends on its category and its position in the page-list. The larger the relevancy value, the better is the result

$$W_i = (V_1, V_2, V_3, V_4) \dots \dots \dots (2)$$

Where i denotes the i^{th} page in the result page-list $R(p)$, n represents the first n pages chosen from the list $R(p)$, and W_i is the weight of i^{th} page as given in (2).

where v_1, v_2, v_3 and v_4 are the values assigned to a page if the page is VR, R, WR and IR respectively. The values are Always $v_1 > v_2 > v_3 > v_4$. Experimental studies show that WPR produces larger relevancy values than the Page Rank.

TABLE 1: Web Mining Categories

Web Mining				
	Web Content Mining		Web Structure Mining	Web Usage Mining
	IR view	DB View		
View of Data	-Unstructured -Structured	-Semi Structured -Web Site as DB	-Link Structure	-Interactivity
Main Data	- Text documents -Hypertext documents	-Hypertext documents	-Link Structure	-Server Logs -Browser Logs
Representation	-Bag of words, n-gram Terms, -phrases, Concepts or ontology -Relational	-Edge labeled Graph, -Relational	-Graph	-Relational Table -Graph
Method	-Machine Learning -Statistical (including NLP)	-Proprietary algorithms -Association rules	-Proprietary algorithms	-Machine Learning -Statistical -Association rules
Application Categories	-Categorization -Clustering -Finding extract rules -Finding patterns in text	-Finding frequent sub Structures -Web site schema Discovery	-Categorization -Clustering	-Site Construction -adaptation and management -Marketing, -User Modeling

COMPARISON - Table 2 shows comparison of all the three algorithms

Algorithm	PageRank	Weighted PageRank	HITS
Mining Technique Used	WSM	WSM	WSM & WCM
Working	Computes Scores at Indexing time. Results are Sorted According to Importance of Pages.	Computes Scores at Indexing Are Sorted According to Page Importance.	Computes Hub & Authority Scores of n Highly Relevant pages on the fly.
I/P Parameters	Backlinks	Backlinks, Forward Links	Backlinks, Forward Links & Content
Complexity	$O(\log N)$	$<O(\log N)$	$<O(\log N)$
Limitations	Query Independent	Query Independent	Topic drift And

			Efficiency Problem
Search Engine	Google	Research Model	Clever

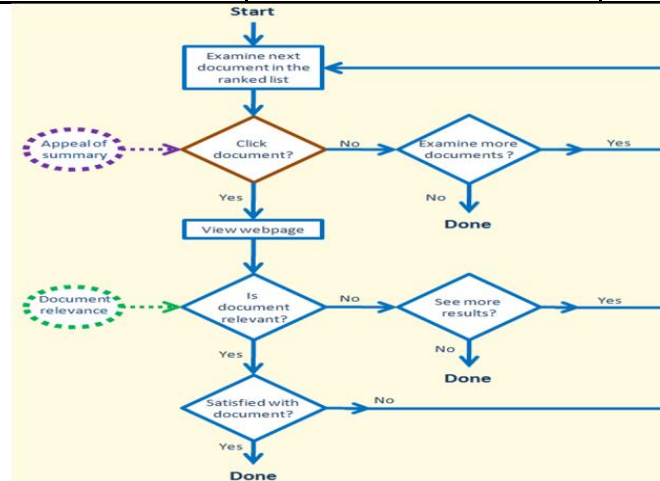


Figure-6: User Model

VI. CONCLUSION

SEO is essential for websites to increase their visibility in search engines. Since a properly indexed websites are an initial step towards better ranking to bring websites more visible in search engines, it is important to optimize a websites properly to increase credibility of a websites. In this research we have presented an approach to perform a detailed study on techniques that can help achieve a real website acquire higher ranking. And I would integrate obtained result from user click logs and URL encoding. I would also like to explore that whether it is possible to achieve top and better ranking in Google, Yahoo and Bing at the same time by using the web structure mining, web usage mining and maintain user logging data to draw comparisons matrix after that we integrate both techniques and obtained web pages will be used for Search engine. Here we also another problem of joint implication of relevance and quality of document scraps by using only search engine click logs. Click logs record the user behavior in response to the retrieved ranked list of documents for every query placed. For each user session, this corresponds to an ordered list of documents on which the user has made a click. As part of this Research work, we are planning further to carry out performance analysis of PageRank and Weighted PageRank and working on finding the ways to categorize the users and web pages on their web personalization to obtain the better PageRank results.

REFERENCES

- [1] G. Rogan; "A study of Search Engine Optimization methods", National University of Ireland, Gateway (2009).
- [2] N. Nazar;, "Exploring SEO Techniques for Web 2.0 Websites", Master of science Thesis in software Engineering and Technology, Department of CS and Engineering Chalmers University Of Technology Goteborg, Sweden, (June 2009)
- [3] R. Vadivel & Dr. K. Baskaran; "Implementing Search Engine Optimization Technique to Dynamic/Model View Controller Web Application", Global Journal of the Computer Science and Technology Vol. 10 Issue 6, V1.0, (July 2010)
- [4] C Duda, G Frey, D. Kossmann & C. Zhau,"AJAXSearch Crawling, Indexing & Searching Web 2.0 Applications", Prfocceding of the VLDB Endowment v.1 n.2, (2008)
- [5] X. Zhang, M. Zuo, & Q. Liu,"Analysis of the Reasons Why Invisible Web Can't Be Seen and its Effective Retrieval Strategies", Innovative Computing Information and Control (ICICIC) '08 3rd International Conference on, vol., no., pp.563-563, 18-20, (June 2008)
- [6] M. Alfano & B. Lenzitti, "A Web Search Methodology for Different User Typologies" International Conference on Computer Systems and Technologies-CompSysTech'09, (2009)
- [7] C. Sheram, "Google Power: Unleash the full potential of Google", McGraws-Hill /Osborne Companies, ISBN 0-07-225787-3, (2005)
- [8] Y. Ru & E. Horowitz, "Indexing the invisible web: a survey", Department of Computer Science, University of Southern California, Los Angeles, California, USA page 2, (2005)
- [9] L. Huang,"Challenging the Invisible Web: Improving Web Meta-Search by Combining Constraint-based Query Translation and Adaptive User Interface Construction", The Department of Computer Science of Technische University, Darmstadt, Germany, (2003)
- [10] J. Kohne,"Optimizing a large dynamically generated website for search engine crawling and ranking", Department of Media- and Knowledge Engineering Technical University of Delft Netherlands, (2006)
- [11] Sadegh Kharazmi, Ali Farahmand Nejad, Hassan Abolhassani," Freshness of Web Search Engines: Improving Performance of Web Search Engines Using Data Mining Techniques", IEEE 2009.
- [12] Nicholas Carroll, Taylor & Francis, "Search Engine Optimization and User Behavior", Encyclopedia of Library and Information Sciences. February 16, 2010.
- [13] Fabricio Benevenutoy, Tiago Rodriguesy, Meeyoung, ChaVirgilio Almeida, "Characterizing User Behavior in Online Social Networks", IMC'09, November 4-6, 2009, Chicago, Illinois, USA.
- [14] Raymond Kosala, Hendrik Blockeel, Web Mining Research: A Survey, ACM SIGKDD, July 2000.
- [15] Rekha Jain, Page Ranking Algorithms for Web Mining, International Journal of Computer Applications (0975 - 8887), Volume 13- No.5, January 2011.
- [16] Aleecia M. McDonald, Lorrie Faith Cranor, Beliefs and Behaviors: Internet Users Understanding of Behavioral Advertising", TPRC, August 16, 2010.

- [17] Christina Lagerstedt, Andreas Aurelius, Hemamali Pathirana, Claus Popp Larsen Netlab Acreo AB Kista, Olle Findahl, Understanding Internet User Behavior: Towards a Unified Methodology, International Journal on Advances in Telecommunications, vol5no3&4, year2012.
- [18] Long Jin, San Diego Yang Chen, Tianyi Wang, Pan Hui, V. Vasilakos, "Understanding User Behavior in Online Social Networks: A Survey", IEEE Communications Magazine September 2013.
- [19] Claude Castelluccia, Technical Perspective Behavioral Tracking on the Internet, Springer Science Business Media B.V. 2012.
- [20] Minky Jindal, Nisha kharb International "Data Mining in Web Search Engine Optimization and User Assisted Rank Results, "Journal of Computer Applications (0975 – 8887) Volume 95– No.8, June 2014.
- [21] Fabrício Benevenuto, Tiago Rodrigues, Meeyoung Cha, Virgílio Almeida, "Characterizing User Behavior in Online Social Networks", IMC'09, November 4–6, 2009, Chicago, Illinois, USA.
- [22] Anna Mastoral, Sarantos Kapidakis, Maria Monopoli, "Exploring Users' Online Search Behaviour: a preliminary study in a library collection"
- [23] Long Jin, San Diego, Yang Chen, Tianyi Wang, Pan Hui and V. Vasilakos, "Understanding User Behavior in Online Social Networks: A Survey", IEEE Communications Magazine September 2013.
- [24] Irene Garrigós, Jaime Gómez "Modeling User Behaviour Aware Web Sites with PRML" WISM'06 1087
- [25] Nicholas Carroll by "Search Engine Optimization and User Behavior" February 16, 2010 And Republished Understanding Information Retrieval Systems: Management, Types, and Standards, 2011.
- [26] Xiaopeng Deng, Chunxiao Xing "A QoS-oriented Optimization Model for Web Service Group", 2009 Eighth IEEE/ACIS International Conference on Computer and Information Science.
- [27] Robert D. Callaway, Michael Devetsikiotis, "Chunk and Object Level Deduplication for Web Optimization: A Hybrid Approach", IEEE ICC 2012 - Communication QoS, Reliability and Modeling Symposium.
- [28] Alejandro A. Vaisman Gabriel Dandretta Mariela Sapia, "Enhancing Web Access Using Data Mining Techniques" Proceedings of the 14th International Workshop on Database and Expert Systems Applications 1529-4188/03/2003 IEEE.
- [29] Pranam kolari and anupam joshi university of maryland, baltimore county "webmining: research and practice", 1521-9615/04/2004 IEEE co published by the IEEE CS and the AIP, July/August 2004.
- [30] Xindong wu, xingquan zhu gong-qing wu, and wei ding, "data mining with big data", IEEE transactions on knowledge and data engineering, vol. 26, no. 1, January 2014.
- [31] Lei xu, chunxiao jiang, jian wang, jian yuan, and yong ren, "information security in big data: privacy and data mining", received September 21, 2014, accepted October 4, 2014, date of publication October 9, 2014, access.2014.2362522.
- [32] A. K. Tripathy, A. K. Singh, "An Efficient Method Of Eliminating Noisy Information In Web Pages for Data mining", Proceedings of the Fourth International Conference on Computer and Information Technology 0-7695-2216-5/ 2004 IEEE.
- [33] Magdalini Eirinaki and M. Vazirgiannis. "Web Mining for Web Personalization" ACM Transactions on Internet Technology (2003).