

## Advance Technique for Website Navigation Improvement Using Machine Learning

Leelashankar Salitra, Prof. Namrata Tapaswi

*IES-IPS Academy , Indore*

*lalitsalitra89@gmail.com, namrastapaswi@gmail.com*

**ABSTRACT :** *Web searching and information retrieval in a website is a extremely recurrent activity in web browsing. When a customer again finding a website particularly after a long time, he comes across some problems between which beings ignorant regarding the information structure of the site is extremely prominent a different problem is that in a website, customer have no technique to systematize the information space or the navigational structure of the website. Our aim is to improve Web navigation efficiency by reorganizing Web structure. To avoid customer to lose their orientation, structure stability is taken into consideration. We are proposing a technique to progress the navigation effectiveness of a website while minimizing changes to its existing structure we present a survey of the use of Web mining for Web transformation approaches.*

**Keywords-** *Clustering, Navigability, Web Mining, Web Personalization, Web Transformation.*

**I.INTRODUCTION :** Website navigation has been appear as one of the mainly significant design features transversely many domains with finance, e-commerce, activity, learning, administration, and medical. A major reason is that the web developers lenient of how a website must be structured can be considerably dissimilar from that of the customer. web site moreover enclose a lot of links and pages each user need dissimilar pages at similar time or identical user might access dissimilar pages at different time. As customer increase over internet we require to build web intelligent we apprehension here concerning intelligent website. To build web site intellectual we have to identify what is content of website, which are customer and how website structured every one this identified as web mining. Now as data mining plan at discover expensive information that is hidden in conservative databases, the promising field of Web mining plan at verdict and extract appropriate information that is hidden in Web associated data. Web mining is a multi-disciplinary attempt that draw method from field similar to information retrieval, statistics, machine learning, natural language dispensation, and others. Basically a Web mining deal among three main areas web content mining, web usage mining and web structure mining. for prefer web structure mining is because web site is huge source of information, except users frequently browsing ineffective page which irritate customer and customer lost attention from searching data over website. A most important cause of [www.ijarse.com](http://www.ijarse.com)

poor web design is to the web developers considerate of how a website must be structured can be significantly dissimilar from those of the customer however, the determine of website efficiency should be the approval of the users somewhat than that of the developers. Thus, Web pages must be prearranged in a method that usually matches the user's replica of how pages be supposed to be prepared. There are two conduct to advance customer navigability web personalization and web conversion. Web personalization deal with customer behavior and customer profile, sessions and history of data moreover called as web logs which is formed by user's action on web site, but alteration come up to mostly focus on increasing technique to totally rearrange the link structure of a website. We accept web transformation method to help user navigation. Proposed approach are links based clustering to modify structure of website as we recognize clustering is one method of data reduction so lager web usage data is utilize as compacted and time of reorganize web site structure is concentrated. In later section we provide details of some popular clustering algorithm and comparison between clustering algorithm with respect to time. Our propose technique addresses how to improve a website without introducing substantial changes.

**II. RELATED WORK:** In this research, we have distinguished a method that discovers the gap among Web site designer prospect and customer behavior. The previous are assessed by measure the inter-page intangible significance and the final by measure the inter-page entrance co-occurrence.

Chen and Young [1] examine the questions of how to get better user navigation in a website with minimal modify to its structure. It complements the literature of alteration approach that focuses on reconstruct the link structure of a website. Their proposed model is appropriate for website continuation and can be functional in a regular manner.

Biswas et al. [2] propose a narrative approach named Dashboard which supports structure custom navigational menus system and permit each user construct her own hierarchical information structure. They was demonstrated how this innovative approach help rapidly re-finding information on re-visitation of a website.

Wang and Yen [3] improves navigation efficiency precisely and relieve the expensive of tedious chore to adapt the structure in transformation. If several constraints, e.g. structural constancy constraint, is relaxed, there is still room for development. Some improvements are important but will saturate or even reduce as after addition a positive number of links.

Lin and Tseng [4] proposed model decreases the information excess and search depth for user surfing the web. A heuristic approach has as well been planned to decrease the necessary computation time. Though, the heuristic approach connecting two successive 0-1 programming models immobile require a extremely long computation time to discover the optimal solution, particularly when the website contains a lot of hyperlinks. To determine the effectiveness problem, proposes an ant colony system to reorder website structure. The proposed algorithm is experienced lengthily with numerical examples. Moreover, an experiential study with a real-world website is conduct to confirm the algorithm applicability.

[www.ijarse.com](http://www.ijarse.com)

Martin-Guerrero et al. [5] estimate the feasibility of implementing a recommender in Web portal is proposed. The primary step of the proposed method is to cluster user data based on simulation in order to make sure that the collaborative filter is robust crosswise a range of user models. These consequences demonstrated the predictive accuracy of cluster based recommender systems with the ART2 neural network algorithm practical to profiles of replicated user data. This predictive accuracy then supports the present of services that are novel to the user.

Kim and Shim [6] present narrative algorithms for extract templates from a huge number of web documents which are generate from heterogeneous templates. We cluster the web credentials based on the resemblance of essential template structures in the documents so that the template for each cluster is extract concurrently. They have developed a novel goodness compute with its quick estimate for clustering and present comprehensive investigation of their algorithm. With real-life data sets corroborate the efficiency and strength of their algorithm compare to the state of the art for template detection algorithms.

**III. PROPOSED METHODOLOGY:** The confront for web structure mining is to arrangement with the configuration of the hyperlinks inside the web itself. Link study is an getting on region of research. However, through the rising attention in web mining, the investigate of structure investigation had improved and these efforts had resulted in a recently rising research region called link mining [1][7], which is situated at the connection of the effort in link analysis, hypertext and web mining, relational knowledge and inductive logic encoding, and graph mining. There is a potentially broad range of function area for this novel area of research, with WWW. The web contain a diversity of objects with about no unite structure, with difference in the authoring style and content a group improved than in normal collection of text documents. The objects in the internet are web pages, and links are in-, out- and co-citation. Attributes contain html tags, word manifestation and anchor texts [2][7]. this variety of objects create novel problems and challenge, since is not probable to directly complete use of obtainable method such as from database supervision or in succession retrieval. Link mining had created a few protests on several of the traditional data mining tasks. as follow link-based categorization is the nearly all new improve of a typical data mining task to connected domains [7]. the task is to focal point on the prediction of the group of a web page, support on words that happen on the page, links among pages, anchor text, html tags and other likely attributes establish on the webpage.

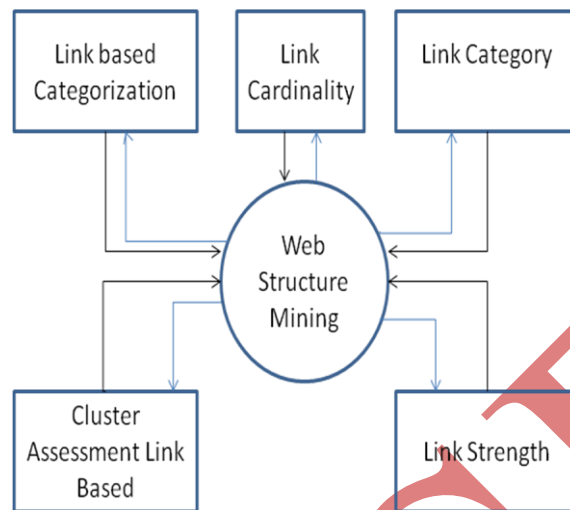


Fig.1 Website navigation improvement

The objective in cluster study is to discover obviously happening sub-classes. the data is segmented into group, where alike substance are group jointly, and different objects are group into dissimilar groups preceding task, link-based cluster study is unverified and can be use to find out hidden patterns beginning data. There are a extensive range of odd jobs about the prediction of the continuation of links, such as predicting the style of link among two entities, or predicting the rationale of a link.

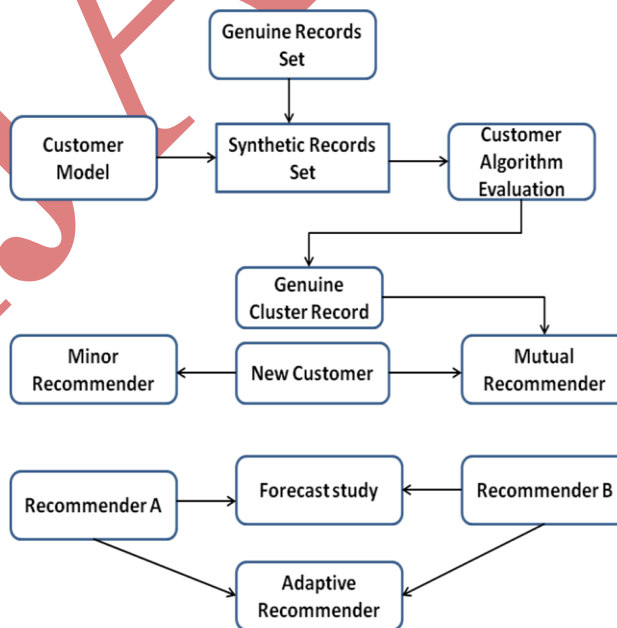


Fig.2 Technique for clustering web page

Links could be connected with weights. The foremost task here is to calculate the quantity of links among objects. There is a lot of conduct to use the link structure of the web to make notions of authority. The major goal in increasing function for link mining is to complete superior use of the accepting of these

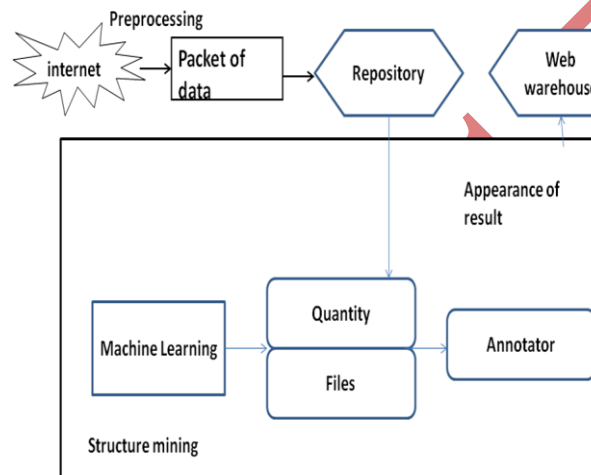


Fig.3 Link extraction technique

Intrinsic social association of the web) pages can't be complex in regulate to adjust the construction. Real-world data lean to be dirty, partial and incompatible. Data preprocessing techniques in machine learning can get improved the quality of the data, thereby segment to recover the precision and capability of the subsequent mining process. Data preprocessing is an important walk in the knowledge ruling process, since superiority decisions be necessary to be base on advantage data. Data used in preprocessing cover server log files, web page content, web site construction and hit count of pages in the web site. Amalgamation of web usage and web structure mining. Data crackdown removes entries unsupportive to data analyzing and mining. It has to eliminate log entries that have status code as failure or error. Secondly some routine search engines make some access report, those have to be identified and detached from the log file. Some of other frequent indicator such as the frequent request for the similar URL from the similar host. a time period among needs too small to detain the contents of a page and a series of needs from one host every one of whose referrer URLs are empty .The dynamic performance of web users below a exacting session can be some of the next These behaviors can be use to make more multifarious navigation behaviors in a single session. the basic behaviors build complex navigations are specified below: A Web customer can initiate session with several single of the probable entry pages of a web site. This performance contain new page

which is not request by some other previous page access from the similar domain in near-time A Web customer can choose the next page have a link from the nearly everyone recently accessed page. A Web customer can press the back button one additional time and thus select as the next page a page have a link as of some one of the before browse pages. The information composed from the web site can facilitate discover the knowledge. This information composed can be used to obtain decision on various factors like the web pages with maximum hit counts will be the accepted pages. What is probable the navigation patterns of users. The time done in on each web page which tells about consequence of the web page. If time used up on challenging web page is irrelevant it identify that the web page does not enclose significant information. The web pages for which no consumer be relevant for is there, identify that page have to be customized. If log file entry says frequently for exacting web page redirect, it should be notify to web site designer owner. following the pattern study is done on web pages, the significant decision can be done concerning structure of the website. The exceptional web pages will be stimulated extremely near to the home page, at next level medium class web pages moved and so on. The pages with extra hit count can be known the preference to be bring closer to the home page provide web site owner designer agrees. We will heap tree can be generated based on hit counts obtainable in the log file throughout exacting session. This heap tree generate will assist us create decision about topology of web site throughout next interval so that the web pages which are extra popular can be brought extremely near to the home parent web page. With this reformation, the web users can gain rapid access to the web pages down with best consumption of bandwidth and server's memory space as every HTTP request will be go into the log file of the server. Access earlier to the almost everybody recently accessed one. A Web customer can conclude his/her session. Adding, deleting and modify links should be careful at the sometime. it's tricky to find out user's require on web site.

#### *Web efficiency modeling and improvement*

In this analysis, we limit our analysis scope by bearing in mind two considerations for webmaster and users.

- (1) Minimize additional work for web users. to protect the site's original design from destructive changes. limited number of links can be added or removed, and also limited number of pages can be created or destroyed.
- (2) Make the website transformation for everyone, especially first-time users and casual users. customization is very useful for experienced users, but does not benefit first-time users.
- (3) To propose a model who find out the instruct of user.

ADVANCE TECHNIQUE FOR WEBSITE NAVIGATION IMPROVEMENT

Home Filter Normalize Get Navigation Performance

Choose File No file chosen Upload

Log File Name	Row Number	Date	Time	Computer Name	S- SiteName	S- IP
E:\Mahendra\WebNavigation\ForensicAnalysis\MyFiles\Logfile.txt.5	11/5/2012 1/1/2000	12:00:00	12:00:10	MLXILAPP20	W3SVC1716039150	172.16.2.10
E:\Mahendra\WebNavigation\ForensicAnalysis\MyFiles\Logfile.txt.6	11/5/2012 1/1/2000	12:00:00	12:00:10	MLXILAPP20	W3SVC1716039150	172.16.2.10
E:\Mahendra\WebNavigation\ForensicAnalysis\MyFiles\Logfile.txt.7	11/5/2012 1/1/2000	12:00:00	12:00:10	MLXILAPP20	W3SVC1716039150	172.16.2.10
E:\Mahendra\WebNavigation\ForensicAnalysis\MyFiles\Logfile.txt.8	11/5/2012 1/1/2000	12:00:00	12:00:21	MLXILAPP20	W3SVC1716039150	172.16.2.10
E:\Mahendra\WebNavigation\ForensicAnalysis\MyFiles\Logfile.txt.9	11/5/2012 1/1/2000	12:00:00	12:00:21	MLXILAPP20	W3SVC1716039150	172.16.2.10
	11/5/2012 1/1/2000					

Fig.4 Finding web navigation

ADVANCE TECHNIQUE FOR WEBSITE NAVIGATION IMPROVEMENT

Home Filter Normalize Get Navigation Performance

Web Navigation

RowId	url link	Count Link	SessionId
1		0	
2		0	0ydmk45z4qz05ygrholyy
3		0	13c2m59ymxq4bec4ounz
4	-> /WEBPAGES/WWForgotPassword.aspx -> /WEBPAGES/WWPartnerLogin.aspx 5. -> /WEBPAGES/WWPartnerLogin.aspx 6. -> /WEBPAGES/WWPartnerLogin.aspx 7. -> /WEBPAGES/WWPartnerLogin.aspx 8. -> /WEBPAGES/WWPartnerLogin.aspx -> /WEBPAGES/Motor/CustRenewMotorPolicy.aspx -> /WEBPAGES/Motor/CustRenewMotorPolicy.aspx 11. -> /WEBPAGES/WWPartnerLogin.aspx 12. -> /WEBPAGES/WWPartnerLogin.aspx	10	14qm1550w8vrx45ckd4t45
5		0	1cgl0ewsp3r1q45w4e54aa-84027421
6	-> /WEBPAGES/Logout.aspx 17. -> /WEBPAGES/WWPartnerLogin.aspx -> /WEBPAGES/Motor/CustRenewMotorPolicy.aspx 19. -> /WEBPAGES/WWPartnerLogin.aspx	0	1kgfrcuwb4mxcagncsajp-85812843

Fig.5 Web site navigation using clustering

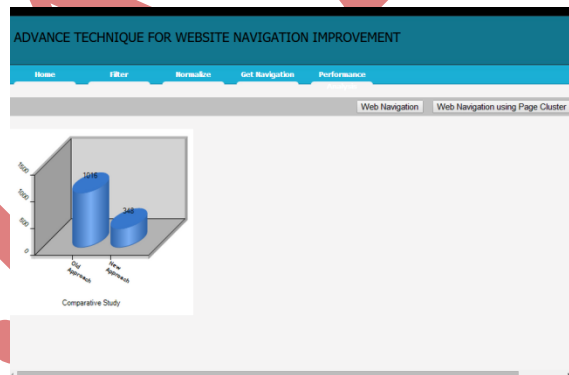


Fig.6 Web navigation using page clustering

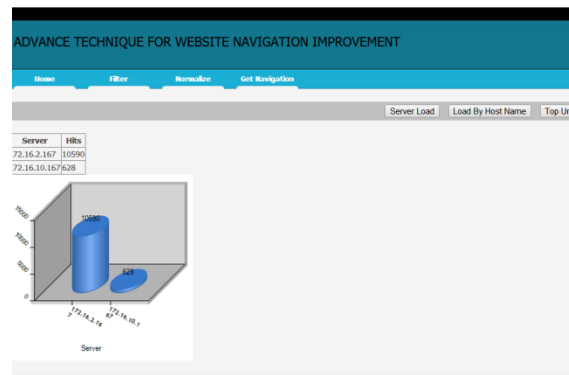


Fig.7 Finding server load and top URL

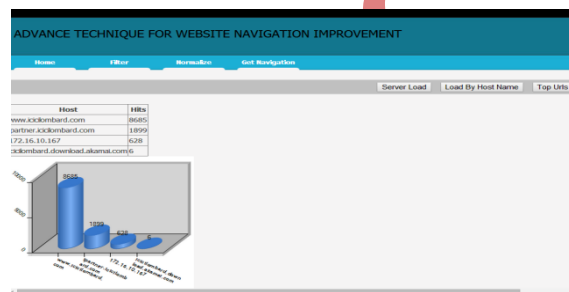


Fig.8 Load on server for top url hits.

**IV. CONCLUSION:** Our aim is to improve Web navigation efficiency by reorganizing Web structure. To avoid users to lose their orientation, structure stability is taken into consideration. This analysis proposes a mathematical programming method to reorganize Web structure in order to achieve better navigation efficiency and also increase our three features which help to improve the user navigation. This analysis has the advantage of improving navigation efficiency mathematically and relieving the designer to modify the structure in transformation. If some constraint, e.g. structural stability constraint, is relaxed, there is still room for improvement. Some improvements are significant but will saturate or even decrease as after adding a certain number of links. To propose a mathematical programming model to improve the navigation effectiveness of a website while minimizing changes to its current structure, a critical issue that has not been examined in the literature.

#### REFERENCES:

- [1] Min Chen and Young U. Ryu, "Facilitating Effective User Navigation through Website Structure Improvement" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 3, MARCH 2013



- [2] Biswas, P , Debnath, M. ; Ghosh, S. ; Rahim, S. ; Mahamud, A.;Chowdhury, N.” Dashboard - a novel approach to re-find information in a website through building personalized navigationnal menus” Computer and Information Technology (ICIT), 2011 IEEE 14th International Conference.
- [3] May Wang , Benjamin Yen ,” Web Structure Reorganization to Improve Web Navigation Efficiency” 11th Pacific-Asia Conference on Information Systems.
- [4] C.C. Lin and L. Tseng, “Website Reorganization Using an Ant Colony System,” Expert Systems with Applications, vol. 37, no. 12, pp. 7598-7605, 2010.
- [5] Jose´ D. Marti´n-Guerrero a,\*, Paulo J.G. Lisboa b, Emilio Soria-Olivas a, Alberto Palomares c, Emili Balaguer c ,” An approach based on the Adaptive Resonance Theory for analysing the viability of recommender systems in a citizen Web porta” Expert Systems with Applications 33 (2007) 743–753.
- [6] C. Kim and K. Shim, “TEXT: Automatic Template Extraction from Heterogeneous Web Pages,” IEEE Trans. Knowledge and Data Eng., vol. 23, no. 4, pp. 612-626, Apr. 2011.
- [7] L. Getoor, Link Mining: A New Data Mining Challenge. SIGKDD Explorations, vol. 4, issue 2, 2003.
- [8] Marcio dos Santos Galli, Eduardo Pezutti Beletato dos Santos,” TagVisor: Extending Web Pages with Interaction Events to Support Presentation in Digital Signage ” WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil. ACM 978-1-4503-2038-2/13/05.
- [9] Jennifer Pearson,” Supporting Effective User Navigation in Digital Documents” CHI 2010, April 10-15, 2010, Atlanta, Georgia, USA. ACM 978-1-60558-930-5/10/04.
- [10] T.L.N.Divya, G.Loshma, Dr. Nagaratna P Hegde,” Template Extraction from Heterogeneous Web Pages Using Text Clustering” International Journal of Computer Trends and Technology- volume3Issue3- 2012.
- [11] C. Kim and K. Shim, “TEXT: Automatic Template Extraction from Heterogeneous Web Pages,” IEEE Trans. Knowledge and Data Eng., vol. 23, no. 4, pp. 612-626, Apr. 2011.
- [12] Y. Yang, Y. Cao, Z. Nie, J. Zhou, and J. Wen, “Closing the Loop in Webpage Understanding,” IEEE Trans. Knowledge and Data Eng., vol. 22, no. 5, pp. 639-650, May 2010.