International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org
Volume 3, Issue 7, July 2014 **ISSN 2319 - 4847**

Facilitating Effective User Navigation by Mining Candidate Links

Minal Gokarn Bade[1], Prof. H.A.Hingoliwala[2]

[1]P.G.Student, JSPM'S JSCOE, Hadapsar, Pune.
[2] Associate Professor, JSPM'S JSCOE, Hadapsar, Pune.

ABSTRACT

It is very important to improve navigation of website since user may lose its interest if navigation is not effective. Websites effectiveness depends on their content and usability needs of their users. Websites are developed by developers, who has own mind set. Websites are developed using requirements rather than providing user friendliness. So designing a website with effective navigation is big challenge. There is different method proposed using user navigation data to re-link web pages. Restructuring whole website may increase the difficulties for the familiar users and it may frustrate them. We propose a mathematical model and algorithm to find out links that are used to improve website structure.

Keywords:- Website structure, web design, user navigation, web mining .

1.INTRODUCTION

Internet are mostly used for information gathering, searching and commercial purpose.There is tremendous increase in number of internet users and it is increased day by day. Online shopping is an example where effective user navigation is very critical and important. Last few years commercial business has grown very much. People are doing online shopping more. So competition for companies doing ecommerce business has increased. The user may go to other website if navigation is not good. Hence companies are interested more to have better navigation of website. Despite that making efficient website is not trivial. The common approach to improve navigation is restructuring website but it is not good approach. User may loose interest because of familiarness caused by restructuring of a website. It is important to increase efficiency of a navigation keeping original structure intact.The reason of a poor navigation of a website is developed by developers understanding and views. Most of the times there is communication gap between developer and actual user. The information which is important for user that is may not be important in views of developer. User get frustrate if the information required is not found in minimum attempt. So website should be able to find required information with minimum navigation. Previous studies have found different issues on websites. In our work we are suggesting solution to improve website navigation.There are different studies have made to improve the efficiency of navigation. Those studies suggest transformation of website. Transformation approach most of time require reorganization of whole website. Reorganization of website results into difficulty for a familiar user, distract user from website, loose business, reduce usability.In the reminder paper, section 2 presents related work and background, section 3 presents mathematical model and proposed algorithm, section 4 explain result and we conclude the paper in section 5

2.RELATED WORK AND BACKGROUND

2.1 Related Work:- Previous studies are divided into web transformation and web personalization. Web personalization means changing the web pages according to particular user or group of users need. Our paper is based on web transformation technique. Wenpu Xing , Ghorbani [1] suggest a weighted PageRank algorithm. This algorithm is the extension of PageRank algorithm. In this paper ranking is given to web pages based on number of time page referred. By using this algorithm, most relevant pages in large number are returned for a search query. Perkowitz and Etzioni [2] present idea of web personalization depending upon occurrence of frequency of pages in user traversal. This paper investigates adaptive websites that automatically improve their organizational presentation by learning from visitor's access patterns. Nakagawa and Mobasher [3] proposed an approach in which degree of connectivity is calculated, then find out location of user in the site. Then based on these two factors recommendation model is dynamically switched. Lin [4] suggested an approach in which cohesion between the web pages are used. Cohesion between pages means relation between pages. If cohesion between pages are more, than they are directly linked. If cohesion between webpage is less, than they are indirectly connected. W. Yan , M. Jacobsen, H. Garcia-Molina and U. Dayal [5] suggest an approach in which users are classified based on pages that user visits. Using web pages log, group of users are determined who visits similar web pages. This data is used to improve navigation convenience. According to users group they recommend links dynamically. Y. Fu, M.Y. Shih, M. Creado and C. Ju [6] investigate adaptive web site that automatically improve organizational presentation by learning from visitor's access pattern. According to access information web pages are

classified as index page and content page. Using that classification data, web site is examined to reorganize web pages in better way. Bamshad Mobasher1, Honghua Dai, Tao Luo, Miki Nakagawa [7] they have suggested two techniques based on clustering of user transactions and clustering of page views, to generate collective profiles that can be used by recommendation system to improve web personalization B. Mobasher , R.Cooley, J. Srivastava [8] propose an approach where full spectrum of web mining techniques and activities are used for web personalization. They propose automatic and dynamic web personalization. The user web pages are grouped and used as a URL reference. In this paper .they have combine transaction clustering, usage clustering and association rule for web personalization. B.Mobesher, R. Cooley and J.srivastava [9] have proposed an approach where they gather information about users offline task and mining data to improve web personalization. They have proposed a technique to gather user profiles using association rule discovery and uses based clustering. C.C.Lin and L.Tseng[10] have proposed an approach to minimize efficiency problem of 0-1 programming model. In this paper they have proposed an ant colony system to understand website structures. R.Gupta, A. Bagchi and S.Sarkar[11] have propose schema using simulated annealing which internally uses data of user preferences to align the pages to improve navigation. This is applicable to wired and wireless devices.

2.2Background:- For effective website navigation we consider two points, first one is information becomes useful only when it is presented in a way consistent with the target users expectations and second one is user get desired data without loss or backtrack.We can say that the website is effective when user find desired page in minimum attempts and without backtracking. Some of the concepts in this paper as below: Session[12]- Number of activities performed from the point he login to point he logout. During one session user gets navigated through several different pages. So, in one session there can be one or more target pages. Target page is a page where user spends more time than threshold time. Group of pages visited by user to reach a single target page is called minisession [12].One way to improve navigation is add extra link for a target page on each page. Such links are called candidate links[12] For example user is at the page 1 then user browses page 4 and then backtracks to page 1, from 1 user visit page 3, again backtrack to 4, then browse 6 and finally reach target page. So, we denote minisession of S = {{1,4},{3},{6,8}}.Now we improve the navigation by adding extra links. So when we add link from 1 to 8, the user can directly reach the target in the first path. Similarly if we add link from 4 to 8, user directly reach target page. Hence this saves users two paths. If we add link from 3 to 8, user reaches the target in second path. So, save users one path. This added relevant links is called candidate links. Table 1 shows the example of Mini Session and Table 2 shows the respective candidate links after adding extra links Adding candidate links on webpage make user to reach to target page faster. But adding to many candidate links on a page may confuse a user. Web page may get loaded with unwanted links and information. And answer will not possible within polynomial time. So our paper focuses on to find out only required candidate links. So by adding only required candidate links we can improve website structure without causing information overload.

Table 1 : Mini session

ID	Mini Session
US1	{{P,Q,R},{S},{T,W}}
US2	{{R,T,S},{Q,U}
US3	{{S,R,Q},{U},{V,T,W}}
US4	{P,R}{Q,U},{T}

Table 2 : Candiadate Link Set

ID	Candidate Links
US1	{{P,W},{Q,W},{R,W},{S,W}}
US2	{{R,U},{T,U},{S,U}
US3	{S,W},{R,W},{Q,W}{U,W}}
US4	{{P,T},{R,T},{Q,T},{U,T}}

3.PROPOSED SYSTEM

In candidate link set first we find out the existent links. Sometimes it is possible that existent links are found in candidate link set. This is because of location of link or visibility of links is not properly placed. So, first we need to improve those links. Next, we need to find out relevant candidate link set. Relevant candidate links means links having larger path than path threshold. For filtering candidate links first use path threshold. Path threshold means maximum number of path allows reaching target page. Path threshold 1 means we consider only those link by which user reach target page in first path. Then we select those links. We cannot add all these links. We can add only those candidate links that are common for all users. For these we need to find out similarity between links. For finding similarity between links we use Dice's coefficient index. We apply index on relevant candidate link set. We obtain the dice's index between 0 and 1. 0 means dissimilar and 1 means exact similar. Then we use KNN classifier. KNN means k nearest neighbour classifier. KNN stores the entire link set and based on similarity measure classify candidate links into two groups. In KNN we use one threshold value to classify candidate links into two groups. One group contains candidate links which are present in more

number of users than threshold value. Other group contains candidate links which are present in less number of users than threshold value. We filter second group. And we use the candidate links in first group means the candidate links present in user more than path threshold to re-link or redesign website structure. In other words, we say that if we add that link then problem for more users to locate target page will be solved.

4.PROGRAMMERS DESIGN

4.1 Mathematical Model:- In this section we will discuss mathematical model of proposed system: Website is represented as a directed graph. Let nodes representing pages and arc representing links. Let n be the number of pages of the website. User may visit more than one page by accessing links during user session. Let Pi={P1,P2,…..,Pm} be the set of m link accessed by user Ui where i={1,2,….,m}. Let Pm be the page user Ui is looking for. Means Pm is the target page. Let U= {U1, U2 ,…,Un} be the set of N users. Let C= {C1, C2 ,…., Cp} be set of p candidate links that need to redesign and relink. We use Dice's coefficient index. Dice's coefficient index is used for checking similarity between two string set.

$$D(A, B) = \frac{2*|A \cap B|}{|A|+|B|}$$

The aim of this paper is to identify links that are used to redesign. Usage pattern is used to analyze user's behavior on the Web. By analyzing user's behaviour we find out target page. And find out candidate links so that user can access target page faster.

4.2 Algorithm:-

Mining Candidate Link Algorithm
Input: Pi – Users Profile data
Output: Links that can be use for redesign
Steps-
1: We identify the usage pattern of users λ from Pi = {P1, P2 ,…., Pm} set for user Ui to get link Pm
2: For every access link set obtain the set of candidate links {C1, C2,….,Cp}
3: For all users and their all access link set obtain the set of candidate links.
4: Obtain the Dice's similarity coefficient for all candidate link set.
5: Apply KNN classifier.
6: Then the links having problem for maximum number of users are selected for redesign the website.

5.RESULT

The experiment was applied on real data set. Real data set was collected from one music machine website. The following snapshot shows that which links needs to be added to improve website structure. We can divide the new links into three tables. First table contains links when no threshold value. Second table contains links having one threshold value and third table contains candidate links having threshold value equal to two. And also snapshot of output contains existing links which are necessary to improve for improve website structure.

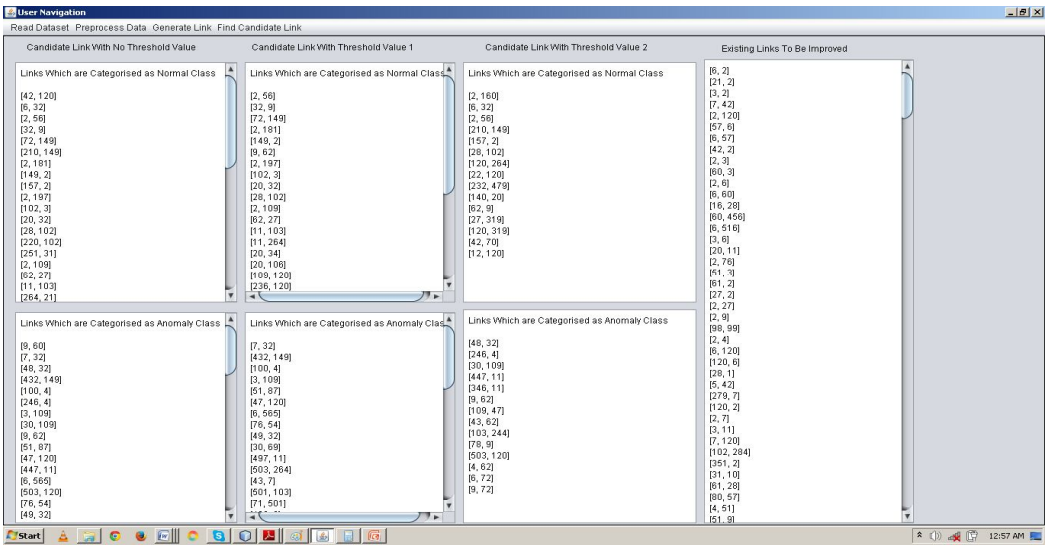

Fig. 1 : Result Snapshot

In table 3, time threshold means user stay on the page more than or equal to time threshold then we consider that page as a target page. 'No. of new links mostly accessed by user' means the number of links which are need to most users to improve user navigation. And that links are used to improve the website structure. 'Number of new links accessed by very few users' is the links which are required for few users to improve user navigation. So we filter out these links. 'Number of links to be improved' means the existing links which are need to improve.

Table 3 : Result

Time Threshold(t)	Path Threshold	No. of New Links (Mostly Accessed By user)	No. of new Links (Accessed by very few User)	No. of Links to be improved	Time
1 min	1	5469	4391	2453	0.114
	2	1256	902	750	0.538
	3	372	278	467	0.427
2 min.	1	5216	3708	2014	0.159
	2	946	837	812	0.659
	3	382	305	481	0.497
5 min	1	4962	2727	2089	0.156
	2	843	703	751	0.567
	3	376	314	468	0.504

6.CONCLUSION

In this paper we propose a mathematical model to improve website structure for effective user navigation by minimum changes to its current structure. We propose an algorithm which gives set of links that can be re-link to improve website structure. This approach is a better solution than whole redesigning of a website. This algorithm is best suited for website with mostly static content such as educational website. This model can be extended to improve web structure of dynamic website

7.ACKNOWLEDGEMENT

The satisfaction achieves after successful completion of any task would be incomplete without mentioning those people who are responsible to complete that task. We are grateful to many people who contributed to the completion of this research. We would like to express our sincere thanks to Prof. M.D.Ingle, Prof. S.M.Shinde and our principal Dr. M.G.Jadhav to providing necessary facilities for completion of paper work..

REFERENCES

[1] Wenpu Xing, Ghorbani , Weighted page rank algorithm, A communication networks and services Research,Proceeding. Second Annual Conference 2004 .
[2] M. Perkowitz and O.Etzioni, Towards adaptive web Sites Conceptual Framework and Case Study, Artificial Intelligence, vol. 118, pp. 245-275,2000.
[3] Nakagawa and Mobasher, A hybrid web personalization Model Based on Site Connectivity, Proc.Web Knowledge Discovery Data Mining Workshop, pp59-70, 2003.
[4] C.C.Lin, Optimal Web Site Recognization Considering Information Overload and Reserch Depth ,European J. Operational Research, Vol.173,no.3,pp 839-848 , 2006.
[5] W.Yan , M. Jacobsen, H.Garcia-Molina and U. Dayal, From User Access Patterns to Dynamic Hypertext Linking, Computer Network and ISDN Systems, vol.28,nos 7-11,pp.1007-1014, 1996.
[6] Y.Fu, M.Y. Shih, M. Creado and C.Ju, (2002) Reorganizing web Sites Based on User Access Patterns, Intelligent Systems in Accounting, Finance and Management, vol.11 , no.1 ,pp.39-53.
[7] B.Mobasher, H.Dai, T.Luo ,and M.Nakagawa, (2002) ,Discovery and Evaluation of Aggregate Usege Profiles for Web Personalization, Data Mining and Knowledge Discovery,vol.6,no.1,pp.61-82.
[8] B.Mobasher, R.Cooley, and J. Srivastava, (2000), Automatic Personalization Based on Web Usage Mining, Comm .ACM ,vol 43,no. 8,pp 142-151.
[9] B.Mobesher, R. Cooley and J.srivastava, (1999) Creating Adaptive web Sites through Usage-Based clustering of URL ,Proc. Workshop Knowledge and Data Eng. Exchange.

[10] C.C.Lin and L.Tseng, (2010) ,Website recognization Using an Ant Colony System, Expert Systems with Applications, vol.37 , no. 12, pp. 7598-7605.

[11] R.Gupta, A. Bagchi and S.Sarkar ,(2007),Improving Linkage of Web pages, INFORMS J.Computing, vol. 19, no.1 ,pp. 127-136.

[12] Min chen, Young U. Ryu (2013), Facilitating Effective UserNavigation Through Web Structure Improvement, IEEETrans. Knowledge and Data Eng., vol. 25 no.3