Optimized Website Structure Improvement for Effective User Navigation

Mr. Samadhan W. Jadhav, Prof. N. L. Bhale

PG Student, Computer Engineering Department, MCERC Nasik, Maharashtra, India
Associate Professor, Computer Engineering Department, MCERC Nasik, Maharashtra, India

---***---

***Abstract*–There is a challenge from long time of designing a website which will facilitate the effective User's navigation. As the users of a website face the difficulties to navigate to their target pages easily, developers decide to reorganize the website. The website needs to be reorganized substantially to navigate the every user efficiently. The completely new structure can be very random and increase the cost of positioning users after the changes. The modifications to the website structure create another challenge that, it changes the locations of familiar items for the existing users. Adding new links to facilitate effective user navigation also disturbs the business logic. In this paper we have proposed a system which is dealing with the above challenges. Specifically, it uses a mathematical programming model with optimized out-degree threshold and providing suggestions to users to reach their target pages. This has optimized changes in the current website structure while improving the users' navigation.**

Key Words: Mathematical Programming, Out-degree, User Navigation, Website design, Web Mining

1. INTRODUCTION

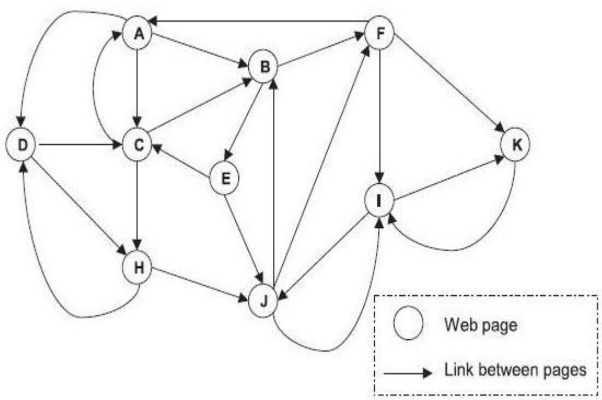
There were 1.73 billion Internet users worldwide as of September 2009, which shows 18percent increase since 2008 [1]. The increased number of Internet users also presents huge business opportunities to firms. The huge increase in the web content, both website users and managers expect high quality content service, while providing users with the information they need conveniently [2]. The link structure of a relatively static website would have great influence to serve quality of contents. Therefore how to test and enhance the website structure becomes a crucial issue. There are many techniques in past such as analyze past user access patterns to discover typical user access patterns, is useful to enhance the link structure of static websites or dynamically insert links to web pages. A method used to redesign a website by identifying user profiles from web usage data. Many techniques have been used to discover the gap between expectations of website designers and users' behavior and suggested areas where website can be improved [4]. The proposed method have used a Mathematical Programming Model, which suggests improvements by suggesting links to add or enhance to improve website structure while minimizing the changes to current website structure. Palmer [3] had shown that if users are able to access desired data without getting lost or need to backtrack, then the website is easy for navigation.

Here backtracks identifies the paths traversed by user to reach the target page. Backtrack is therefore defined as a user's revisit to a page. The perception is that users will backtrack from the point, where the expected page is not found. Thus, a path is a set of pages visited by a user sequentially, without backtracking. This concept is similar to the concept of maximal forward reference described in Chen et al. [5] essentially; each backtracking point is the end of a path. Figure 1 shows backtracking can be used to find the Mini-sessions and Target Pages, user is searching for. As the point of backtrack indicates selection of new path to reach the desired page, one Mini-session can consist of many paths traversed to see a target page. That is one session of user is a collection of number of Mini-sessions for their corresponding target pages. Consider Target Page: K
Mini-sessions are {(A, D, H), (D, C, B, E, J), (B, F, K)}
Backtrack Points: H and J

2. LITERATURE SURVEY

There has been a tremendous research performed in the past to improve user navigation. Some of these techniques have improved user search up to some extent but still there is a huge scope in the future. The techniques discovered previously based upon restructuring of the website using user's usage patterns, user access patterns, page classification, site reorganization etc [7]. These techniques have played an important role to help facilitating effective user navigation but still there are some challenges

Such as…



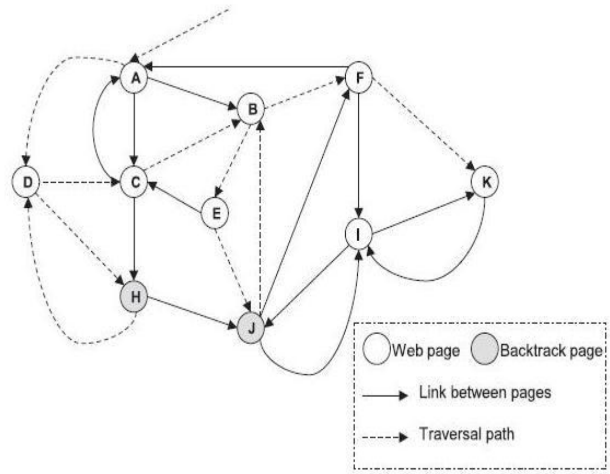


Figure1: A website with 10 pages.

_ User access pattern are not considered
_ Out-degree threshold is not optimized
_ User target identification accuracy is low
_ User interest levels are not considered

The solutions drawn for any of the above challenge will definitely improve the results of system for effective user navigation on a website with optimizing the changes in current structure.

2.1 Mathematical Programming model

The Mathematical Programming model proposed by Min Chen and Young U. Ryu, [6][8][9][10] used here is consists of three steps and are described as follows:

1. Apply the MP model on the training data (website structure) to obtain the set of new links and links to be improved.

2. Acquire from the testing data(log files) the mini sessions that can be improved, i.e. two or more paths, path's length, i.e., number of paths, and the set of candidate links which can be used to improve them.

3. For each mini session, check whether any candidate link matches one of the links, that is, the results from the training data. If yes, remove all pages visited after the source node of the session for the improved website.

2.2 Existing System

Figure 2.1 shows the block diagram of existing system [3]. It uses web log files to obtain user search data. This data is analyzed to obtain User Sessions, which are further processed to find the Target pages and Mini-sessions. The Target pages are the end of users search, whereas the number of paths traversed to reach that target page forms a Mini-session. The length of a Mini - session, if not falls within given Path Threshold then such Mini-sessions are the relevant Mini-sessions for further processing to find Candidate links and relevant Candidate Links. The found Relevant Candidate links are analyzed using an Objective function, which verifies them against the out-degree threshold.

The links which qualifies through the objective function are suggested to add or improve by the web developer to improve given Mini-sessions. The link considered to add if exists in website, then it is simply enhanced or improved, otherwise web developer have to update web link structure, causing frequent maintenance of the website.

2.3 Proposed System

The Proposed system provides alternative paths which user may haven't traversed during his navigation. Such paths avoid addition of new links which further avoids frequent maintenance of the website. It also provides paths as per the specified path threshold. This prevents changes in website structure by adding new links and provides relief to developer from frequent maintenance.

2.4 Problem Definition:

”To Optimize the Website Structure Improvement for the Effective User Navigation of a Website by using a Mathematical Programming Model with Optimized Out-degree threshold which permits optimal changes to website's current structure and prevents reorganizing it substantially.”

3. MATHEMATICAL MODELLING

This phase describes the designing and modeling phase of the website structure improvement process by optimizing alterations to its current structure and Out-degree threshold.

3.1 Model:

Let ’S’ be a Website Structure Improvement system for effective user navigation, such that,

S = {L, P, M, S, T, I, O, C, W, E}

Where,
L = {l0, l1, ..., ln}; is a set of links connecting web pages of a website.
P = {p0, p1 ,..., pn}; is a set of paths traversed by user while reaching a target page.
M = {m0, m1, ..., mn}; is a set of mini sessions of an user.
S = {s0, s1,..., sn}; is a set of sessions of many user logins.
T = {t0, t1, ..., tn}; is a set of Target pages.
I = {i0, i1, ..., in}; is a set of In-degrees of Web pages.
O = {o0, o1, ..., on}; is a set of Out-degrees of web pages.
C = {c0, c1, ..., cn}; is a set of candidate links.
W = {w0, w1, ..., wn}; is a set of web pages of a website.
E = {e0, e1, ..., en}; is a set of links to be added/improved in site restructuring process.

Let Ls be the link suggested by System S to improve and if Ls ∈ C for target page t ∈ W and Ot less than Out − degree Threshold, then Ls ∈ E.

3.2 Functional Dependency

The sequential algorithm will execute function one time to get the required result, which may contain 9 commands or relevant algorithms as per the efficiency issue given below.
F0: get_Data()- accepts data set to process.
F1: find_Session()- returns sessions from the given data set.
F2: find_Minisession- identifies mini sessions from each session.
F3: find_Paths()- returns paths traversed in each mini session.
F4: find_Target()- identifies and returns the target pages from each Session and mini session.
F5: find_Degree()- returns In-degree and Out-degrees of web pages.
F6: find_Clinks()- returns candidate links for the given Target page.
F7: obj_Fun()- returns the flag whether new link will be added to site structure.
F8: put_Data()- displays results to show how the restructuring will improve user navigation.

Functional dependency of above functions is as shown in Table 1.

Table 1: Functional dependency

	F0	F1	F2	F3	F4	F5	F6	F7	F8
F0	1	0	0	0	0	0	0	0	0
F1	1	1	0	0	0	0	0	0	0
F2	0	1	1	0	0	0	0	0	0
F3	0	0	1	1	0	0	0	0	0
F4	0	0	1	1	1	0	0	0	0
F5	0	0	0	0	1	1	0	0	0
F6	0	0	1	1	1	0	1	0	0
F7	0	0	0	0	0	1	1	1	0
F8	0	1	0	0	0	0	0	1	0

4. ALGORITHM

Input: Web log files and Website structure.

Steps:

1. Accept the Web log files i.e. testing data as input.
2. Process the Web log files to identify target pages, paths, path lengths, out-degrees, nodes etc.
3. Find out available links or training data of website current structure by traversing the complete website.
4. Process the training data (website structure) to obtain the set of links to be used to reach a target page.
5. Suggest existing or alternative links if available, otherwise suggest new links to add within given Path Threshold, which should also satisfy the set Out-degree thresholds of nodes or pages which are considered to be connected using new links.
6. Each mini session is checked for whether any candidate link matches one of the existing link, that is, the results from the training data. If yes, the existing link is improved or alternative links are suggested to facilitate users' navigation.

5. RESULTS

"Dataset" is the numeric representation of parsed links of a website as shown below. It has format of
Node (outgoing links), pages reachable via that node.

5.1 Dataset representing website:

Total Nodes/Pages: 19
Node (outgoing links) Pages reachable
1 (1) 15
2 (1) 15
3 (1) 15
4 (1) 15
5 (5) 15 9 10 11 5
6 (5) 15 9 10 11 5
7 (4) 15 9 10 11
8 (1) 15
9 (5) 15 9 10 11 5
10 (5) 15 9 10 11 5
11 (4) 15 9 10 11
12 (4) 15 9 10 11
13 (4) 15 9 10 11
14 (4) 15 9 10 11
15 (5) 15 9 10 11 5
16 (1) 15
17 (1) 15
18 (1) 15

5.2 Weblog file

(1, 15), (15.9), (9, 11), (11, 10)
(3, 15), (15, 5), (5, 10)
(2, 15), (15, 10)

Consider Path Threshold = 2

Therefore relevant Mini-sessions are: Relevant Mini-session 1: (1, 15), (15. 9), (9, 11), (11, 10)
Relevant Mini-session 2: (3, 15), (15, 5),(5, 10)

Candidate Links for: Target Page: 10

For Mini-session1: (15, 10), (9, 10), (11, 10)
For Mini-session2: (15, 10), (5, 10)

Relevant Candidate Links: (Path threshold 2)
For Mini-session1: (15, 10), (9, 10)
For Mini-session2: (15, 10), (5, 10)

Relevant Candidate Links: (Path threshold 1)
For Mini-session1: (15, 10)
For Mini-session2: (15, 10)

Now decide Path threshold and links to be added or improve.
For Path threshold 2 Links (15, 10), (9, 10) are only enhanced and not newly added to improve Mini-session1 as they are matched with available links i.e. they are existing links.

Similarly, Links (15, 10), (5, 10) are enhanced and not newly added to improve Mini-session1 as they are matched with available links i.e. they are existing links. Hence optimizing the changes to current website structure by avoiding the addition of new links.

For the same target pages alternate available paths can be suggested as per the set Path Threshold.
For Example; consider target page 10 and the all available alternate paths to page 10 are as given in the table 2.

Table 2: Available paths to target page as per the set Path threshold.

Target Page	Path Threshold – 1	Path Threshold – 2	Path Threshold - 3
10	{5,10}	{5,9}, {9,10}	{5,9}, {9, 15}, {15, 10}
	{6,10}	{5,11}, {11, 10}	{5,9}, {9, 9}, {9, 10}
	{7,10}	{6, 9}, {9, 10}	{5,9}, {9, 10}, {10, 10}
	{9,10}	{6, 5}, {5, 10}	{5,9}, {9, 11}, {11, 10}
	{10,10}	{7, 9}, {9, 10}	{5,9}, {9, 5}, {5, 10}
	{11,10}	{7,11}, {11, 10}	{5,11}, {11, 15}, {15, 10}
	{12,10}	{9,11}, {11, 10}	{6,9}, {9, 11}, {11, 10}
	{13,10}	{11, 9}, {9, 10}	{7,11}, {11,9}, {9, 10}
	{14,10}	{12,11}, {11, 10}	{12,9},{9, 11}, {11, 10}
	{15,10}	{13, 9}, {9, 10}, and {14,11}, {11,10}	And many other available……

Along with this the addition of new link is if mandatory to improve the user navigations, then that link is added from a page/node to target page by checking it's new out-degree against the set Out-degree threshold.

The current out-degrees are already computed while generating dataset. Now figure below shows
the set out-degree threshold and nodes under that
Threshold to choose from while adding new links.

Table 3: Nodes having Out-degree <= Ot (Set Out-degree Threshold)

Sr. No.	Set Out-degree Threshold	Number of nodes < Ot	Nodes
1.	1	8	1, 2, 3, 4, 8, 16, 17, 18
2.	2	8	1, 2, 3, 4, 8, 16, 17, 18
3.	3	8	1, 2, 3, 4, 8, 16, 17, 18
4.	4	13	1, 2, 3, 4, 7, 8, 11, 12, 13, 14, 16, 17, 18

It is possible here to set lowest Out-degree threshold, because rarely there is a need to add new links and for that as well we have computed many nodes/pages to choose from while adding new links. **Hence optimizing the Out-degree threshold.**

6. CONCLUSION

The proposed system reveals the way to improve a website's structure for effective user navigation, while optimizing the cost of users' disorientation due to substantial changes in the current structure, a challenge in literature. This project implements a Mathematical Programming Model (MPP) by optimizing Out-degree threshold. The results shows that the modules of system are used to analyze the current website structure by displaying the available links as well as to determine the direct links to target pages, which will be further used to recommend the links to be
Added/enhanced as per the path threshold of mini sessions and optimized Out-degree threshold of web pages respectively. The implemented software also displays paths to target pages as per the set path threshold.

Future Work:

This work is a motivation for several researchers. Basically in this paper we have considered a website to be improved for effective user navigations. The system can be extended by integrating the system with Personalization approaches.

REFERENCES

1. Pingdom, Internet 2009 in Numbers, [Online] Available at: http://royal.pingdom.com/2010/01/22/internet-2009-innumbers/, 2010.
2. Y. D. Dhyani, W.K. Ng, S.S. Bhowmick, A Survey of Web Metrics, ACM Computing Surveys, Vol. 34, No. 4, pp. 469-503, 2002.
3. J. Palmer, Web Site Usability, Design, and Performance Metrics, Information Systems Research, Vol. 13, No. 2, pp. 151-167, 2002.
4. V. McKinney, K. Yoon, and F. Zahedi, The Measurement of Web Customer Satisfaction: An Expectation and Disconfirmation Approach, Information Systems Research, vol. 13, no. 3, pp. 296 315, 2002.
5. M.S. Chen, J.S. Park, and P.S. Yu, Efficient Data Mining for Path Traversal Patterns, IEEE Trans. Knowledge and Data Eng., vol. 10, no. 2, pp. 209-221, Mar./Apr. 1998.
6. Facilitating Effective User Navigation through Website Structure Improvement Journal of Advances in Computer, Electrical Electronics Engg., Volume 3 , Issue 1; Spl. Issue of IC3T 2014 @ISSN: 2248-9584.
7. Optimizing User Navigation with Pattern based Web Site Restructuring Scheme, International Journal of Innovative Research in Computer and Communication Engineering, Vol.2, Special Issue 1, March 2014.
8. Effective User Navigation through Website Structure Improvement, IJCST Vol. 4, Issue Spl-4, Oct - Dec 2013.
9. IJCTT Effective User Navigability through Website Structure Reorganizing Using Mathematical Programming Model, International Journal of Computer Trends and Technology (IJCTT) - volume 10 number 3 -Apr 2014.
10. 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.