International Journal of Computer Science
and Engineering (IJCSE)
ISSN(P): 2278-9960; ISSN(E): 2278-9979
Vol. 3, Issue 5, Sep 2014, 37-42
© IASET

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

IMPROVING USER NAVIGATION THROUGH WEBSITE LINK STRUCTURE IMPROVEMENT

VINAY K S[1], KUMARSWAMY H[2], MAHESH M[3] & K M NIRANJAN[4]

[1]M.Tech (CSE) 4th Semester, S. J. M. I. T, CHITRADUGA, Karnataka, India

[2]Associate Professor, Department HOD, IS & E, S. J M. I. T, CHITRADURGA, Karnataka, India

[3]Assistant Professor, CS & E, S. J. M. I. T, CHITRADURGA, Karnataka, India

[4]Associate Professor, Department of Mathematics, U. B .D. T, DAVANAGERE, Karnataka, India

ABSTRACT

Locating a useful information in website has been a problem and designing well-structured websites to facilitate effective user navigation has long been a big challenge. This paper provides suggestions to improve user navigation by performing minimum changes to website link structures. Since the users view could be disoriented due to complete restructure and radical changes in the website. The proposed approach consists of two steps viz. preprocessing and website reorganization, which is based on mathematical programming model. User access pattern and website link structure were considered to enhance the link structure which resulted effective user navigation in a website.

KEYWORDS: Website Design, User Navigation, Web Mining and Mathematical Programming

INTRODUCTION

The free-for-all Internet has led towards a massive publication of information. The Web nowadays seems to serve as a worldwide free library. This jumbled nature of the WWW is mirrored in the structure of websites, making it essentially messy. Most Web structures are large and complicated and users often miss the goal of their target, or receive ambiguous results when they try to navigate through them.

A primary cause of poor website design is that the web developer's understanding of how a website should be structured can be considerably different from those of the users. Such differences result in cases where users cannot easily locate the desired information in a website. This problem is difficult to avoid because when creating a website, web developers may not have a clear understanding of user's preferences and can only organize pages based on their own judgments. However, the measure of website effectiveness should be the satisfaction of the users rather than that of the developers. Thus, Web pages should be organized in a way that generally matches the user's model of how pages should be organized. There are two approaches to reorganize the

Website structure one is personalization and the other is transformation. In personalization the structure is changed according to individual's interest. But the transformation approach tends to make change for a general group of users.

This paper is concerned with transformation approaches. Transformation approach referred as to modify the site structure to ease the navigation for all users. The literature considering transformations approaches mainly focuses on developing methods to completely reorganize the link structure of a website.

Our model is suitable for those websites whose contents are not changing as per user requirement & data within the websites are static. For examples companies, hospitals, universities, banks, tourist websites. The presented model may not be appropriate for websites that contents dynamic pages and volatile contents. This is because in dynamic web pages access patterns changes frequently, therefore the weblog data is not used to improve the web site structure. In this we perform various tests on a data set collected from a real website. The result proves that our model can improve the site structure with minimum changes. In this paper, we explore the problem of improving user navigation on a website with minimal changes to the current structure then we calculate the out-degree as an objective function instead of fixed constraints. This allows a page to have more links than the out-degree threshold if the cost is reasonable and hence offers a good balance between minimizing changes to a website and reducing information overload to users.

RELATED WORK

Web transformation, involves changing the structure of a website to facilitate the navigation for a large set of users instead of personalizing pages for individual users. [1] Describe an approach to reorganize web pages so as to provide users with their desired information in fewer clicks. However, this approach considers only local structures in a website rather than the site as a whole, so the new structure may not be necessarily optimal. [2] Propose a heuristic method based on simulated annealing to re-link web pages to improve navigability. This method makes use of the aggregate user preference data and can be used to improve the link structure in websites for both wired and wireless devices. However, this approach does not yield optimal solutions and takes relatively a long time (10 to 15 hours) to run even fora small website. [3] Develops integer programming models to reorganize a website based on the cohesion between pages to reduce information overload and search depth for users. In addition, a two-stage heuristic involving two integer programming models is developed to reduce the computation time. However, this heuristic still requires very long computation times to solve for the optimal solution, especially when the website contains many links. Besides, the models were tested on randomly generated websites only, so its applicability on real websites remains questionable. To resolve the efficiency problem in [4] propose an ant colony system to reorganize website structures. Although their approach is shown to provide solutions in a relatively short computation time, the sizes of the synthetic websites and real website tested in are still relatively small, posing questions on its scalability to large-sized websites.

In this paper, a Web Enhancement Model is proposed to improve user traversal path in a website with making fewer changes to its structure. The presented model is used for enhancing the web site structure and can be applied for small as well as large websites. The presented model is scalable according to the requirement of web developers and it is well suited for website maintenance.

PROPOSED SYSTEM

The system first extracts website link structure from the website created by the author and generate the Weighted Directed Graph Model for the website. It then extracts user access pattern information from user web logs to describe user preference and navigation convenience. Based on user access pattern extracted and website link structure, the system can evaluate website link structure and give suggestions for improvement using the mathematical programing model. Authors will make adaptation according to the system’s suggestions to the website structure for better user’s access of the web pages in the site.

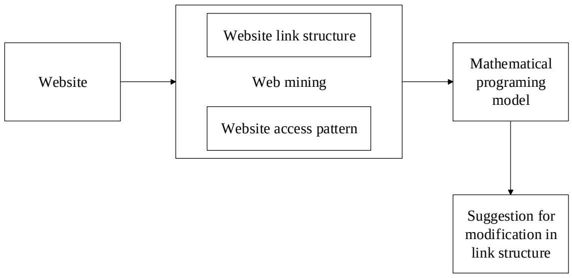

Figure 1: Block Diagram of the System

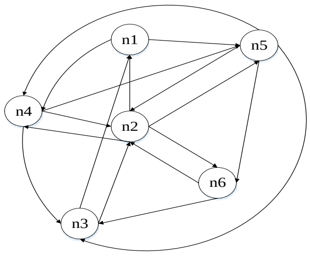

Figure 2: The Directed Graph of a Website

First website structure is converted into directed graph as in above figure 1 and this graph can be solved to get the optimum solution. We convert our website as graph optimization problem in which nodes represents web pages and the line between two nodes represents Links which connects two different web pages.

$$\begin{array}{c} \\ n1 \\ n2 \\ n3 \\ n4 \\ n5 \\ n6 \end{array} \begin{array}{c} \begin{array}{cccccc} n1 & n2 & n3 & n4 & n5 & n6 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

Figure 3: The Connectivity Matrix

A matrix is generated for the graph which gives the information of connectivity of a website. Value 1 indicates the connection between the row node and the column node and the summation of the value in a row gives the out-degree of a node.

Second to analyze the interaction between users and a website, the log files must be broken up into user sessions. A session may include one or more target pages, as a user may visit several targets during a single session. Since the metric used in our analysis is the number of paths traversed to find one target, the term mini session is used to refer to a group of pages visited by a user for only one target. Thus, a session may contain one or more mini sessions, each of which comprises a set of paths traversed to reach the target. The page-stay timeout heuristic is used to demarcate mini sessions. Specifically, to identify whether a page is the target page by evaluating the time spent on that page is greater than a timeout threshold. Then the proposed model inputs preprocessed data from web log files and web site structure, and output produced is a modified website structure. The presented model includes matrix representation of a graph, user's session, mini session. Selection of candidate link, relevant candidate link determined and suggestion for modification in website structure is given.

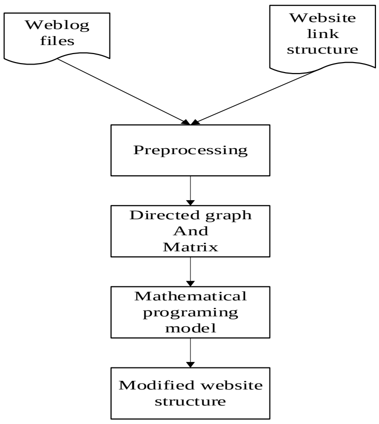

Figure 4: Detailed Architecture of System

The presented model improve the web site structure without modifying whole structure, it only reorganizes the web pages such way that user get targeted page with very few clicks. Modified web structure provides better user navigation on a website.

MATHEMATICAL PROGRAMNG MODEL

The problem of improving the user navigation on a website while minimizing the changes to its current structure can then be formulated as the mathematical programming model below:

$$\sum_{(i,j)\in E} x_{i.j}\left[1-\boldsymbol{\lambda}_{i,j}(1-\varepsilon)\right]+m\sum_{i\in N_E} p_i$$

The objective function minimizes the cost needed to improve the website structure, where the cost consists of two components: 1) the number of new links to be established (the first summation), and 2) the penalties on pages containing excessive links, i.e., more links than the out-degree threshold (C_i), in the improved structure (the second summation).

$$C_{kr}^{S}=\sum_{(i,j)\epsilon E} a_{ijkr}^{S}.x_{ij}\ ; r=1,2,\ldots\ldots,L_p(k,S)$$

k = 1, 2, 3... Lm(s),$\forall S\in T^R$

The above equation is the first constrain its value is 1if in mini session S, a link from r^{th} page in k^{th} path to the target is selected; 0 otherwise. C_{kr}^{S} is a variable which will be set to 1 only if a_{ijkr}^{S}=1and x_{ij}=1,for some i,j ∈ E.

xij - represents 1 if the link from page i to j selected; otherwise 0.

m - Represents multipliers for the penalty term into the objective function.

pi - represents number links that exceeds the out degree threshold.

λij - represents 1 if the i has a link to page j in the current structure; otherwise 0.

Ɛ - is a very small number, in the objective function to let the model select existing links.

N_E – the set of source nodes of links in set E.

a_{ijkr}^{S}– 1 if *i* is the r^{th} page in k^{th} path and *j* is the target page in mini session S; 0 otherwise.

$L_p(k,S)$– Length of k^{th} path in S mini session.

DISCUSSIONS

1. **Mini Session:** A user in a session traverses the website and locates the many target pages. To access the traversal paths used to reach the individual targets system divides the session into a mini session for each targets. The traversal paths are determined as when the user backtracks from a page and returns to the previous page, it indicates the target is not fond in that path hence the path ends.

2. **Target Identification:** The target page is identified by the term page-stay timeout heuristic. The intuition is that the user stays longer time in target page than the other pages because the user finds useful information in that page.

3. **Path Threshold:** The goal for user navigation that the improved structure should meet. This is the maximum number that the system sets for each target page, if the user fails to reach the target page within this threshold value it means that the link structure to that page should be improved.

4. **Pattern Analysis:** The aim of this analysis is to obtain information that offer valuable insights about users‟ navigational behavior. For example we can understand the number of users that started from a page and proceeded through some certain pages and finally visited their goal page. Also, we can obtain some information about page popularity or some pages that contain the most information for a visitor. The most common form of pattern analysis is combining WUM tools with a knowledge query mechanism such as SQL. Content and structure information can be used to filter out patterns containing pages of a certain usage type, content type, or pages that match a certain hyperlink structure.

EXPERIMENTS AND PERFORMANCE EVALUATIONS

Extensive experiments were conducted, both on a data set collected from a real website. We set the out-degree threshold the path threshold (b) and the multiplier for the penalty term (m) to examine how results change with respect to these parameters. The math programs were coded in AMPL and solved using CPLEX/AMPL 8.1 on a PC running Windows XP on an Intel Core 2 Duo E6300 processor. The times for generating optimal solutions varied from 0.109 to 0.938 seconds, indicating that our model is very effective and practical for real-world websites. We have reported the times taken to solve the math programs only.

Results of Evaluation on Improved Website Using Percentage of Mini Sessions Enhanced for T=5 Minute

Table 1: Results of Evaluation

m	be	Mini Session Enhanced (%)		
		b=1	b=2	b=3
0	1	71.54	51.12	36.45
	2	86.26	72.04	57.35
	3	87.95	75.72	65.01
1	1	70.80	48.23	34.89
	2	85.15	69.11	53.97
	3	87.57	72.47	61.38
2	1	70.80	47.89	31.17
	2	85.29	68.48	53.28
	3	87.38	71.32	59.66

We evaluate performance using two metrics: the average number of paths per mini session and the percentage of mini sessions enhanced to a specified threshold. The first metric measures whether the improved structure can facilitate users to reach their targets faster than the current one on average, and the second metric measures how likely users suffering navigation difficulty can benefit from the improvements made to the site structure.

We claim that improved and newly added links could guide users to find their target pages more efficiently to some extent. This is because: 1) our method establishes efficient paths to target pages that were not available in the website structure before optimization, and 2) our method suggests improving links that would lead to users' target pages efficiently but missed by users (since they did not know what these links would lead to), so that more efficient navigation can be facilitated.

CONCLUSIONS

In this paper, we have proposed a web enhancement mathematical programming model to facilitate better navigation on the website by altering its link structure with minimal changes. This model is suitable for static websites whose contents are fixed. It is not suitable for dynamic website because of their volatile contents. It gives optimum solution than other techniques and is effective for n real web sites. This model will be improved with incorporating with data mining techniques.

REFERENCES

1. J. Hou and Y. Zhang, "Effectively Finding Relevant Web Pages from Linkage Information," IEEE Trans. Knowledge and Data Eng., vol. 15, no. 4, pp. 940-951, July/Aug. 2003.
2. R. Gupta, A. Bagchi, and S. Sarkar, "Improving Linkage of Web Pages," INFORMS J. Computing, vol. 19, no. 1, pp. 127-136, 2007.
3. C.C. Lin, "Optimal Web Site Reorganization Considering In-formation Overload and Search Depth," European J. Operational Research, vol. 173, no. 3, pp. 839-848, 2006.
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. R. Srikant and Y. Yang, "Mining Web Logs to Improve Web Site Organization," Proc. 10th Int'l Conf. World Wide Web, pp. 430-437, 2001.
6. 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.