ENHANCING THE WEB SITE STRUCTURE TO PROVIDE EASY TRAVERSAL ON A WEB SITE WITH MINIMUM CHANGES TO ITS CURRENT STRUCTURE

Bokefode J.D. *, Ubale S. A. **, Modani D. G. ***, Bhandare P.S. ****

*Research Scholar, Department of Computer Sciences and Engineering, Sinhgad College of Engineering, korti, Pandharpur, Solapur University, INDIA
**Assistant Professor, Department of Computer Sciences and Engineering, Sinhgad College of Engineering, korti, Pandharpur, Solapur University, INDIA
***Assistant Professor, Department of Computer Sciences and Engineering, Gharda Institute of Technology, Lavel, Chiplun, Mumbai University, INDIA
****Research Scholar, Department of Computer Sciences and Engineering, Sinhgad College of Engineering, korti, Pandharpur, Solapur University, INDIA

ABSTRACT

This paper provides a way for improving a website with minimum changes. Creating a new websites to provide better navigation becomes a challenging. A primary reason is that the webs developers’ may not have a clear understanding of user preferences and can only organize webpages based on their own judgment. For designing websites number of methods area available to connect webpages to improve user navigation on a websites, the completely reorganized new structure of a websites is not easy to predict. In this paper a model for web enhancement is introduced to improve the user navigation on a website with minimum changes to its current structure. Results from various tests are conducted on a globally available data set indicate that the presented model improves the user navigation with less changes. The presented model can also be used for larger data sets.

Key Words: Website structure improvement, web mining, directed tree, user navigation, and transformation.

1. INTRODUCTION

The growth of the internet is does not have any limitations. Now there are uncountable pictures, audios, videos, and other data available via internet and still it increasing. Finding required
data or any content from websites has become a very difficult task. In order to satisfy the increasing demands from online customers, Firms or organizations are heavily investing in the development and maintenance of their website. Web developer’s doest not have any idea about user’s requirements and can only organize pages based on their own judgment.

In this paper, we are concerned primarily 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. But this approach has lots of limitations. First, the complete reorganization could radically change the location of familiar items, the new website may disorient users [13]. Second, the reorganized website structure is not easy to predict, and the cost of unorientation about where you are and how to precede users after the changes remains unanalyzed. This is because a website’s structure is designed by experienced web Developers. Finally, Website reorganization models may change the current structure; therefore they are not used to improve the web site structures.

Perceiving the drawbacks of website reorganization approaches, we comprehend the raised question of how to improve the website structure rather than reorganizing it. A web enhancement model facilitates good user navigation without making more changes to its current structure. Our model is suitable for those websites whose contents are not changing as per user requirement & data within the websites are static. For example, 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 [14]. 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.

2. LITERATURE SURVEY

To resolve these questions number of techniques is proposed and they broadly classified into two types [1]: First is Personalization approach and second is Transformation approach. In personalization approach, for making easier access to particular user, it dynamically rebuilds the pages based on user profile and access paths. In Transformation approach we modify the website structure to make easier access and good navigation for all users.

Web personalization is the process of “altering” webpages to presuming the needs of particular users based on the user profile data and the way in which users’ prefer to access the web pages in a websites [6].

In [1], Perkowitz and Etzioni introduce an approach that synthesizes index pages automatically based on the co-occurrence frequency of pages in user log data which contain links to pages which belongs to particular topics, to providing easier access and good navigation for all users. In [7], [8], [9], [10], Mobasher and Yan introduce an approach that create the group or clusters of users profiles from weblog data and then dynamically reconstruct web links for users who belongs to different groups or clusters based on their access patterns. In [11], Nakagawa and Mobasher designed a hybrid personalization system that cans dynamically switching between the degree of connectivity and the user’s current location in the site to enhance the traversal path for that user. See [6] and [12], for understanding the web personalization approaches.
Web transformation, includes the making changes into website structure to provide easier access and good navigation for all users [4] instead of personalizing webpages for particular users.

In [3], Fu describe an approach to reconstruct webpages to provide users with their required contents in minimum clicks. But, this approach takes only local structures in a website instead of full structure. So the generated new structure does not give optimal navigational path. In [2], Gupta proposed a heuristic method based on simulating or altering the web links to improve the user traversal path. This method is applicable for both wired and wireless devices to change the web links structure in websites, for that it uses the frequently accessed data by the user. But, this approach takes very long time to small websites therefore it does not give optimal solutions. In [5], Lin design integer programming models, this model is based on the cohesion between pages to reduce depth of searching for users and overload of contents to reconstruct a website. But, this technique also requires very long time to large websites which contains many links therefore it does not give optimal navigational path. This model is not tested on real websites so it is appropriate or not for real web sites is still questionable. In [4], to resolve the problem in [5] Lin and Tseng design an ant colony system to reconstruct websites structures to provide easier access and good navigation for all users. But, it is not scalable for large websites.

The following table describes the summery of different methods which are used to improve the web site structure with their advantages and disadvantages.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>synthesizing index pages</td>
<td>Effectively provide the required web pages for that user</td>
<td>used for particular user</td>
<td>Give optimal solution to particular user</td>
</tr>
<tr>
<td>Clustering of user profiles</td>
<td>Dynamically generate links for user who belongs to particular group</td>
<td>used for particular user</td>
<td>Give optimal solution for particular group</td>
</tr>
<tr>
<td>Heuristic Method</td>
<td>Create or modify the structure of a website</td>
<td>it required 10 to15 hours and applicable for small website</td>
<td>Give optimal solution for all users.</td>
</tr>
<tr>
<td>Integer Programming Model</td>
<td>It gives better performance when web site contains many links.</td>
<td>Not applicable for real websites</td>
<td>Give optimal solution for all users but have low efficiency.</td>
</tr>
<tr>
<td>Ant Colony System</td>
<td>Used for real website and synthetic websites and computational time is low</td>
<td>Not Scalable on large sized- websites</td>
<td>Give optimal solution for all users for small websites</td>
</tr>
</tbody>
</table>

In this paper, we hypnotized Web Enhancement Model 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 web site maintenance.
3. PROBLEM STATEMENT

Complete reorganization of a website may lead to number of problems. First, it change the location of familiar items. Second, new websites may disorient users. Third, the reorganized structure is highly unpredictable. So; our aim is designing well-structured websites to provide better navigation on a website for all users with fewer changes to its current structure.

4. METHODOLOGY

In order to provide better navigation on a website for all users, the website structure must be reconstructed in a way that the different access paths required to get the desired contents in the newly improved structure have minimum traversal cost than the specified path threshold value. In this paper, we convert website structure into directed graph and this graph can be solved to get the optimum solution. We convert our website as graph optimization problem in which nodes represent web pages and the line between two nodes represents Links which connects two different web pages.

We designed a web enhancement model to reconstruct website structures, as shown in Figure 1. The input of the model is web log files and web site structure, and output is modified website structure. The presented model includes six parts: data pre-processing, directed graph, matrix representation of a graph, web enhancement, modified structure and Pattern Visualization.

4.1 Pre-processing

The raw web log data and web site structure is not suitable to convert into directed graph and it is not used directly for further processing, so it is required to process it to get useful data. Pre-processing involves data cleaning, user identification, session identification and path complement.
4.2 Directed Graph
For converting web site structure into directed graph following process will be followed.
Let root of the directed tree: t be the first requested page in user session;
Let pi the current request page and pj the next request page;
For each pi in user session
  If pj isn’t in the tree
    Create a new branch under pi with a solid line;
  Else
    Draw a dotted line from pi to pj;
After converting the website structure into directed graph it will be reduced into matrix representation.

![Directed Graph Example](image)

4.3 Matrix Representation of website structure
To determine whether the path is exist between two nodes or not, we use Warshall's Algorithm. In this we begin with creating adjacency matrix for a given graph. In this, instead of using weights it will use Boolean operators. If there is a path between two nodes enter a 1 in matrix otherwise enter 0 if path does not exists. Let N be the set of all webpages and $A_{ij}$, where $i,j \in N$, denote page connectivity in the current structure, with $A_{ij} = 1$ indicating page i has a link to page j, and $A_{ij} = 0$ otherwise. The current out-degree for page i is denoted by $W_i = \sum_{j=1}^{n} A_{ij}$.

![Adjacency Matrix Example](image)

Fig. 2 A website with 6 pages.

Fig. 3. Matrix Representation of website structure
4.4 Web Enhancement Model

The problem of improving user traversal on a website with fewer changes to its current structure can then be formulated using following formula.

\[
\text{Minimize } \sum_{(i,j) \in N} X_{ij}(1 - A_{ij}(1 - \varepsilon)) + m \sum_{i \in N} \pi_i
\]

Where,

- \(X_{ij}\) represents 1 if the link from page \(i\) to \(j\) selected; otherwise 0.
- \(m\) represents multipliers for the penalty term into the objective function.
- \(\pi_i\) represents number links that exceeds the out degree threshold.
- \(A_{ij}\) represents 1 if the \(i\) has a link to page \(j\) in the current structure; otherwise 0.
- \(\varepsilon\) is a very small number, in the objective function to let the model select existing links.
- \(N\) represents set of all links.

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).

4.5 Modified Web Structure

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.

5. METHODOLOGY

For reduction of the problem size we use mini sessions. Mini session represents the traversal path to visit targeted page. Mini session is based on a backtracking concept, if we are not getting solution by moving towards the forward direction then backtracks. Then from the obtained mini session we calculate the relevant candidate links which are required to visit the targeted page and remove irrelevant mini sessions based on the path threshold \((b)\) and out degree threshold \((C)\). Path threshold is nothing but the maximum number of nodes allowed to reach the target page in a mini session and out degree threshold represents the maximum number of links that page can have without being penalizing the web pages. Mini sessions represented by \(s\). The math programs were coded in AMPL and solved using CPLEX/AMPL 8.1 on a PC running Windows 7 on an Intel Core 2Duo E6300 processor. In web enhancement model we vary the path threshold \((b)\), and the multiplier for the penalty term \((m)\) to examine, the out-degree threshold \((C)\), how results change with respect to these parameters. We applied presented web enhancement model on the data set obtained from the matrimony site. Following table shows the result of our experiment.
Table 2
Result from the obtained data sets

<table>
<thead>
<tr>
<th>Time threshold(t)</th>
<th>Multiplier for penalty term(m)</th>
<th>Path threshold (b)</th>
<th>Out-degree threshold(C)=20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>No. of new Links</td>
</tr>
<tr>
<td>1 min.</td>
<td>0</td>
<td>1</td>
<td>9,179</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>1,645</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>596</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>9,181</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>1,559</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>609</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1</td>
<td>9,200</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1,559</td>
</tr>
<tr>
<td>3 min.</td>
<td>0</td>
<td>1</td>
<td>7,850</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>1,520</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>637</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>7,897</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>1,523</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>640</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1</td>
<td>7,896</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1,560</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>618</td>
</tr>
</tbody>
</table>

In Table 2, the column “No. of new links” indicates how many new links need to be added into the current structure in order to achieve the users navigation goal specified in the column “Path threshold (b)” for all mini sessions. The column “No. of excessive links” reports the number of new links added to pages that have excessive links, i.e., have more than C links. For example, if we set time threshold to $t=3$, out-degree threshold to $C=20$, path threshold to $b=1$, and multiplier ($m=1$), then we need to establish 7,897 new links and improve 1,702 existing links. Among all new links, 6,291 are added to pages having 15 more links. Approximately one new link is needed per four mini sessions.

6. FUTURE WORK

The results of experiments are stored in database. These results are so hard for analysis directly. Therefore we need to convert the abstract data into visual spatial forms by using information visualization technique. In future experiment we try to transform our results into spatial form so we easily see the improved structure of a web site structure.
7. CONCLUSION

In this paper, we have hypnotized a web enhancement model to provide better navigation on the website without reorganizing it. This model is suitable for static websites whose contents are fixed. It gives optimum solution than other techniques and is effective for real web sites. This model will be improved with incorporating with data mining techniques.

8. ACKNOWLEDGMENT

I would like to express my special thanks of gratitude to my guide prof. Ubale S.A. Madam who gave me the golden opportunity to write this important paper on the Web Enhancement, I am really thankful to them.

REFERENCES