Effective User Navigation Through Website Structure Improvement

T.Vijaya Madhavi, M.Vanaja
1Dept. of IT, Sir C R Reddy College of Engineering, Eluru, AP, India

Abstract
Designing well-structured websites to facilitate effective user navigation has long been a challenge. A primary reason is that the web developers' understanding of a given page should be structured can be considerably different from that of the users. While various methods have been proposed to restructure web pages to improve navigability using user navigation data, the completely reorganized new structure can be highly unpredictable, and the cost of disorienting users after the changes remains unanalyzed. This paper addresses how to improve a website without introducing substantial changes. Specifically, we propose a mathematical programming model to improve the user navigation on a website while minimizing alterations to its current structure. Results from extensive tests conducted on a publicly available real data set indicate that our model not only significantly improves the user navigation with very few changes, but also can be effectively solved. Evaluation results confirm that the user navigation on the improved structure is indeed greatly enhanced. More interestingly, we find that heavily disoriented users are more likely to benefit from the improved structure than the less disoriented users.

Keywords
Website Design, User Navigation, Web Mining, Mathematical Programming

I. Introduction
The advent of the Internet has provided an unprecedented spending by at least 11 percent, compared to that in 2006. Despite the heavy and increasing investments in website design, it is still revealed, however, that finding desired information in a website is not easy [4] and designing effective websites is not a trivial task [5-6]. Galletta et al. [7] indicate that online sales lag far behind those of brick-and-mortar stores and at least part of the gap might be explained by a major difficulty users encounter when browsing online stores. Palmer [8] highlights that poor website design has been a key element in a number of high profile site failures. McKinney et al. [9] also find that users having difficulty in locating the targets are very likely to leave a website even if its information is of high quality.

A primary cause of poor website design is that the web developers’ understanding of how a website should be structured can be considerably different from those of the users [10-11]. 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 users’ 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 [12].

Previous studies on website have focused on a variety of issues, such as understanding web structures [13], finding relevant pages of a given page [14], mining informative structure of a news website [6, 15] and extracting template from web pages [7]. Our work, on the other hand, is closely related to the literature that examines how to improve website navigability through the use of user navigation data. Various works have made an effort to address this question and they can be generally classified into two categories [11]: to facilitate a particular user by dynamically reconstituting pages based on his profile and traversal paths, often referred as personalization, and to modify the site structure to ease the navigation for all users, often referred as transformation.

In this paper, we are concerned primarily with transformation approaches. The literature considering transformations approaches mainly focuses on developing methods to completely reorganize the link structure of a website. Although there are advocates for website reorganization approaches, their drawbacks are obvious. First, since a complete reorganization could radically change the location of familiar items, the new website may disorient users [8]. Second, a reorganized website structure is highly unpredictable, and the cost of disorienting users after the changes remains unanalyzed. This is because a website’s structure is typically designed by experts and bears business or organizational logic, but this logic may no longer exist in the new structure when the website is completely reorganized. Besides, no prior studies have assessed the usability of a completely reorganized website, leading to doubts on the applicability of the reorganization approaches. Finally, since website reorganization approaches could dramatically change the current structure, they cannot be frequently performed to improve the navigability.

II. Metric for Evaluating Navigation Effectiveness

A. The Metric
Our objective is to improve the navigation effectiveness of a website with minimal changes. Therefore, the first question is, given a website, how to evaluate its navigability. Marsico and Levialdi [10] point out that information becomes useful only when it is presented in a way consistent with the target users’ expectation.

Fig. 1: A Website With 10 Pages

Palmer [1] indicates that an easy-navigated website should allow users to access desired data without getting lost or having to backtrack. We follow these ideas and evaluate a website’s navigation effectiveness based on how consistently the information is organized with respect to the user’s expectations. Thus, a well-structured website should be organized in such a way that the discrepancy between its structure and users’ expectation of the
structure is minimized. Since users of informational websites typically have some information targets [9, 12], i.e., some specific information they are seeking, we measure this discrepancy by the number of times a user has attempted before locating the target.

B. An Example
We use an example to illustrate the aforementioned concepts and how to extract the metric from weblog files. To analyze the interaction between users and a website, the log files must be broken up into user sessions. Cooley et al. [4] define a session as a group of activities performed by a user during his visit to a site and propose timeout methods to demarcate sessions from raw log files. In this definition, 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, we use a different term mini session 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.

III. Computational Experiments and Performance Evaluations
Extensive experiments were conducted, both on a data set collected from a real website and on synthetic data sets. We first tested the model with varying parameters values on all data sets. Then, we partitioned the real data into training and testing data. We used the training data to generate improved site structures which were evaluated on the testing data using two metrics that are discussed in detail later. Moreover, we compared the results of our model with that of a heuristic.

A. Real Data Set

1. Description of the Real Data Set
The real data set was collected from the Music Machines website (http://machines.hyperreal.org) and contained about four million requests that were recorded in a span of four months. This data set is publicly available and has been widely used in the literature [9-10]. Table 6 shows the number of pages in the website that had out-degrees within a specified range. This website has in total 916 pages, of which 716 have an out-degree of 20 or less, with the majority (83 percent) of the remaining pages having 40 links or less. Before analysis, we followed the log preprocessing steps described in [9] to filter irrelevant information from raw log files. These steps include: 1) filter out requests to pages generated by Common Gateway Interface (CGI) or other server-side scripts as we only consider static pages that are designed as part of a website structure, 2) ignore un-successful requests (returned HTTP status code not 200), and 3) remove requests to image files (.gif, .jpg, etc.), as images are in general automatically downloaded due to the HTML tags rather than explicitly requested by users.

IV. Discussion

A. Mini Session and Target Identification
We employed the page-stay timeout heuristic to identify users’ targets and to demarcate mini sessions. The intuition is that users spend more time on the target pages. Page-stay time is a common implicit measurement found to be a good indicator of page/document relevance to the user in a number of studies [14]. In the context of web usage mining, the page-stay timeout heuristic as well as other time-oriented heuristics are widely used for session identification [15], and are shown to be quite robust with respect to variations of the threshold values [6].

B. Searching Sessions Versus Browsing Sessions
While the majority of users typically have one or more goals and search for particular pieces of information when navigating a website (“searching” sessions), some users might simply browse for general information (“browsing” sessions). Though exact distinction between these two web surfing behaviors is often impossible by only looking at the anonymous user access data from weblogs, certain characteristics can help differentiate the two types of sessions. For example, some visits are clearly purposeless and finish abruptly at pages that cannot be target pages, so these sessions are more likely to be browsing sessions. To the best of our knowledge, there is no algorithm developed to distinguish between the two types of sessions and further investigation on this question is needed. While we did not explicitly separate searching sessions from browsing sessions, the preprocessing steps can help eliminate many purposeless browsing sessions. As a result, the final improved website structure mainly “shortens” searching sessions and also reduces purposeful browsing sessions.

C. Implications of This Research
This research contributes to the literature on improving web user navigation by examining this issue from a new and important angle. We have performed extensive experiments on both real and synthetic data sets to show that the model can be effectively solved and is highly scalable. In addition, the evaluation results confirm that users can indeed benefit from the improved structure after suggested changes are applied. There are several important implications from this research.

D. Choice of Parameter Values for the Model
Path threshold: The path threshold represents the goal for user navigation that the improved structure should meet and can be obtained in several ways. First, it is possible to identify when visitors exit a website before reaching the targets from analysis of weblog files [4, 8]. Hence, examination of these sessions helps make a good estimation for the path thresholds. Second, surveying website visitors can help better understand users’ expectations and make reasonable selections on the path threshold values. For example, if the majority of the surveyed visitors respond that they usually give up after traversing four paths, the path threshold should be set to four or less. Third, firms like com Score and Nielsen have collected large amounts of client-side web usage data over a wide range of websites. Analyzing such data sets can also provide good insights into the selection of path threshold values for different types of websites.

F. Multiplier for the Penalty Term
The use of the penalty term can prevent the model from adding new links to pages that already have many links. This helps keep the information load low for user at the cost of inserting more new links into other pages with small out-degrees. Generally, if website has both pages with small out-degrees and pages with very large out degrees.

G. Evaluation Procedure
We used simulations to approximate the real usage and to evaluate how the user navigation could be enhanced in the improved website structure. The use of simulation for website usability
evaluation is very popular and has been widely used in modeling users’ choices in web navigation and usability test. However, simulation studies often have to make simplifying assumptions in order to simulate real-life scenarios, posing questions on the generalizability of the results. In the context of our simulation approach, we assume that users would find their target pages effectively through a new/improved link if it exists. In practice, certain criteria related to the visual design of web interfaces need to be followed in order to effectively apply the suggested changes to a website. We note that there exist an abundant literature on both webpage design [6,8], and hyperlink design. Though we did not explicitly consider design issues in this paper, we do assume that Webmasters follow the guidelines and suggestions from such studies when creating and editing links and designing web pages. Consequently, in the simulation approach used for user navigation evaluation, we assume that new links are carefully designed and existing links are appropriately edited. In addition, they should also be placed in proper places for users to easily locate. Thus, these links should provide users with accurate knowledge on the contents on the other end of a link and help them make correct selections.

V. Conclusions

Our model is particularly appropriate for informational websites whose contents are relatively stable over time. It improves a website rather than reorganizes it and hence is suitable for website maintenance on a progressive basis. The tests on a real website showed that our model could provide significant improvements to user navigation by adding only few new links. Optimal solutions were quickly obtained, suggesting that the model is very effective to real-world websites. In addition, we have tested the MP model with a number of synthetic data sets that are much larger than the largest data set considered in related studies as well as the real data set. The MP model was observed to scale up very well, optimally solving large-sized problems in a few seconds in most cases on a desktop PC.

To validate the performance of our model, we have defined two metrics and used them to evaluate the improved website using simulations. Our results confirmed that the improved structures indeed greatly facilitated user navigation. In addition, we found an appealing result that heavily disoriented users, i.e., those with a higher probability to abandon the website, are more likely to benefit from the improved structure than the less disoriented users. Experiment results also revealed that while using small path thresholds could result in better outcomes, it would also add significantly more new links. Thus, Webmasters need to carefully balance the tradeoff between desired improvements to the user navigation and the number of new links needed to accomplish the task when selecting appropriate path thresholds. Since no prior study has examined the same objective as ours, we compared our model with a heuristic instead. The comparison showed that our model could achieve comparable or better improvements than the heuristic with considerably fewer new links.

The paper can be extended in several directions in addition to those mentioned in Section 6. For example, techniques that can accurately identify users targets are critical to our model and future studies may focus on developing such techniques. As another example, our model has a constraint for out-degree threshold, which is motivated by cognitive reasons. The model could be further improved by incorporating additional constraints that can be identified using data mining methods. For instance, if data mining methods find that most users access the finance and sports pages together, then this information can be used to construct an additional constraint.

References