Volume 4, Issue 10, October 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

# Website Structure Transformation by Page Access Connectivity Mining for Optimal User Navigation

**L. Dilip**
M.Tech in Computer Science & Engineering
JNTUA college of Engineering,
Pulivendula, Andhrapradesh, India

**Smt. S. Jessica Saritha**
Assistant professor, dept. of CSE
JNTUA college of Engineering,
Pulivendula, Andhrapradesh, India

***Abstract— Creating well-structured sites to enhance reliable user consultation has long been a concern. A main purpose is that the web designer's knowledge of how a website must organize might significantly different from that of the customers. Although several techniques have been suggested to relink websites to enhance navigability with user routing information, the entirely reorganized unique design could be extremely volatile, and the price of disorienting consumers after the modifications remains unanalysed. This report covers how to enhance a webpage without providing significant modifications. Particularly, we suggest a mathematical development design to enhance the user menu on a webpage while reducing changes to its actual design. Outcomes from extended tests performed on a widely obtainable real data set suggest that our design not only considerably enhances the user menu with few modifications, but also might efficiently fixed. Assessment outcomes validate that the user menu on the enhanced structure is certainly highly improved. Most curiously, we choose that greatly disoriented customers are more probably to reward from the enhanced structure than the lower disoriented customers.***

***Keywords— Website Design, User Navigation, Web Mining, Mathematical Programming s***

## I. INTRODUCTION

The arrival of the Net has supplied an extraordinary taking by at minimum 11 percent, than that in 2006. Besides the significant and growing investments in webpage layout, it is even presented, anyhow, that choosing preferred data in a webpage is hard [4] and creating efficient web sites is not a insignificant process [5-6]. Galletta et al. [7] suggest that online selling lag much about those of brick-and-mortar storage and about portion of the gap can be revealed by a significant problem customers notice when exploring online vendors. Palmer [8] shows that terrible website layout has become a key component in a quantity of large profile website problems. McKinney et al. [9] even choose that customers having problem in finding the objectives are extremely prone to leave a web site even when its data is of elevated excellence.

A key result of terrible webpage layout is that the online designers' knowledge of how a webpage must be organized could be significantly assorted from that from the customers [10-11]. These variations effect in instances where customers can not definitely select the preferred data in a webpage. This issue is hard to prevent due to when generating a webpage, web designers cannot have an obvious knowledge of customers' choices and can just arrange web pages according to their own assessment. Anyhow, the evaluate of webpage efficiency must be the approval of the customers compared to that of the designers. Hence, Websites must be arranged such that usually meets the user's design of how content must be arranged [12].

Earlier reports on webpage has centered on a selection of concerns, like comprehending web architecture [13], choosing appropriate content of a provided website [14], mining helpful design of a news web site [6, 15] and removing theme from websites [7]. Our move, however, is intently regarding the literary works that analyzes how to enhance webpage navigability thru the usage of customer routing information. Assorted works have prepared a hassle to handle this query and they may be usually categorized into two types [11]: to enhance a specific user by dynamically reconstituting content according to his profile as well as traversal routes, frequently introduced as customization, also to change the website design to alleviate the routing for many customers, usually introduced as change.

In this report, we are involved mainly with transformation strategies. The literature regarding transformations strategies primarily concentrates on providing techniques to completely rearrange the link system of a webpage. Though there are recommends for webpage reorganization strategies, their issues are noticeable. Initially, since a full reorganization could significantly alter the location of common objects, the new webpage may disorient customers [8]. Secondly, the reorganized webpage design is definitely volatile, and the cost of disorienting customers after the transforms keeps unanalyzed. This is considering a website's design is usually created by specialists and contains business or business logic, but this logic may not exist in the new design when the webpage is entirely reorganized. Furthermore, no earlier reports have evaluated the functionality of an entirely reorganized webpage, causing concerns on the usefulness of the reorganization techniques. Ultimately, since webpage reorganization techniques could significantly alter the updated design, they can't be commonly executed to enhance the navigability.

 

## II. LITERATURE REVIEW

An Website navigation among all design aspects has long been one of the crucial design principle for any domain like finance, entertainment, e-commerce, government, education, or medical [10]. Website design quality has been researched from qualitative as well as quantitative perspective [12]. In literature, website design has been studied in different viewpoints; however it has long been termed as a hypermedia or a database. In engineering perspective, web engineering and IDEAL consider modelling and further enhancement. In human computer communication's perspective for website design the interface elements like graphical, layout, usability design and analysis play a major role. System design aspects like hardware design, cache scheduling, etc. influence the website performance. Structure design like hyperlinks configuration and information structural design affect website navigation.

## III. WEBSITE EVALUATION

All Information systems evaluation can be classified as design evaluation and outcome evaluation [11]. Website design from the standpoint of human-computer interaction studies two aspects mainly i.e. usability and interface design. Previous research's strategies cumulative finding for website usability or interface design from a cognitive or empirical perspective recommend high-level design principles.

### *Interface-Based Evaluation*

A major part of previous research has considered for website usability of interface design a cognitive or empirical approach. The findings have led to specific design principles or guidelines. Zhang and Dran recommend for website design and evaluation a hygiene and motivator model with focus on requirements for interface design [13]. Ivory surveys developed automated website evaluation for both research and practice and which was in its infancy stage and concluded the subject requires persistent research and evaluation. The challenges faced in website design are far greater compared to designing conventional systems. Further developments in the field of website design led to new development tools and methodologies. These developments gained considerable attention from researchers as well as practitioners. However advances in website evaluation and enhancement practices remained slowly.

### *Structure-Based Evaluation*

Analytical modelling strategies have been studied for website design also. Let's consider an analytical survey of Web metrics. The survey examines essential graph characteristics relevant for website design. Quantifying Web graph properties, classifies them into a set of important metrics including page significance, page similarity, search and retrieval, usage characterization and information theoretic properties for improved Web information availability and use. Wu et al. studies Web graph properties, to find relevant Web graph properties by concentrating on compactness and stratum which together give high-level guidelines for website designers. Brin, models website structural designs with graph to get ranking of pages. Monika and Ricca research Web hyperlinks and compare various ranking approaches [14]. Web mining methods were used to examine content, structure as well as usage. Kumar et al. propose a stochastic model for developing a Web graph having edges which rely statistically on each other and allows the development of dynamic new vertices over time. Sarukkai develops a Markov Chain model by analyzing popular content-access behaviour's by visitors for link prediction and path analysis. Zhou et al. investigate individual search behaviours and patterns on a website to assess the design of its link frameworks.

## IV. IMPROVEMENT OF NAVIGATION EFFICIENCY

Navigation efficiency is dependent on design at various levels i.e. at the system level, structure level and at the page level. System level design is defined by the web application workload and CPU requirements. Page design adheres to user interface guidelines to search and display relevant amount of content. Structural level design is given priority in Web usage mining analysis. Customization and transformation however exist as the two basic rules for improving browsing efficiency.

In the last section, transformation has been given lesser attention for improving Website structure; however Website reorganization is crucial for logical design and also to meet user's requirements.

### *Definition of Navigation Efficiency*

Navigation is the basic tasks of web search and information finding activity, (e.g. browsing for product information or a person's name) involves 25% of the total search activity [15]. Dictionary defines efficiency as the property of being efficient or the degree to which this quality is practiced. It defines efficiency as ratio of useful or effective output vs. the total input in any system or the ratio of the output energy of a machine to the input energy given for its operation. Navigation efficiency can be similarly defined by considering the average information accumulated in navigation. The average performance is the expected performance. Example the typical number of clicks needed to lead a visitor from one page to page. The definition of efficiency considers page importance, volume of information and page loading time / number of clicks. Related metrics are analyzed, correlations studied, for improving navigation efficiency. Operation efficiency or performance efficiency is time needed to complete the task, reflects Navigation efficiency. Operating efficiency is associated with a group of users when browsing a website and a websites operating efficiency is defined as a measurement of the amount of data transferred through users' interactive search manners. Efficiency = Shortest path (shortest path from first page to target page) / user operating expenses. Overall weighted distance between web pages has led to optimization methods making popular pages more accessible.

 

## V. METHODOLOGY

This study puts forward a definition for navigation efficiency. Navigation efficiency is defined by assessing best, worst and average cases. Ex; an optimistic view proposes efficiency as a model of the least navigation path from home page to destination. A pessimistic view, defines navigation efficiency modelled as the longest path between any two nodes. This study uses the definition of average information accumulation performance for navigation efficiency. Metric used in this research as well as the correlations between interrelated metrics are explained. For showing various possible navigation objectives several computation are explained for representing different possible navigation objectives. Website stability is considered upmost to ease website maintenance according to user's various requirements. Heuristics were developed considering problem properties to increase the efficiency objective function. The strategy is adaptable to varied user required specific constraints as well as stability constraints.

## VI. WEB EFFICIENCY MODELLING AND IMPROVEMENT

In this study Web navigation is transformed with definite user requirements into a graph problem. The study scope is limited to two factors for webmasters and users.

(1) Minimize extra work for webmaster. Original design of the site is protected from detrimental changes. The links that can be added or deleted is limited and also pages that can be created or destroyed are limited.

(2) Transformation of website should be done for everybody, both for first-time and casual users. Customization is beneficial to experienced users, but not for first-time users.

### *Modelling and Evaluation*

In this study we try to enhance web navigation efficiency considering website structure stability as a key factor. Previous studies have also proposed, user requirements modelling and transformation into common design and further into graphical problems (Yen 2007). In case the constraints are inconsistent, we can deactivate some and assign lower priority. Some studies propose binary search using inconsistent group of constraints and deactivation of all constraints generating an inconsistency. This study, limits itself to only consistent constraints. Navigation modelling along optimistic, pessimistic or average conditions, we can consider information accumulation along shortest, longest, or average path length. To determine the worst case scenario in navigation, we can consider a longest path of navigation; however this scenario has been confirmed as NP and difficult to be resolved.

### *Efficiency Improvement*

To enhance efficiency of website navigation as a whole, we have to consider efficiency of every route. Identification of inefficient routes, will improve the efficiency on the whole, while not affecting the efficiency on other paths. Efficiency of one route can be improved by taking the average information accumulation between two nodes of the route. This can be done by adding a link to form a path reducing the length of the route. Link addition can decrease repeated path, decrease accumulation of unnecessary or repeated information and maintain necessary information. Heuristic can be designed to identify such links.

By adding a link the information collected on shortest route will remain unchanged. Addition of hotlinks (shortcut) will decrease average number of steps in searching. However as the number of shortcuts (hotlinks) increases, the complexity of graph increases and the search time may not decrease.

The methods of adding, deleting, and other hybrid approaches can be used to get structure with better average navigation efficiency. However due to page limits constraints, we consider only adding a link. This study proposes, a branch-and-bound approach implicitly enumerating all the possible solutions and decreasing the integer constraints. Even though it can be very time-consuming, but is in principle capable of yielding an exactly optimal solution. The main purpose of the method is to cut computation by eliminating in the tree nodes / subsets, having zero possibility of an optimal solution. Solution of a problem with branch-and-bound method can be expedited with addition of side constraints making the viable set in limited size and with no integer constraint.

## VII. DISCUSSION

### *A. Mini Session and Target Identification*

Here the page-stay timeout heuristic is used to find users targets and to separate mini sessions presuming users devote more time on the target pages. Page-stay time is a general implicit metric serving as an effective demonstrator of relevance of page/document to the user in various studies [14]. Web usage mining, models on the page-stay timeout heuristic and other time-oriented heuristics used widely for session identification [15], are very efficient even with respect to variations of the threshold values [6].

### *B. Searching Sessions versus Browsing Sessions*

Although most of the users regularly have predefined objective and search for relevant information while navigating a website ("searching" sessions), some users could very well browse for common information ("browsing" sessions). Exactly distinguishing these two browsing patterns is often unlikely. Though looking at the anonymous user access data from weblogs, certain qualities can help differentiate the two types of sessions. For example, a few visits are evidently without any purpose and finished at pages that are not relevant pages and such visits are likely browsing sessions. There is no standard algorithm developed to our knowledge to distinguish the two session's types and further study on this topic is needed. Though this study has not clearly separated searching sessions from browsing sessions, the steps discussed can

 

help remove many aimless browsing sessions. The result, the final enhanced website structure implicitly "shortens" search sessions as well as decreases purposeful browsing sessions.

### *C. Implications of This Research*

This study contributes to the literature on improving web user navigation by showing the case from a new and important perspective. Extensive experiments performed using real as well as synthetic data sets, show that the model is effective and are also highly scalable. The evaluation results confirm users benefit from the enhanced structure after applying the suggested changes. This research has several important implications.

### *D. Choice of Parameter Values for the Model*

Path threshold: The path threshold shows the objective for user navigation that the improvised model should fulfill and could be gained in various ways. Analysis of weblog files will first make it possible to find when visitors leave a website before locating targets [4, 8] giving us a good evaluation of the path thresholds. The analysis also helps perceive users' expectations and make sensible selections of the path threshold values. As an example, if the majority of a sample survey of visitors usually gives up after navigating four paths, then we can set path threshold to four or less. Third, firms such as com Score and Nielsen have gathered huge amounts of client-side web usage data over a wide selection of websites. Data analysis of this information can lead to efficient ways to select path threshold values of various types of websites.

### *F. Multiplier for the Penalty Term*

The penalty term used can prevent the model from putting new links to pages already having several links. This assists in maintaining the information load low for user at the cost of inserting more new links into other pages with small out- degrees, especially if website has both pages with small out-degrees and pages with very large out degrees.

### *G. Evaluation Procedure*

We used simulations to estimate the real usage and to evaluate how user navigation could be improved in the modified website structure. Studying website usability evaluation using simulation is an accepted practice in modeling users' requirements in web navigation and usability test. Simulations studies however in order to simulate real life situations often have to make simplifying assumptions, raising questions on the generalizability of the results. Here in our context, in the simulation approach we presume that users through a new/ improved link if it exists will locate their target pages effectively. In order to effectively apply the suggested changes to a website, in principle certain criteria related to the visual design of web interfaces need to be followed. Here we mention that there is abundant literature on webpage design [6, 8], as well as hyperlink design. Though design issues have not been considered in this paper, it's assumed that Webmasters follow the guidelines and suggestions of such studies while designing web pages and creating/ editing links. Similar guidelines for developing and editing links have to be adhered while evaluating user navigation using simulation approach. The links should help the users make correct selections by placing them in relevant places for users to easily locate and provide them with relevant contents.

## VIII. CONCLUSIONS

Our model is specifically suitable for informational websites where the contents are relatively stable over a period of time. It enhances the website, however does not reorganize it and hence is ideal for website maintenance on an ongoing basis. The trails on a real website demonstrated that our model can offer substantial improvements to user navigation by adding only few new links. It helped in gaining optimal solutions quickly, implying that the model is efficient for real-world websites. The MP model was tested with multiple sets of synthetic data much larger than the largest data set considered in similar studies as well as the real data set. The MP model effectively scaled up, solving optimally large-sized issues in few seconds in majority of scenarios on a desktop PC.

To authenticate the efficiency of our model, we have described two metrics to evaluate the enhanced website with simulations. The results proved that the improved framework immensely promoted user navigation. An apparent result is that, greatly disoriented users or those having a greater probability to leave the website benefited more from the improved structure compared to their less disoriented counterparts. The results also reveal that using small path thresholds not only result in better outcomes but also would add considerably more new links. Here Webmasters have to intelligently balance the tradeoff between desired modifications to the user navigation and the number of new links required to achieve the task while choosing suitable path thresholds. As no previous research has evaluated similar objective as ours, we reviewed our model with a heuristic as an alternative. The results demonstrate that our model can achieve similar or better results than the heuristic with significantly fewer new links.

## REFRENCES

[1] Pingdom,"Internet 2009 in Numbers", [Online] Available: http://royal.pingdom. com/2010/01/22/internet-2009-in- numbers/, 2010.

[2] J. Grau,"US Retail e-Commerce: Slower But Still Steady Growth", [Online] Available: http://www.emarketer.com/ Report.aspx?code=emarketer_ 2000492, 2008.

[3] Internetretailer, "Web Tech Spending Static-But High-for the Busiest E-Commerce Sites", [Online] Available: http://www. internetretailer.com/ dailyNews.asp?id = 23440, 2007.

 

[4] D. Dhyani, W.K. Ng, S.S. Bhowmick,"A Survey of Web Metrics", ACM Computing Surveys, Vol. 34, No. 4, pp. 469¬503, 2002.
[5] X. Fang, C. Holsapple,"An Empirical Study of Web Site Navigation Structures' Impacts on Web Site Usability", Decision Support Systems, Vol. 43, No. 2, pp. 476-491, 2007.
[6] J. Lazar,"Web Usability: A User-Centered Design Approach", Addison Wesley, 2006.
[7] D.F. Galletta, R. Henry, S. McCoy, P. Polak,"When the Wait Isn't So Bad: The Interacting Effects of Website Delay, Familiarity, and Breadth", Information Systems Research, Vol. 17, No. 1, pp. 20-37, 2006.
[8] J. Palmer,"Web Site Usability, Design, and Performance Metrics", Information Systems Research, Vol. 13, No. 2, pp. 151-167, 2002.
[9] Expectation and Disconfirmation Approach", Information Systems Research, Vol. 13, No. 3, pp. 296-315, 2002.
[10] T. Nakayama, H. Kato, Y. Yamane,"Discovering the Gap between Web Site Designers' Expectations and Users' Behavior", Computer Networks, Vol. 33, pp. 811-822, 2000.
[11] M. Perkowitz, O. Etzioni,"Towards Adaptive Web Sites: Conceptual Framework and Case Study", Artificial Intelligence, Vol. 118, pp. 245-275, 2000.
[12] J. Lazar,"User-Centered Web Development", Jones and Bartlett Publishers, 2001
[13] Y. Yang, Y. Cao, Z. Nie, J. Zhou, J. Wen,"Closing the Loop in Webpage Understanding", IEEE Trans. Knowledge and Data Eng., Vol. 22, No. 5, pp. 639-650, May 2010.
[14] J. Hou, 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.
[15] J. Hou, 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

 