International Journal of Science Engineering and Advance Technology

Enlightening Website Structure for Enabling Client Steering Effectually

S Naga Kartheek Dhulipala#1, Potti Kusuma Tejaswini#2
#1 Student of M.Tech Department Of Computer Science Engineering G.V.R&S college of Engineering & Technology, GUNTUR
#2 Department of Computer Science Engineering G.V.R&S college of Engineering & Technology, GUNTUR.

ABSTRACT—Designing well-Constructiond websites to facilitate Operative user Steering has long been a challenge. A primary reason is that the web developers' understanding of how a website should be Constructiond can be considerably different from that of the users. While various methods have been proposed to relink webpages to improve navigability using user Steering data, the completely reorganized new Construction 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 Steering on a website while minimizing alterations to its current Construction. Results from extensive tests conducted on a publicly available real data set indicate that our model not only significantly improves the user Steering with very few changes, but also can be Operatively solved. We have also tested the model on large synthetic data sets to demonstrate that it scales up very well. In addition, we define two evaluation metrics and use them to assess the performance of the improved website using the real data set. Evaluation results confirm that the user Steering on the improved Construction is indeed greatly enhanced. More interestingly, we find that heavily disoriented users are more likely to benefit from the improved Construction than the less disoriented users.

Index Terms—Website design, user Steering, web mining, mathematical programming

.1 INTRODUCTION

THE advent of the Internet has provided an unprecedented platform for people to acquire knowledge and explore information. There are 1.73 billion Internet users worldwide as of September 2009, an increase of 18 percent since 2008 [1]. The fast-growing number of Internet users also presents huge business opportunities to firms. According to Grau [2], the US retail ecommerce sales (excluding travel) totaled $127.7 billion in 2007 and will reach $218.4 billion by 2012. In order to satisfy the increasing demands from online customers, firms are heavily investing in the development and maintenance of their websites. InternetRetailer [3] reports that the overall website operations spending increased in 2007, with one-third of site operators hiking 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 Operative websites is not a trivial task [5], [6]. Galletta et al. [7] indicate that online sales lag far behind those of brickandmortar 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 [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 Operativeness should be the satisfaction of the users rather than that of the developers. Thus, Webpages should be organized in a way that generally matches the user's model of how pages should be organized [12].

In this paper, we are concerned primarily with transformation approaches. The literature considering transformations approaches mainly focuses on developing methods to completely reorganize the Second, the reorganized website Construction is highly understanding of how a website should be website, leading to doubts on the applicability of the reorganization Constructiond can be considerably different from those of the users approaches. Finally, since website reorganization approaches could dramatically change the current Construction, they cannot be experiments on synthetic data indicate that our model also scales up frequently performed to improve the navigability.

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 Out-Degree Statistics improved site Constructions 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.

Real Data Set

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 [19], [29]. 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 [29] 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 Construction, 2) ignore unsuccessful 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 [33].

We utilized the page-stay time to identify target pages and to demarcate mini sessions from the processed log files. Three time thresholds (i.e., 1, 2, and 5 minutes) were used in the tests to examine how results changes with respect to different parameter values. Furthermore, we adapted the algorithm proposed in [40] to identify the backtracking pages in mini sessions, which are then used to demarcate the paths traversed to reach the targets. Table 7 lists the number of mini sessions comprising a given number of paths ð> 1Þ for different time thresholds.

Results and Analysis —Real Data Set

We set " ¼ 1:0E 8 and vary the out-degree threshold ðCÞ, the path threshold ðbÞ, and the multiplier for the penalty term ðmÞ to examine how results change with respect to these parameters.

Out-degree	Number of pages
>100	1
81–100	2
61–80	10
41–60	21
21–40	166
11–20	538
0–10	178
Total	916

Table 8 reports the experiment results. 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 Path Characteristics of Mini have reported the times taken to solve the math solutions varied from 0.109 to 0.938 seconds, evaluate

TABLE 7

Number of paths	Number of mini sessions		
	t=1 min.	t=2 min.	t=5 min.
2	27,140	23,485	20,964
3	4,457	4,242	4,075
4	1,340	1,469	1,427
5	477	590	652
6–10	395	498	525
>10	3	8	7
Total	33,812	30,292	27,650

The times taken for preprocessing steps and obtaining values of a ^{S}ijkr are not included, as they can be done very quickly in practice. Note that the size of the real website considered in our paper is significantly larger than the average website size [52] as well as those used in related papers addressing the website reorganization problem. For example, Gupta et al. [19] and Lin and Tseng [28] report results based on websites with only 427 and 146 pages, respectively.

CONCLUSIONS

In this paper, we have proposed a mathematical programming model to improve the Steering Operativeness of a website while minimizing changes to its current

Construction, a critical issue that has not been examined in the literature. Our model is particularly[1] 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 Developments to user Steering by adding only few new links. Optimal solutions were quickly obtained, suggesting that the model is very Operative to realworld 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 largesized 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 the improved website using simulations. Our results confirmed that the improved Constructions indeed greatly facilitated user Steering. 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 Construction 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 Developments to the user Steering 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 Developments 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 outdegree 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 [65]. 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

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

[2] J. Grau, "US Retail e-Commerce:Slower But Still Steady Growth, http://www. emarketer. com/ Report. aspx ?cod e=emarketer_2000492

[3] D. Dhyani, W.K. Ng, and S.S. Bhowmick, "A Survey of Web Metrics," ACM Computing Surveys, vol. 34, no. 4, pp. 469-503, 2002.

[4] X. Fang and C. Holsapple, "An Empirical Study of Web Site Steering Constructions' Impacts on Web Site Usability," Decision Support Systems, vol. 43, no. 2, pp. 476-491, 2007.

[5]J. Lazar, Web Usability: A User Centered Design Approach.Addison Wesley,2006.

[6]D.F. Galletta, R. Henry, S. McCoy, and 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. 2037, 2006.

[7]J. Palmer, "Web Site Usability, Design, and Performance Metrics," Information Systems Research, vol. 13, no. 2, pp. 151 167, 2002.

[8]V. McKinney, K. Yoon, and F. Zahedi, "The Measurement of WebCustomer Satisfaction: An Expectation and Disconfirmation Approach," Information Systems Research, vol. 13, no. 3, pp. 296315, 2002.

[9]T. Nakayama, H. Kato, and Y.Yamane, "Discovering the Gap between Web Site Designers' Expectations and Users'Behavior," Computer Networks, vol. 33, pp. 811-822, 2000.

[10]M. Perkowitz and O. Etzioni, "Towards Adaptive Web Sites:Conceptual Framework and Case Study," Artificial Intelligence, vol. 118, pp. 245 275, 2000.

[11]J. Lazar, User-Centered Web Development.Jones and Bartlett Publishers, 2001.

[12] Y. Yang, Y. Cao, Z. Nie, J. Zhou, and J. Wen, "Closing the Loop in Webpage Understanding,"

IEEE Trans. Knowledge and Data Eng., vol. 22, no. 5, pp. 639 650, May 2010.

[13]J. Hou and Y. Zhang, "Operatively Finding Relevant Web Pages from Linkage Information," IEEE Trans. Knowledge and Data Eng., vol. 15, no. 4, pp. 940-951, July/Aug. 2003.

[14]H.Kao, J.Ho, and M. Chen, "WISDOM: Web Intrapage Informative Construction Mining Based on Document Object Model," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 5, pp. 614-627, May 2005.

[15]H. Kao, S. Lin, J. Ho, and M. Chen, "Mining Web Informative Constructions and Contents Based on Entropy Analysis," IEEE Trans. Knowledge and Data Eng., vol. 16, no. 1, pp. 41-55, Jan. 2004.

[16]C. Kim and K. Shim, "TEXT: Automatic Template Extraction fromHeterogeneous Web Pages," IEEE Trans.Knowledge and Data Eng., vol. 23, no. 4, pp. 612-626, Apr. 2011.

[17]M. Kilfoil et al., "Toward an Adaptive Web: The State of theArt and Science," Proc. Comm. Network and Services Research Conf., pp. 119130, 2003.

18]R. Gupta, A. Bagchi, and S. Sarkar, "Improving Linkage of Web Pages,"INFORMS J. Computing, vol. 19, no. 1, pp. 127-136, 2007.

[19]C.C. Lin, "Optimal Web Site Reorganization Considering Information Overload and Search Depth, " European J. Operational Research, vol. 173, no. 3, pp. 839-848, 2006.

AUTHORS PROFILE:

S NAGA KARTHEEK DHULIPALA is a student of Computer Science Engineering from G.V.R&S college of Engineering & Technology Presently pursuing M.Tech (CSE) from this college. He received MCA from ANU in the year of 2010.

POTTI KUSUMA TEJASWINIis aAssistant Professor at G.V.R&S college of Engineering & Technology, Guntur. She received M.Tech in Computer science engineering from JNTUK. She gained 3 years of Experience on Teaching. She is a good Researcher in Computer network and Data Mining.