International Journal of Innovative Research in Advanced Engineering (IJIRAE)
Volume 1 Issue 2 (April 2014)

A Survey on Improving User Navigation through Website Structure Improvement

S.R. Nagpurkar
Dept.of Comp.Engg.PVPIT, Pune
snagpurkar@gmail.com

Y.B.Gurav
Dept. of Comp.Engg.PVPIT, Pune
ybgurav@gmail.com

Abstract— Website navigation has been seemed as one of the most important design features across many domains, including finance, e-commerce, entertainment, education, government, and medical A main reason is that the web developers indulgent of how a website should be structured can be significantly different from that of the users. While various method have been proposed to relink WebPages 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 .The start of the Internet has provided an extraordinary platform for people to gain knowledge and explore information. How to efficiently and effectively retrieve required Web pages on the Web is becoming a challenge.There are 1.73 billion Internet users worldwide as of September 2009, an increase of 18 percent since 2008. The fast-growing number of Internet users also presents huge business opportunities to firms.

Keywords— Website design, user navigation, web mining, web navigation

I. INTRODUCTION

The two predominant paradigms for finding information on the Web are navigation and search [Olston and Chi 2003]. Most Web users typically use a Web browser to navigate a Web site. They start with the home page or a Web page found through a search engine or linked from another Web site, and then follows the hyperlinks they think relevant in the starting page and the subsequent pages, until they have found the desired information in one or more pages. They may also use search facilities provided on the Web site to speed up information searching. For a Web site consisting of a very large number of Web pages and hyperlinks between them, these methods are not sufficient for users to find the desired information effectively and efficiently.

It is still an issue for users to find the desired information on the Web site effectively and efficiently. When users get frustrated in their attempts to find the desired information, it only takes them one click to leave the Web site. Thus how to help users navigate a Web site and find the desired information effectively and efficiently is crucial to the success of the Web site. It is not always easy for users to find the desired information on the Web site effectively and efficiently. When they get frustrated in their attempts to find the desired information, it only takes them one click to leave the Web site. Thus how to help users navigate a Web site and find the desired information effectively and efficiently is crucial to the success of the Web site.

A most important cause of poor website design is that the web developers' perceptive of how a website should be structured and can be considerably different from those of the users. Such differences result in cases where users cannot easily find the desired information in a website. This issue is difficult to handle because when creating a website, web developers may not have a clear under- standing of users' preferences and can only organize pages based on their own ideas. However, the measure of website effectiveness 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 Therefore The problem of improving user navigation on a website with minimal changes to the current structure, is an important issue.

Previous studies on website has 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 [15], [16], and extracting template from WebPages [17]. 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. 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 [18].Second, the 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.

II. LITERATURE SURVEY

Min Chen and Young U. Ryu [1] proposed an approach of mathematical programming model to improve the navigation effect of the website minimizing changes to its current structure. Their model was particularly suitable for informational websites whose contents are relatively stable over time. It improves the performance of website rather than reorganizes and therefore suitable for website maintenance on a progressive basis The Mathematical Programming model was observed to scale up very well, optimally solving large-sized problems in a few seconds in most cases on a desktop PC Perkowitz and Etzioni [02] describe an approach that automatically synthesizes index pages which contain links to pages pertaining to particular topics based on the co-occurrence frequency of pages in user traversals, to facilitate user navigation. However this method is web personalization

The methods proposed by Mobasher et al. [13], [14], [15] and Yan et al. [16] create clusters of users profiles from weblogs and then dynamically generate links for users who are classified into different categories based on their access patterns. These methods are web personalization based.

Nakagawa and Mobasher [13] develop a hybrid personalization system that can dynamically switch between recommendation models based on degree of connectivity and the user's position in the site. For reviews on web personalization approaches, see [18] and [19].

Web transformation, on the other hand, involves changing the structure of a website to facilitate the navigation for a large set of users [28] instead of personalizing pages for individual users. Fu et al. [29] describe an approach to reorganize WebPages 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.

Gupta et al. [19] propose a heuristic method based on simulated annealing to relink WebPages 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 for a small website.

Lin [20] 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.

Lin and Tseng [20] 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 [20] are still relatively small, posing questions on its scalability to large-sized websites.

There are several remarkable differences between web transformation and personalization approaches. First, transformation approaches create or modify the structure of a website used for all users, while personalization approaches dynamically reconstitute pages for individual users. Hence, there is no predefined/built-in web structure for personalization approaches.

In order to understand the preference of individual users, personalization approaches need to collect information associated with these users (known as user profiles). This computationally intensive and time-consuming process is not required for transformation approaches.

Transformation approaches make use of aggregate usage data from weblog files and do not require tracking the past usage for each user while dynamic pages are typically generated based on the users' traversal path. Thus, personalization approaches are more suitable for dynamic websites whose contents are more volatile and transformation approaches are more appropriate for websites that have a built-in structure and store relatively static and stable contents.

Jia-Ching Ying, Chu-Yu Chin, Vincent S. Tseng [23] proposes a special data structure named Ideal-Tree (Inverted-data-base Expectable Tree) to avoid the effort of scanning database. Meanwhile, an efficient mining algorithm named Ideal-Tree-Miner is proposed for mining web navigation patterns with dynamic thresholds. Based on the discovered patterns, they give a navigation prediction model. Mining Web Navigation Patterns with Dynamic Thresholds for Navigation

Dean and Henzinger alsPredictiono proposed another simple algorithm to find relevant pages from page similarities. The page source of this algorithm, however, only consists of the sibling pages of the given page and many important semantically relevant pages might be neglected. And the similarity between a page and the given page is measured by the number of their common parent pages, named cocitation degree. The pages that have higher cocitation degrees with the given page are identified as relevant pages. Although this algorithm is simple and efficient, the deeper relationships

among the pages cannot be revealed. For example, if two or more pages have the same cocitation degree with the given page, this algorithm could not identify which page is more related to the given page.

Reis et al .used a restricted tree-edit distance to cluster documents and, in it is assumed that labelled training data are given for clustering. However, the tree edit distance is expensive and it is not easy to select good training pages. Crescenzi et al.focused on document clustering without template extraction.

They targeted a site consisting of multiple templates. From a seed page, WebPages are crawled by following internal links and the pages are compared by only their link information. However, if web pages are collected without considering their method, pages from various sites are mixed in the collection and their algorithm should repeatedly be executed for each site. Since pages crawled from a site can be different by the objectivity of each crawler, their algorithm may require additional crawling on the fly.

[25] Propose a hierarchical network search engine that clusters hypertext documents to structure given information space for supporting various services like browsing and querying. All hypertext documents in a certain information space (e.g one website) were clustered into a hierarchical form based on contents as well as link structure of each hypertext document. By considering about links within the same website, related documents in the same website could be grouped into one cluster. However, our target is not general situation but search results classification, which clusters search results into more narrow and detailed groups.

There are several remarkable differences between web transformation and personalization approaches. First, transformation approaches create or modify the structure of a website used for all users, while personalization approaches dynamically reconstitute pages for individual users. Hence, there is no predefined/built-in web structure for personalization approaches.

In order to understand the preference of individual users, personalization approaches need to collect information associated with these users (known as user profiles). This computationally intensive and time-consuming process is not required for transformation approaches.

Transformation approaches make use of aggregate usage data from weblog files and do not require tracking the past usage for each user while dynamic pages are typically generated based on the users' traversal path. Thus, personalization approaches are more suitable for dynamic websites whose contents are more volatile and transformation approaches are more appropriate for websites that have a built-in structure and store relatively static and stable contents.

Dean and Henzinger [30] also proposed another simple algorithm to find relevant pages from page similarities. The page source of this algorithm, however, only consists of the sibling pages of the given page and many important semantically relevant pages might be neglected. And the similarity between a page and the given page is measured by the number of their common parent pages, named cocitation degree. The pages that have higher cocitation degrees with the given page are identified as relevant pages. Although this algorithm is simple and efficient, the deeper relationships among the pages cannot be revealed. For example, if two or more pages have the same cocitation degree with the given page, this algorithm could not identify which page is more related to the given page. Crescenzi et al. studied initially the data extraction problem and Yossef and Rajagopalan introduced the template detection problem. Previously, only tags were considered to find templates but Arasu and Garcia-Molina observed that any word can be a part of the template or contents. they detects elements of a template by the frequencies of words. Reis et al. used a restricted tree-edit distance to cluster documents and, in it is assumed that labeled training data are given for clustering. However, the tree edit distance is expensive and it is not easy to select good training pages. Crescenzi et al. focused on document clustering without template extraction. They targeted a site consisting of multiple templates. From a seed page, WebPages are crawled by following internal links and the pages are compared by only their link information. However, if web pages are collected without considering their method, pages from various sites are mixed in the collection and their algorithm should repeatedly be executed for each site. Since pages crawled from a site can be different by the objectivity of each crawler, their algorithm may require additional crawling on the fly. Lerman et al. Proposed systems to identify data records in a document and extract data items from them. Zhai and Liu proposed an algorithm to extract a template using not only structural information, but also visual layout information. Chakrabarti et al. solved this problem by using an isotonic smoothing score assigned by a classifier. For XML documents, Garofalakis et al. solved the problem of DTD extraction from multiple XML documents. While HTML documents are semi structured, XML documents are well structured, and all the tags are always a part of a template. The solutions for XML documents fully utilize these properties. In the problem of the template extraction from heterogeneous document, how to partition given documents into homogeneous subsets is important.

III. IMPLICATION OF THIS SURVEY

This survey contributes to the literature on improving web user navigation by examining issue from a new and important angle. As time passes and the need for information changes, websites also need to be regularly maintained and improved. However, the current literature focuses on how to restructure a website. Webmasters need to carefully balance

the tradeoff between desired improvements in the user navigation and the changes needed to accomplish the task when selecting appropriate path thresholds. This is particularly important when a website is improved on a regular basis.

IV. CONCLUSIONS

In this paper we have presented a survey of research papers analysing different techniques for Web mining, the application of data mining and Web transformation techniques. A change in any of the constituent technique for Web Mining accounts for a change in paradigm consequently affecting the approach of web analyser to examine the data for identifying the navigation pattern. So a good understanding of the data preparation technique

REFERENCES

[1] Min Chen and Young U. Ryu Facilitating Effective User Navigation Through Website Structure Improvement" IEEE Transaction on knowledge and Data Engineering, Vol. 25, no. 3, March2013
[2] M. Perkowitz and O. Etzioni, "Towards Adaptive Web Sites: Conceptual Framework and Case Study," Artificial Intelligence, vol. 118, pp. 245-275, 2000.
[3] J. Lazar, User-Centered Web Development. Jones and Bartlett Publishers, 2001.
[4] 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.
[5] 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
[6] H. Kao, J. Ho, and M. Chen, "WISDOM: Web Intrapage Informative Structure Mining Based on Document Object Model," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 5, pp. 614-627, May 2005.
[7] H. Kao, S. Lin, J. Ho, and M. Chen, "Mining Web Informative Structures and Contents Based on Entropy Analysis," IEEE Trans.Knowledge and Data Eng., vol. 16, no. 1, pp. 41-55, Jan. 2004.
[8] C. Kim and K. Shim, "TEXT: Automatic Template Extraction from Heterogeneous Web Pages," IEEE Trans. Knowledge and Data Eng., vol. 23, no. 4, pp. 612-626, Apr. 2011.
[9] V. McKinney, K. Yoon, and F. Zahedi, "The Measurement of Web-Customer Satisfaction: An Expectation and Disconfirmation Approach," Information Systems Research, vol. 13, no. 3, pp. 296- 315, 2002
[10] M. Kilfoil et al., "Toward an Adaptive Web: The State of the Art and Science," Proc. Comm. Network and Services Research Conf.,pp. 119-130, 2003.
[11] R. Gupta, A. Bagchi, and S. Sarkar, "Improving Linkage of Web Pages," INFORMS J. Computing, vol. 19, no. 1, pp. 127-136, 2007
[12] 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.
[13] M. Eirinaki and M. Vazirgiannis, "Web Mining for Web Personalization," ACM Trans. Internet Technology, vol. 3, no. 1,pp. 1-27, 2003.
[14] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa, "Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization,"Data Mining and Knowledge Discovery, vol. 6, no. 1, pp. 61-82, 2002.
[15] B. Mobasher, R. Cooley, and J. Srivastava, "Automatic Personalization Based on Web Usage Mining," Comm. ACM, vol. 43, no. 8,pp. 142-151, 2000.
[16] B. Mobasher, R. Cooley, and J. Srivastava, "Creating Adaptive Web Sites through Usage-Based Clustering of URLs," Proc. Workshop Knowledge and Data Eng. Exchange, 1999.
[17] W. Yan, M. Jacobsen, H. Garcia-Molina, and U. Dayal, "From User Access Patterns to Dynamic Hypertext Linking," Computer Networks and ISDN Systems, vol. 28, nos. 7-11, pp. 1007-1014, May 1996.
[18] M. Nakagawa and B. Mobasher, "A Hybrid Web Personalization Model Based on Site Connectivity," Proc. Web Knowledge Discovery Data Mining Workshop, pp. 59-70, 2003.
[19] M. Eirinaki and M. Vazirgiannis, "Web Mining for Web Personalization," ACM Trans. Internet Technology, vol. 3, no. 1,pp. 1-27, 2003.
[20] B. Mobasher, "Data Mining for Personalization," The Adaptive Web: Methods and Strategies of Web Personalization, A. Kobsa, W.Nejdl, P. Brusilovsky, eds., vol. 4321, pp. 90-135, Springer-Verlag,2007.
[21] .C. Lin and L. Tseng, "Website Reorganization Using an Ant Colony System," Expert Systems with Applications, vol. 37, no. 12,pp. 7598-7605, 2010.
[22] Y. Fu, M.Y. Shih, M. Creado, and C. Ju, "Reorganizing Web Sites Based on User Access Patterns," Intelligent Systems in Accounting,Finance and Management, vol. 11, no. 1, pp. 39-53, 2002.
[23] Jia-Ching Ying, Chu-Yu Chin,Vincent S. Tseng "Mining Web Navigation Patterns with Dynamic Thresholds for Navigation
[24] Prediction"Granular Computing (GrC), 2012 IEEE International Conference 2012
[25] Ron Weiss et. al. 96 Hypursuit: A Hierarchical Network Search Engine that Exploits Content-Link Hypertext Clustering. AC M Conference on Hypertext, Washington USA,1996