International Journal of Research Studies in Science, Engineering and Technology
Volume 1, Issue 9, December 2014, PP 110-117
ISSN 2349-4751 (Print) & ISSN 2349-476X (Online)

Speed Up Effective User Navigation through Website Structure Improvement

[1.] Vaseem Basha Shaik, [2.] Shakeel Ahamed Pathan

[1.] (M.Tech), Department of CSE, QUBA College of Engineering and Technology, Nellore, India
[2.] Associate Professor, Department of CSE, QUBA College of Engineering and Technology, Nellore, India

Abstract: *Designing well-structured websites to facilitate effective user navigation has long been a test. A main reason is that the web developers' understanding of how a website should be structured can be considerably different from that of the users. While various methods have been projected to relink web pages to improve navigability using user navigation data, the totally reorganized new structure can be extremely 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 wide-ranging 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 successfully solved. 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 navigation on the improved structure is indeed greatly enhanced. More interestingly, we find that heavily unsettled users are more likely to benefit from the improved structure than the less disoriented users.*

1. INTRODUCTION

The approach of the Internet has given a remarkable stage to individuals to secure information and investigate data. There are 1.73 billion Internet clients worldwide as of September 2009, an increment of 18 percent since 2008. The quickly developing number of Internet clients additionally introduces immense business chances to firms. As indicated by Grau, the US retail e-trade deals (barring travel) totaled $127.7 billion in 2007 and will reach $218.4 billion by 2012. So as to fulfill the expanding requests from online clients, firms are intensely putting resources into the advancement and upkeep of their sites. Internet retailer reports that the general site operations using expanded in 2007, with one-third of webpage administrators trekking using by no less than 11 percent, contrasted with that in 2006. An essential driver of poor site configuration is that the web designers' understanding of how a site ought to be organized can be respectably not quite the same as those of the clients. Such contrasts bring about situations where clients can't without much of a stretch find the coveted data in a site. This issue is hard to dodge on the grounds that when making a site, web engineers might not have a reasonable understanding of clients' inclination and can just compose pages focused around their own particular judgments. Be that as it may, the measure of site viability ought to be the fulfillment of the clients instead of that of the engineers. Subsequently, Webpages ought to be composed in a manner that by and large matches the client's model of how pages ought to be sorted out. Past studies on site has concentrated on a mixture of issues, for example, comprehension web structures, discovering applicable pages of a given page, mining educational structure of a news site, and concentrating layout from pages. Our work, then again, is nearly identified with the writing that inspects how to enhance site safety through the utilization of client route information. Different works have attempted to address this inquiry and they can be for the most part characterized into two classifications: to encourage a specific client by rapidly reconstituting pages focused around his profile and traversal ways, regularly alluded as personalization, and to alter the site structure to facilitate the route for all clients, frequently alluded as change.

In this paper, we are concerned basically with change approaches. The writing considering changes approaches essentially concentrates on creating routines to totally reorganize the connection structure of a site. In spite of the fact that there are supporters for site revamping methodologies, their

downsides are self-evident. To start with, since a complete rearrangement could profoundly change the area of recognizable things, the new site may perplex clients. Second, the redesigned site structure is very eccentric, and the expense of confusing clients after the progressions stays unanalyzed. This is on account of a site's structure is commonly composed by masters and bears business or hierarchical rationale, however this rationale might no more exist in the new structure when the site is totally rearranged. Furthermore, no earlier studies have surveyed the ease of use of a totally rearranged site, prompting questions on the pertinence of the revamping methodologies. At long last, since site rearrangement methodologies could drastically change the current structure, they can't be often times performed to enhance the traversability. Perceiving the downsides of site rearrangement approaches, we address the inquiry of how to enhance the structure of a site instead of revamp it significantly. Particularly, we create a mathematical programming (MP) demonstrate that encourages client route on a site with negligible progressions to its present structure. Our model is especially proper for instructive sites whose substance are static and generally steady about whether. Cases of associations that have instructive sites are colleges, vacation destinations, clinics, government orgs, and sports associations. Our model, in any case, may not be suitable for sites that simply utilize element pages or have unstable substance. This is on account of a consistent state may never be arrived at in client access designs in such sites, so it may not be conceivable to utilize the weblog information to enhance the website structure.

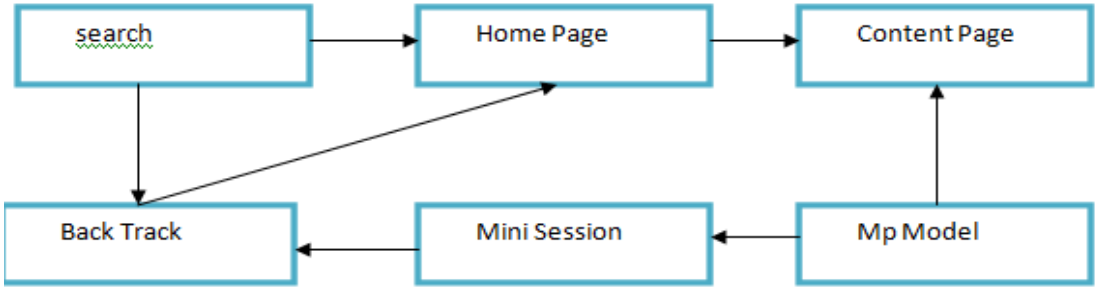
Existing System: A primary cause of poor website design is that the web developers' understanding of how a website should be structured can be considerably different from those of the users. Such differences result in cases where users cannot easily locate the desired information in a website. This problem 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.

Proposed System: I 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. 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 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.

2. METHODOLOGY

Problem Definition: An essential driver of poor site outline is that the web engineers' understanding of how a site ought to be organized can be extensively not the same as those of the clients. Such problems bring about situations where clients can't undoubtedly spot the sought data in a site. This issue is hard to dodge on the grounds that when making a site, web engineers might not have an agreeable understanding of clients' inclination and can just arrange pages focused around their judgments. Notwithstanding, the measure of site adequacy ought to be the fulfillment of the clients instead of that of the designers. Therefore, Web pages ought to be sorted out in a manner that for the most part matches the client's model of how pages ought to be composed. Results from broad tests directed on a freely accessible true information set demonstrate that our model not just altogether enhances the client route with not very many progressions, additionally can be adequately illuminated. What's more, we characterize two assessment measurements and use them to evaluate the execution of the enhanced site utilizing the true information set. Assessment results affirm that the client route on the enhanced structure is for sure enormously improved. All the more interestingly, we find that intensely perplexed clients are more inclined to profit from the enhanced structure than the less confused clients.

Architecture



Architecture will says clearly about the flow of navigation among the pages or modules. By straight forwardly, search engine will have two ways in the architecture. One that is homepage of our project and another is history of our link that I was used in the past. Here we are implementing mini sessions for tracking the sessions and sessions contains the sequence of links and those aren't looped .After home page will navigate the Content page and Content page consists of many links in one web page. Link are choosing based on our wish and we will have a cross checking slot i.e., MP model .MP model plays vital role to navigate the links that consists of necessary information.MP model gives the results of most hitting URLs and related URLs. Finally, we are tracking the all information by implementing Mini sessions.

Modules:

- ✓ Web Personalization.
- ✓ Web Transformation.
- ✓ Maximal Forward Reference.
- ✓ Mini Sessions.
- ✓ Out-Degree Threshold.

Web Personalization: Web personalization is the methodology of "Tailoring" pages to the needs of particular clients utilizing the data of the clients' navigational conduct and profile information. Perkowitz and Etzioni depict an approach that naturally blends list pages which contain connections to pages relating to specific themes focused around the co-event recurrence of pages in client traversals, to encourage client route. The strategies proposed by Mobasher et al. what's more Yan et al. make groups of clients profiles from weblogs and afterward alterably produce joins for clients who are arranged into distinctive classifications focused around their right to gain entrance designs.

Web Transformation: Web transformation, then again, includes changing the structure of a site to encourage the route for a huge set of clients as opposed to customizing pages for individual clients. Fu et al. depict a methodology to redesign site pages in order to give clients their sought data in less clicks. Then again, this methodology considers just nearby structures in a site as opposed to the site all in all, so the new structure may not be fundamentally ideal. Gupta et al. [19] propose a heuristic strategy focused around reproduced tempering to relink website pages to enhance safety. This strategy makes utilization of the total client inclination information and can be utilized to enhance the connection structure in sites for both wired and remote gadgets.

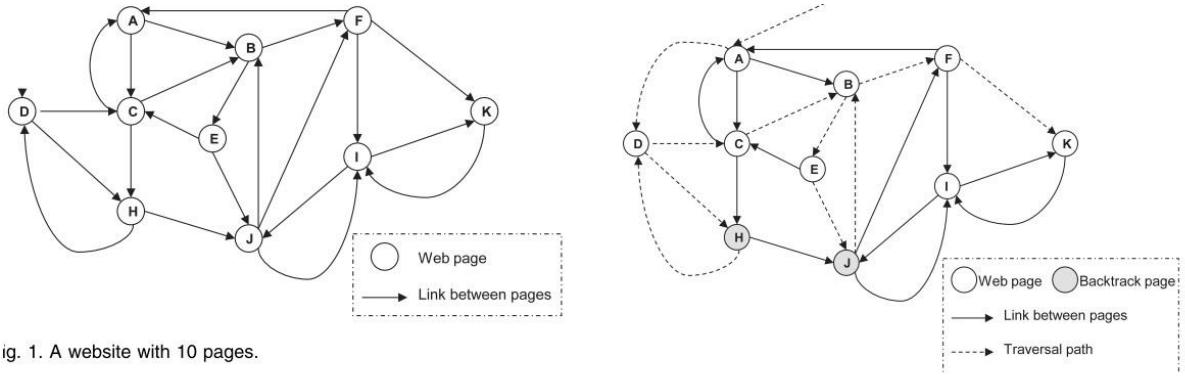
Maximal Forward Reference: We utilize backtracks to recognize the ways that a client has crossed, where a backtrack is characterized as a client's return to an awhile ago skimmed page. The instinct is that clients will backtrack in the event that they don't discover the page where they expect it . Hence, a way is characterized as a succession of pages went to by a client without backtracking, an idea that is like the maximal forward reference characterized in Chen et al. Basically, each one backtracking point is the end of a way. Consequently, the more ways a client has crossed to achieve the focus on, the more discrepant the site structure is from the client's desire.

Mini Sessions: Review that a scaled down session is important just in the event that its length is bigger than the comparing way edge. Therefore, just significant smaller than normal sessions need to be considered for development and this prompts countless small sessions (meant as TI) being wiped out from thought in our MP model.

OUT-DEGREE Threshold: web pages can be by and large characterized into two classes : record pages and substance pages. A file page is intended to help clients better explore and could incorporate numerous connections, while a substance page contains data clients are intrigued by and ought not have numerous connections. Therefore, the out-degree edge for a page is exceedingly reliant on the reason for the page and the site. Normally, the out degree edge for record pages ought to be bigger than that for substance pages.

3. METRIC FOR EVALUATING NAVIGATION EFFECTIVENESS

The Metric: The 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 navigation effectiveness. Marsico and Levialdi point out that information becomes useful only when it is presented in a way consistent with the target users' expectation. Palmer 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 i.e., some specific information they are seeking, we measure this discrepancy by the number of b times a user has attempted before locating the target. Our metric is related to the notion of information scent developed in the context of information foraging theory Information foraging theory models the cost structure of human information gathering using the analogy of animals foraging for food and is a widely accepted theory for addressing the information seeking process on the web . Information scent refers to proximal cues (e.g., the snippets of text and graphics of links) that allow users to estimate the location of the "distal" target information and determine an appropriate path . Users are faced with a decision point at each page; they use information scent to evaluate the likely effort and the probability of reaching their targets via each link and make navigation decisions accordingly. Consequently, a user is assumed to follow the path that appears most likely to lead him to the target. This suggests that a user may backtrack to an already visited page to traverse a new path if he could not locate the target page in the current path. Therefore, we use the number of paths a user has traversed to reach the target as an proximate measure to the number of times the user has attempted to locate one target. We use backtracks to identify the paths that a user has traversed, where a backtrack is defined as a user's revisit to a previously browsed page. The intuition is that users will backtrack if they do not find the page where they expect it . Thus, a path is defined as a sequence of pages visited by a user without backtracking, a concept that is similar to the maximal forward reference defined in Chen et al. Essentially, each backtracking point is the end of a path. Hence, the more paths a user has traversed to reach the target, the more discrepant the site structure is from the user's expectation.



ig. 1. A website with 10 pages.

4. MATHEMATICAL PROGRAMMING

In this section we will discuss mathematical model of proposed system: Website is represented as a directed graph. Let nodes representing pages and arc representing links. Let n be the number of pages of the website. User may visit more than one page by accessing links during user session. Let Pi={P1,P2,…..,Pm} be the set of m link accessed by user Ui where i={1,2,….,m}. Let Pm be the page user Ui is looking for. Means Pm is the target page. Let U= {U1, U2 ,…,Un} be the set of N users. Let C= {C1, C2 ,…., Cp} be set of p candidate links that need to redesign and relink. We use

Dice's coefficient index. Dice's coefficient index is used for checking similarity between two string set.

$$D(A, B) = \frac{2 * |A \quad B|}{|A| \ |B|}$$

The aim of this paper is to identify links that are used to redesign. Usage pattern is used to analyze user's behavior on the Web. By analyzing user's behaviour we find out target page. And find out candidate links so that user can access target page faster.

Algorithm:-

Mining Candidate Link Algorithm

Input: Pi – Users Profile data

Output: Links that can be use for redesign

Steps-

1: We identify the usage pattern of users from Pi = {P1, P2 ,…., Pm} set for user Ui to get link Pm

2: For every access link set obtain the set of candidate links {C1, C2,….,Cp}

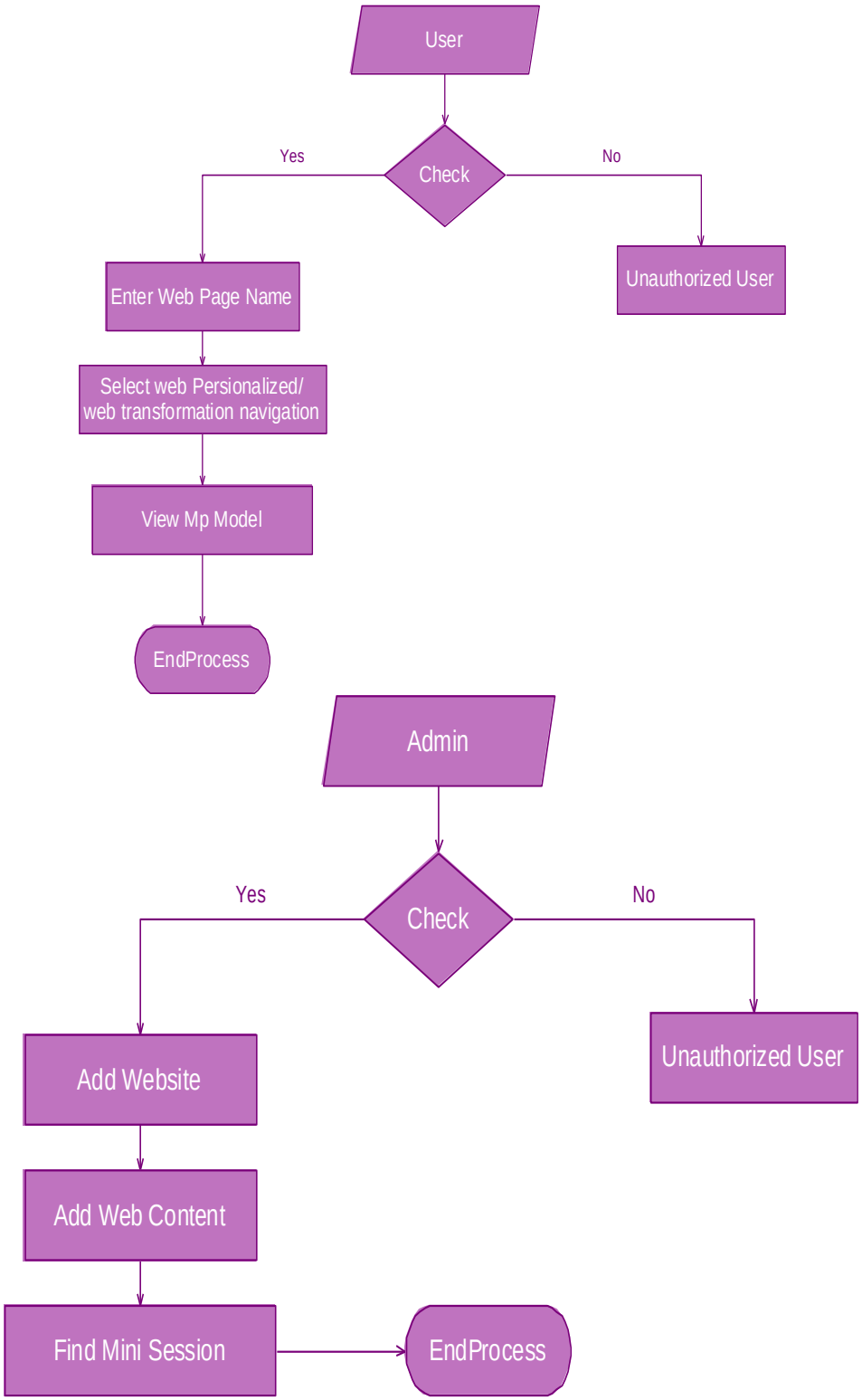
3: For all users and their all access link set obtain the set of candidate links.

4: Apply KNN classifier.

5: Then the links having problem for maximum number of users are selected for redesign the website

Evaluation Procedure: We used simulations to approximate the real usage and to evaluate how the user navigation could be enhanced in 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 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 Webpages. 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. Because of the assumption made for the new and improved links, the claimed benefit can be interpreted as the upper bound and optimal benefit of our model. However, we would like to claim that improved and newly added links could guide users to find their target pages more efficiently to some extent. This is because: 1) our method establishes efficient paths to target pages that were not available in the website structure before optimization, and 2) our method suggests improving links that would lead to users' target pages efficiently but missed by users (since they did not know what these links would lead to), so that more efficient navigation can be facilitated. Since our evaluation is simulation based, a usability study involving real users may help strengthen the results of our study and deserves further investigation. However, we note that such usability studies are generally more expensive and time consuming in the context of website evaluation [60], and hence are usually conducted on small sized websites [5]. In contrast, simulation can be easily implemented, quickly performed for various parameter settings, and tested on a large scale. Thus, the simulation 586 IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 3, MARCH 2013 studies in our paper complement usability studies by offering its own distinct advantages.

System Design:



5. CONCLUSION

In this paper, we have proposed a mathematical programming model to enhance the route adequacy of a site while minimizing progressions to its present structure, a basic issue that has not been analyzed in the writing. Our model is especially suitable for educational sites whose substance is moderately steady about whether. It enhances a site as opposed to rearranges it and subsequently is suitable for site upkeep on a dynamic premise. The tests on a true site demonstrated that our model could give huge enhancements to client route by including just few new connections. Ideal arrangements were immediately acquired, proposing that the model is exceptionally successful to true sites. Also, we

have tried the MP model with various engineered information sets that are much bigger than the biggest information set considered in related studies and the true information set. The MP model was seen to scale up extremely well, ideally tackling expansive measured issues shortly much of the time on a desktop PC. Our results affirmed that the enhanced structures for sure significantly encouraged client route. Furthermore, we discovered an engaging come about that vigorously perplexed clients, i.e., those with a higher likelihood to desert the site, are more inclined to profit from the enhanced structure than the less confused clients. Trial comes about additionally uncovered that while utilizing little way limits could bring about better conclusions, it would likewise include essentially all the more new connections. Hence, Webmasters need to deliberately adjust the trade off between fancied enhancements to the client route and the quantity of new connections required to perform the undertaking when selecting fitting way edges. Since no earlier study has analyzed the same destination as our own, we contrasted our model and a heuristic. The correlation demonstrated that our model could accomplish practically identical or preferred enhancements over the heuristic with extensively less new connections.

REFERENCES

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

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

[3] Y. Yang, Y. Cao, Z. Nie, J. Zhou, and J. Wen, "Closing the Loop in Webpage Understanding," IEEE Trans. Knowledge and Data Eng., May 2010, vol. 22, no. 5, pp. 639-650.

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

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

[6] 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. , Jan. 2004, vol. 16, no. 1, pp. 41-55.

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

[8] M.Kilfoil et al., "Toward an Adaptive Web: The State of the Art and Science," Proc. Comm. Network and Services Research Conf., 2003, pp. 119-130.

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

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

[11] M. Eirinaki and M. Vazirgiannis, "Web Mining for Web Personalization," ACM Trans. Internet Technology, 2003, vol. 3, no. 1, pp. 1-27.

[12] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa, "Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization," Data Mining and Knowledge Discovery, 2002, vol. 6, no. 1, pp. 61-82.

[13] B. Mobasher, R. Cooley, and J. Srivastava, "Automatic Persona- lization Based on Web Usage Mining," Comm. ACM, 2000, vol. 43, no. 8, pp. 142-151.

[14] 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.

[15] W. Yan, M. Jacobsen, H. Garcia-Molina, and U. Dayal, "From User Access Patterns to Dynamic Hypertext Linking," Computer Net- works and ISDN Systems, May 1996, vol. 28, nos. 7-11, pp. 1007-1014.

[16] Jiawei Han and Micheline Kamber , Data Mining, Concepts and Techniques, Elsevier, II Edition, 2008

[17] By Grady Booch , James Rumbaugh, Ivar Jacobson The Unified Modeling Language User Guide 2nd Edition, Pearson Education

[18] Programming world wide web-Sebesta, PearsonCore servlets and java server pages Volume 1: core Technologies, Marty Hall and Larry Brown Pearson

[19] Software Engineering, A practitioner's Approach- Roger S. Pressman, 7th edition. McGrawHill International Edition.

[20] Effective Methods for Software Testing, 2nd Edition, William E. Perry, Second Edition, Wiley India, 2006