

# Survey on Methods for Enhancing Web Navigation

*Adhokshaj Haridas*<sup>1</sup>, *Prof. Vinod Nayyar*<sup>2</sup>

<sup>1</sup> Dept. of M.Tech CSE  
Abha Gaikwad-Patil College Of Engineering,  
Nagpur, Maharashtra, India  
adhokshaj66@gmail.com

<sup>2</sup> Astit. Professor in Dept. of M Tech CSE  
Abha Gaikwad-Patil College Of Engineering  
Nagpur, Maharashtra, India  
vinodnayyar5@gmail.com

**Abstract:** *Using the web usage mining effectively as an underlying mechanism for web personalization and recommender system can enable us to achieve efficient web navigation system. Using the user usage patterns as a tool of prior information we can very well keep an intelligent base of most common user choices. Also by using the effective web structuring methods and web usage mining we will have a system that will ensure a completely different web navigation experience. Specifically, the proposed changes to the existing website will improve the user navigation on a website while minimizing alterations to its current structure. The users who are more affected by inefficiency of website are likely to benefit more from the improved structure.*

**Keywords:** Web usage mining, Web navigation, Web structure.

## 1. Introduction

In today's day and age it is imperative to go online for variety of reasons as all the things that an individual would want to achieve has to have online solution. As such is the present day situation software industry, specially the web developers are needed to work extensively on such a need of the everyday internet user. For some users using the net is part and parcel of their job and they cannot do without websites that are of their interest. E-commerce websites are the best example of this category since all the business of the company is wholly and solely dependent on their website performance.

When we consider such commercial firms that have their everyday working heavily dependent on the basis of their website performance we got to admit to the fact that the best possible web structuring, easy to use website design, user friendly user interface(UI) would be their primary need. Having understood such a basic requirements for such a particular firm or company we got to also consider the fact that what is good for a website developer may not be good enough for a user who is using such a website for the first time. Web developers generally develop a website using their own strategy and conviction to do the job assigned with utmost professionalism. They more often end up making a very useful website but this alone in any way may not be the only reason for a commercial website to be a huge hit amongst customers.

A primary cause of poor website design is that the web developer's 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 user preference 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 developer's. Thus, Web pages should be organized in a way that generally matches the user's model of how pages should be organized.

Previous studies on website has focused on a variety of issues, such as understanding web structures, finding relevant pages of a given page, mining informative structure of a news website and extracting template from web pages. Our work, examines how to improve website navigability through the use of user navigation data. It they can be generally classified into two categories: 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.

## 2. Literature Survey

1) **An empirical study of web site navigation structure's impact on web site usability.**

Web sites are intended to facilitate knowledge acquisition, often in the interest of supporting decision making. Based on taxonomy of factors influencing Web site usability, hypotheses are developed about usability of alternative navigation structures. These hypotheses are tested via experiments that measure user performance in accomplishing knowledge acquisition tasks and user perceptions of usability. Two rounds of experimentation are performed for both simple and relatively complex task sets. Results show that a usage-oriented hierarchy or a combined hierarchy is a navigation structure associated with significantly higher usability than subject-oriented hierarchies, for both simple and relatively complex knowledge acquisition tasks.

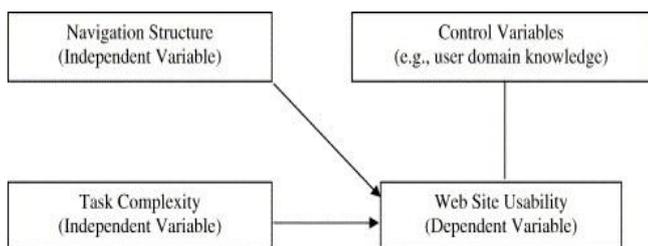


Figure 1. Research Model

## 2) Knowledge Discovery from Web Logs

Web usage mining is obtaining the interesting and constructive knowledge and implicit information from activities related to the World Wide Web. Web servers trace and gather information about user interactions every time the user requests for particular resources.

Evaluating the Web access logs would help in predicting the user behavior and also aid in formulating the web structure. Based on the applications point of view, information extracted from the web usage patterns possibly directly applied to competently manage activities related to e-business, e-services, e-education, on-line communities and so on. On the other hand, since the size and density of the data grows rapidly, the information provided by existing web log file analysis tools may possibly facilitate insufficient information and hence more intelligent mining techniques are required.

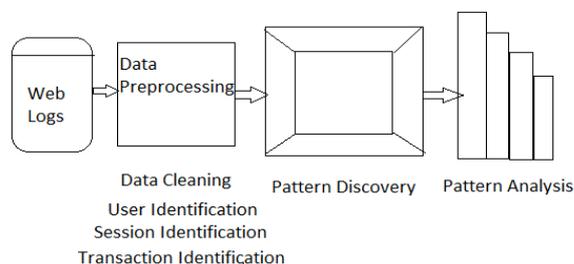


Figure 2. Phases of Web Usage Mining

Web log files play a significant role in the Mining for web usage. The knowledge that can be obtained from web log files is the user's navigation pattern. The difficulty in obtaining such knowledge is that the users are constantly shifting their focus and different users have different navigational behavior with different needs associated with them. The navigation pattern

knowledge can be used to help users by predicting their future request and it will help on the personalization of websites.

## 3) Research Framework for Consumer Satisfaction with Internet Shopping

E-shopping lends a great convenience to shoppers, yet its electronic format changes information-gathering methods traditionally used by customers. This change raises questions concerning customer satisfaction with the online purchasing process. e-shopping involves a number of phases, including the information phase, in which customers search for information regarding their intended purchases. This concept helps in understanding e-shopping and develops theoretically justifiable constructs for measuring Web-customer satisfaction during the information phase.

By synthesizing the expectation-disconfirmation paradigm, we separate Web site quality into information quality (IQ) and system quality (SQ). The measurements for these constructs are developed and tested in a two-phase study. In the first phase, the IQ and SQ dimensions are identified, and instruments for measuring them are developed and tested. In the second phase, using the salient dimensions of Web-IQ and Web-SQ as the basis for formulating first-order factors, we develop and empirically test instruments for measuring IQ and SQ-satisfaction. The analysis of the measurement model indicates that the proposed metrics have a relatively high degree of validity and reliability.

## 4) Towards adaptive Web sites : Conceptual framework and case study

Today's Web sites are intricate but not intelligent; while Web navigation is dynamic and idiosyncratic, all too often Web sites are fossils cast in HTML. In response, this technical idea investigates adaptive Web sites: sites that automatically improve their organization and presentation by learning from visitor access patterns. Adaptive Web sites mine the data buried in web server logs to produce more easily navigable web sites.

To demonstrate the feasibility of adaptive Web sites, the paper considers the problem of index page synthesis and sketches a solution that relies on novel clustering and conceptual clustering techniques. Our preliminary experiments show that high-quality candidate index pages can be generated automatically, and that our techniques outperform existing methods (including the A priori algorithm, K-means clustering, hierarchical agglomerative clustering, and COBWEB) in this domain. Five desiderata for an adaptive web site

(a) Avoid creating work for visitors (e.g., filling out questionnaires). Visitors to a site are turned off by extra work, especially if it has no clear reward, and may opt out rather than participate. If the site cannot improve itself without feedback, it will fail if users do not assist.

(b) Make the Web site easier to use for everyone, including first-time users, casual users, etc. Customization can be genuinely useful for repeat visitors, but does not benefit first-time users. In addition, one user's customizations do not apply to other users; there is no sharing or aggregation of information

across multiple users. Transformation has the potential to overcome both limitations.

(c) Minimize additional work for the webmaster. Although human-authored meta information (e.g., XML annotations) may facilitate sites ability to adapt, we should weigh the benefits of this additional information against the cost of authoring it.

(d) Protect the site's original design from destructive changes. For safety, we limit ourselves to nondestructive transformations: changes to the site that leave existing structure intact. We may add links but not remove them, create pages but not destroy them, add new structures but not scramble existing ones.

(e) Keep the human webmaster in control. Clearly, the human webmaster needs to remain in control of the Web site in the foreseeable future both to gain her trust in automatic adaptive techniques and to avoid "disasters".

### 5) Closing the Loop in Webpage Understanding

A novel framework called WebNLP, which enables bidirectional integration of page structure understanding and text understanding in an iterative manner, fine-grained and informative blocks is presented here. The intra-page informative structures of pages in a news Web site contain only anchors linking to news pages or bodies of news articles. An intra-page informative structure mining system called WISDOM (Web Intra-page Informative Structure Mining based on the Document Object Model) which applies Information Theory to DOM tree knowledge in order to build the structure is proposed. WISDOM splits a DOM tree into many small sub-trees and applies a top-down informative block searching algorithm to select a set of candidate informative blocks. The structure is built by expanding the set using proposed merging methods.

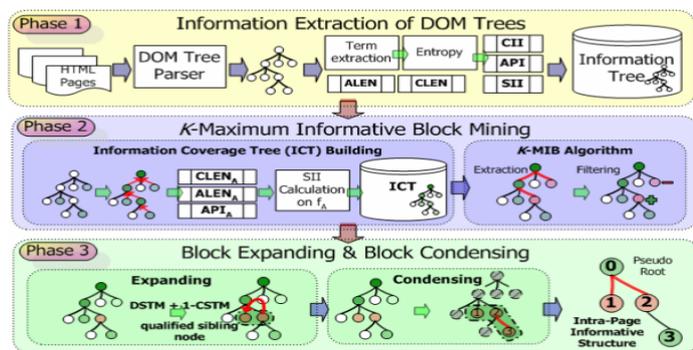


Figure 3. WISDOM system flow

WISDOM automatically extracts and recognizes ISs of each page in a Web site according to the knowledge in the tree structures of pages. As shown in Figure 3, WISDOM consists of three phases: (1) information extraction from DOM trees, (2) k-maximum informative block mining, and (3) block expansion and condensation. In the first phase, we extract useful features from the information of the original DOM tree.

These features can be classified into two types of information: node information and structure information. In the second phase, we aggregate the node information to build the Information Coverage Tree (ICT). According to the ICT, we devise a greedy algorithm, i.e., k-maximum informative block mining algorithm (k-MIB), to extract sub-trees that contain richer information. The extracted sub-trees are either better information hubs or better information authorities, depending

on the criteria employed in the greedy algorithm. They form the skeleton set of the IS of a page. Then expanding the skeleton set by assembling neighboring sub-trees that contain similar features corresponding to the original skeleton sub-trees. After condensing the expanded set by removing dummy nodes, the assembled forest (or tree), in essence the IS of a page, is constructed.

### 6) Improving Linkage of Web Pages

Organizations maintain informational websites for wired devices. The information content of such websites tends to change slowly with time, so a steady pattern of usage is soon established. User preferences, both at the individual and at the aggregate level, can then be gauged from user access log files. A heuristic scheme based on simulated annealing that makes use of the aggregate user preference data to re-link the pages to improve navigability is proposed in this paper. This scheme is also applicable to the initial design of websites for wireless devices.

Using the aggregate user preference data obtained from a parallel wired website, and given an upper bound on the number of links per page, our methodology links the pages in the wireless website in a manner that is likely to enable the "typical" wireless user to navigate the site efficiently.

### 7) Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization

Web usage mining, possibly used along with standard approaches to personalization such as collaborative filtering, can help address some of the shortcomings of these techniques, including reliance on subjective user ratings, lack of scalability, and poor performance in the face of high-dimensional and sparse data. However, the discovery of patterns from usage data by itself is not sufficient for performing the personalization tasks. The critical step is the effective derivation of good quality and useful (i.e., actionable) "aggregate usage profiles" from these patterns. This idea presents and experimentally evaluate two techniques, based on clustering of user transactions and clustering of page views, in order to discover overlapping aggregate profiles that can be effectively used by recommender systems for real-time Web personalization. The evaluation is both in terms of the quality of the individual profiles generated, as well as in the context of providing recommendations as an integrated part of a personalization

engine.

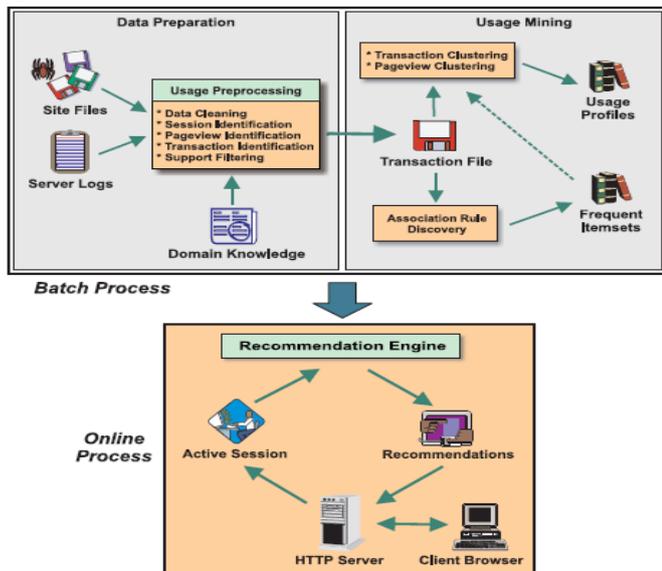


Figure 4 A general framework for personalization based on web usage mining

A general framework for personalization based on aggregate usage profiles is shown in Figure 4. The above framework distinguishes between the offline tasks of data preparation and usage mining, and the online personalization components. The data preparation tasks result in aggregate structures such as a user transaction file capturing meaningful semantic units of user activity to be used in the mining stage. Given the preprocessed data, a variety of data mining tasks can be performed. It is focused on specific clustering techniques for the discovery of aggregate usage profiles. In the online component of the system, the Web server keeps track of the active server session as the user's browser makes HTTP requests. In this stage, the data cleaning tasks involve the removal of erroneous or redundant references, as well as the detection and removal of robot navigation references. Page view identification is the task of determining which page file accesses contribute to a single browser display, and is heavily dependent on the intra-page structure, and hence requires detailed site structure information. Only relevant page views are included in the transaction file.

It describes an approach for automatically classifying visitors of a web site according to their access patterns. User access logs are examined to discover clusters of users that exhibit similar information needs; e.g., users that access similar pages. This may result in a better understanding of how users visit the site, and lead to an improved organization of the hypertext documents for navigational convenience.

### 8) Mining Web Logs to Improve Website Organization

Many websites have a hierarchical organization of content. This organization may be quite different from the organization expected by visitors to the website. In particular, it is often unclear where a specific document is located. Here we propose an algorithm to automatically find pages in a website whose location is different from where visitors expect to find them.

The key insight is that visitors will backtrack if they do not find the information where they expect it: the point from where they backtrack is the expected location for the page. The

algorithm for discovering such expected locations that can handle page caching by the browser are presented. Expected locations with a significant number of hits are then presented to the website administrator. Also presented algorithms for selecting expected locations (for adding navigation links) to optimize the benefit to the website or the visitor.

Three approaches are considered here for recommending additional links to the web site administrator (or automatically adding links):

1. First Only: Recommend all the pages whose frequency of occurrence in the first expected location is above an administrator-specified threshold.
2. Optimize Benefit: Recommend the set of pages that optimize benefit to the website, where benefit is estimated based on the fraction of people who might give up on not finding a page.
3. Optimize Time: Recommend the set of pages that minimize the number of times the visitor has to backtrack, i.e., the number of times the visitor does not find the page in an expected location.

### 3. Challenges

The preprocessed web log data is being used by the web usage mining algorithms. Web server serves the log files as and when needed. But certain reasons due to which the actual logs are not collected are:

- A) The cache present on client browser, majority of the request, if it is present in the cache are not sent to web server.
- B) Most often the user does not visit the home page of a website. Users directly navigate to a particular page, by getting the URL from search engines. This reduces the hit count of index page.
- C) In general the web pages designed by server side scripting languages like PHP, JSP or ASP.NET use inner page. That is, one page consisting of more than one page. In such a case the request for main page records two entries in access log. Identifying the inner page becomes a difficult task .
- D) In some websites, web pages take query string as argument to the URL. It is cumbersome to count the page access of the web page without the argument. In web usage mining the pattern extraction algorithms are applied on the log data after they are processed. So, preprocessing is very much important and must be carried out carefully.

While preprocessing the web access log the above points need to be critically taken into account so that it will produce a good set of access logs for pattern extraction. These are the few challenges that arise amongst other few which needs to be addressed for improving the website performance.

### 4. A different approach to the same problem - Proposed solution

A mathematical programming (MP) model to improve the navigation efficiency of a website while reducing changes to its original. The following contributions would hopefully better

the overall performance of the websites in general. Firstly, the problem of improving user navigation on a website with minimal changes to its current structure.

Mathematical programming (MP) model which not only successfully, accomplishes the task but also generates the optimal solutions surprisingly fast. Secondly, we model the out-degree as a cost term in the objective function instead of as hard constraints. This allows a page to have more links than the out-degree threshold if the cost is reasonable and hence offers a good balance between minimizing alterations to a website and reducing information overload to users. Thirdly, we propose two evaluation metrics and use them to assess the improved structure to confirm the validity of our model.

The evaluation procedure that, we propose to develop will provide a framework for evaluating website structures in an efficient manner.

## 5. Conclusion

Companies, organizations and individuals alike are increasingly gathering information for their best interest. This is done by way of, efficient web mining function which along with effective data extraction algorithm enhances the overall website performance. There is a need to everyday improve the efficiency and the performance of a website in general. New methods need to be adopted and integrated within the framework of any website in particular to enhance its overall response. In this survey we have studied such various methods that can be implemented for enhancing the website efficiency we also present a more effective way forward in this regard through this paper.

## References

- [1] Jennifer English, Marti Hearst, Rashmi Sinha, Kirsten Swearington, and Ping Yee “Examining the Usability of Web Site Search” School of Information Management & Systems, University of California, Berkeley, Berkeley, CA 94720-4600
- [2] S. Chitra, Dr. B. Kalpana “Knowledge Discovery from Web Logs” Global Journal of Computer Science and Technology Volume 11 Issue 17 Version 1.0 October 2011
- [3] Christy M K Cheung, Matthew K O Lee "Research Framework for Consumer Satisfaction with Internet Shopping" City University of Hong Kong, China. Sprouts: Working Papers on Information Systems, 5(26). <http://sprouts.aisnet.org/5-26> (2005).
- [4] Mike Perkowitz, Oren Etzioni “Towards Adaptive Web Sites: Conceptual Framework and Case Study” Department of

Computer Science and Engineering, Box 352350, University of Washington, Seattle, WA 98195

- [5] 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.
- [6] Hung-Yu Kao, Jan-Ming Ho, and Ming-Syan Chen “WISDOM: Web Intra-page Informative Structure Mining based on Document Object Model”
- [7] Ergin Elmacioglu, Min-Yen Kan, Dongwon Lee, Yi Zhang “Web Based Linkage”, WIDM’07Lisboa, Portugal., November 9, 2007
- [8] 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.
- [9] Tak Woon Yan, Matthew Jacobsen, Hector Garcia-Molina, Umeshwar Dayal “From User Access Patterns to Dynamic Hypertext Linking”
- [10] Ramakrishnan Srikant, Yinghui Yang “Mining Web Logs to Improve Website Organization” WWW10, May 1-5, 2001, Hong Kong, ACM 1-58113-348-0/01/0005.

## Author Profile



**Adhokshaj Haridas** received B.E in Information Technology from RTMNU, Nagpur and is presently pursuing his Masters degree in Computer Science & Engineering from Abha Gaikwad-Patil College of engineering, Nagpur. Data Warehousing and data mining is the domain of his interest. Apart from that he has vivid interest in web-technologies.



**Vinod Nayyar** received his M Tech in Computer Science & Engineering from Yeshwantrao Chavan College of Engineering, Nagpur. He has more than 8 year experience in teaching. He is presently working as an Assistant Professor in Abha Gaikwad-Patil College of engineering, Nagpur. He holds a profound interest and knowledge of website structuring and related technologies.