ISSN (Online): 2349-7084
GLOBAL IMPACT FACTOR 0.238
DIIF 0.876

INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING IN RESEARCH TRENDS
VOLUME 2, ISSUE 8, AUGUST 2015, PP 481-485

Expedite Effective Client Navigation System through Improved Website Structure

[1]Mohammad Raziuddin ,[2]T. Venkata Ramana

[1]Assistant Professor, NMREC
[2] Professor & HOD, CSE, SLC's IET
raziuddin.5807@gmail.com, meetramana_12@yahoo.co.in

Abstract:-Planning well-structured websites to encourage successful client navigation has long been a test. A fundamental reason is that the web designers' comprehension of how a site ought to be organized can be extensively unique in relation to that of the clients. While different strategies have been anticipated to relink web pages to enhance navigations utilizing client route information, the completely revamped new structure can be to a great degree eccentric, and the expense of muddling clients after the progressions stays unanalyzed. This paper addresses how to enhance a website without presenting considerable changes. In particular, we propose a numerical programming model to enhance the client navigation on a website while minimizing modifications to its present structure. Results from far reaching tests directed on an openly accessible genuine information set demonstrate that our model not just fundamentally enhances the client route with not very many changes, additionally can be effectively comprehended. Also, 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.

Keywords: Website Design, User Navigation, Web Mining, Mathematical Programming

1. INTRODUCTION:

There are lakhs of user for website since it is large source of information, web site also contain many links and pages every user require different pages at same time or same user may access different pages at different time. As user increases over www we need to make web intelligent we concern here about intelligent website. To make web site intelligent we must know what is content of website, which are users and how website structured all this known as web mining. The advent of the Internet has provided an unprecedented 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 effective websites is not a trivial task [5-6]. Data mining combines data analysis techniques with high-end technology for use within a process. The primary goal of data mining is to develop usable knowledge regarding future events.

The steps in the data mining process are:

- Problem definition
- Data collection and enhancement
- modeling strategies
- Training, validation, and testing of models
- Analyzing results
- Modeling iterations
- Implementing results.

Galletta etal. [7] Indicate that online sales lag far behind those of brick-and-mortar stores and at least part of the gap might be explained by a major difficulty user's 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 having difficulty in locating the targets are very likely to leave a website even if its information is of high quality Web structure mining can be defined as mining of links between pages, which is also called as hyperlinks which enable user to access web sites in form of URL

ISSN (Online): 2349-7084
GLOBAL IMPACT FACTOR 0.238
DIIF 0.876

INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING IN RESEARCH TRENDS
VOLUME 2, ISSUE 8, AUGUST 2015, PP 481-485

and navigate user. In web structure mining developer uses the data from web usage and change structure of web site, pages which is most visited and user spent more time is linked to the start page.

The goal of a Web site is to meet the needs of its users. As a result, as the interests of its users change over the time, a static Web site that does not change itself will soon become outdated and less useful. Accordingly, a Web site must constantly examine site use, and modify itself accordingly to best serve its users. In other words, Web sites should be adaptive. An adaptive Web site has been defined as a Web site that semi-automatically improves its organization and presentation by learning from visitor access patterns (Perkowitz and Etzioni, 1998). In this paper, an attempt is made to build adaptive Web sites, which improve their navigation based on access patterns of its users. An approach for reorganizing Web sites based on user access patterns is proposed. Our goal is to build adaptive Web sites by evolving site.

Structure to facilitate user access. To be more specific, we aim to build Web sites that provide users with the information they want with fewer clicks. This minimizes the effort on the user's side. By analyzing the usage of a Web site and the structure of the Web site, modifications to the Web site structure are found to accommodate changes in access patterns of its users. These modifications will be suggested to the Webmaster for consideration and implementation.

Best Practices for Site Navigation

Create a naturally flowing hierarchy. As what I've stated earlier, your website should flow naturally where users can access first from general content to the more specific content they want on your site.

- Use mostly text navigation instead of images or animation.
- Put an HTML site map page on your site, and use an XML Sitemap file.

Webmasters are also advised to have useful 404 page that guides the user back to a relevant section or page with a link back the home page in case site visitor encounters a broken link or types in an incorrect URL. If a search engine comes across such an error, it can have a negative impact on your search engine visibility. Google provides a 404 widget that you can embed in your 404 page to automatically.

There are two ways to improve user navigability web personalization and web transformation.web personalization deals with user behavior and user profile, sessions and history of data also called as web logs which is created by user's activity on web site, but transformations approaches mainly focuses on developing methods to completely reorganize the link structure of a website. We adopted web transformation technique to facilitate user navigation.

Fig. 1(a). Normal website structure.

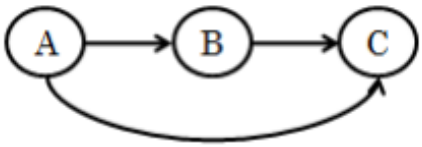

Fig. 1(b). Reorganize website structure.

In fig-1(a) it shows website is like graph and nodes A, B, C are pages and links between pages are edges through which user can navigate, if we found from weblogs that Page 'C' is access very frequently website is reorganized and new link is created from first page so user can access page 'C' in less clicks and time shown in Fig-1(b).First consideration is, in –links and out-links of web pages. Second is the traversing path of user and finally user access pattern. All technique can be used with data mining techniques to facilitate user navigation, since web is huge and user always wants response as fast as possible so our goal is to reorganization website by changing its structure so user should reach target in less clicks and in less time. Proposed strategies are links based clustering to change structure of website as we know clustering is one way of data reduction so lager web usage data is use as compact and time of reorganizing web site structure is reduced. In later section we provide details of some popular clustering algorithm and comparison between clustering algorithm with respect to time.

2. RELATED WORK.

This paper is about survey of web structure mining and clustering techniques over web pages and hyperlinks, as structure mining is useful for organization if done according to user need, so to facilitate user we considered structure mining by performing data mining techniques on weblogs also known as part of web usage mining. About web usage mining, author in [1] explains about weblogs like who accessed order of page request, total time for page view. This paper includes several pre-processing like.

1. Data cleaning-It is method of removing irrelevant items or logs like removing of file with .gif and .jpg extensions.

2. User identification-It involves USER ID for each user to provide uniqueness even different users are on same IP.

3. Session identification- This is defines according to time i.e. time between page request and page close or time out.

4.Path completion- It is defined as if some information or page is important and mostly accessed but not recorded in logs and not linked cause problem .

5.Formatting- It is method of converting transactions or logs it to a format of data mining like removal of numeric value for determining association rules.

Access Information Collection:

In this step, the access statistics for the pages are collected from the sessions. The data obtained will later be used to classify the pages as well as to reorganize the site. The sessions obtained in server log preprocessing are scanned and the access statistics are computed. The statistics are stored with the graph that represents the site obtained in Web site preprocessing. The obvious problem is what should be done if a page happens to be the last page of a session. Since there is no page requested after that, we really couldn't tell the time spent on the page. Therefore, we keep a count for the number of times that the page was the last page in a session. The following statistics are computed for each page:

- Number of sessions in which the page was accessed;
- Total time spent on the page;
- Number of times the page is the last requested page of a session.

Page Classification:

In this phase, the pages on the Web site are classified into two categories: index pages and content pages (Scime and Kerschberg, 2000). An index page is a page used by the user for navigation of the Web site. It normally contains little information except links. A content page is a page containing information the user would be interested in. Its content offers something other than links. The classification provides clues for site reorganization. The page classification algorithm uses the following four heuristics.

(1) File type.

An index page must be an HTML file, while a content page may or may not be. If a page is not an HTML file, it must be a content page. Otherwise its category has to be decided by other heuristics.

(2) Number of links.

Generally, an index page has more links than a content page. A threshold is set such that the number of links in a page is compared with the threshold. A page with more links than the threshold is probably an index page. Otherwise, it is probably a content page

(3) End-of-session count.

The end-of-session count of a page is the ratio of the number of time it is the last page of a session to the total number of sessions. Most Web users browse a Web site to look for information and leave when they find it. It can be assumed that users are interested in content pages. The last page of a session is usually the content page that the user is interested in. If a page is the last page in a lot of sessions, it is probably a content page; otherwise, it is probably an index page. It is possible that a specific index page is commonly used as the exit point of a Web site. This should not cause many errors at large.

(4) Reference length.

The reference length of a page is the average amount of time the users spent on the page. It is expected that the reference length of an index page is typically small while the reference length of a content page will be large. Based on this assumption, the reference length of a page can hint whether the page should be categorized as an index or content page. A more detailed explanation is given below, followed by a page classification algorithm based on these observations and heuristics

Metric for Evaluating Navigation Effectiveness

The Metric

Our 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 [10] point out that information becomes useful only when it is presented in a way consistent with the target users' expectation.

3. METHODOLOGY

The procedure for the quality assessment of website structure involves three modules: establishment of sitemap, computing path length metric and evaluating structural complexity of website. All these modules are incorporated in a web program.

3.1. Establishment of sitemap

Every website must have sitemap to know the organization of web pages in the website structure. The

ISSN (Online): 2349-7084
GLOBAL IMPACT FACTOR 0.238
DIIF 0.876

INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING IN RESEARCH TRENDS
VOLUME 2, ISSUE 8, AUGUST 2015, PP 481-485

sitemap shows all web pages in a hierarchical tree with home page as root of the tree. A web tool PowerMapper is used in the procedure to construct a sitemap for the website. It selects URL address of website and generates the tree structure for all web pages of website. In this process only markup files (html, asp, php, xml, etc.,) are considered and remaining components like graphic files script files, etc., are not included because these files do not have any significance in website structure. The sitemap of a website may be organized into various levels depending on its design. Some websites have one or two levels and some may have three or more levels. A snapshot of Aligarh Muslim University's website sitemap is shown in figure 3

3.2. Evaluating Path length metric

A path length is used to find average number of clicks per page. The path length of the tree is the sum of the depths of all nodes in the tree. It can be computed as a weighted sum, weighting each level with its number of nodes or each node by its level using equation (1). The average no. of clicks is computed using equation (2). The width of a tree is the size of its largest level and the height of a tree is the length of its longest root path.

Path length = Σ li.mi (1)

Where li is level number i, mi is number of nodes at level i.

Avg no. of clicks = path length/n (2)

Where n is the number of nodes in the tree an example tree is shown in figure3.

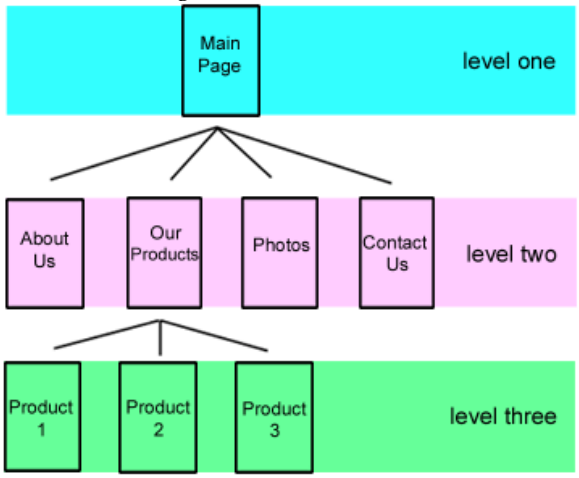

Fig 2. A Tree with 3 Levels.

3.3. Structural complexity

The structural complexity of website is determined with Mc. Cab's cyclomatic complexity metric [2]. This metric is used to know navigation path for a desired web page. The cyclomatic complexity metric is derived in graph theory as follows. A tree graph is constructed with home page as root. The tree consists of various sub trees and leaf nodes. An example tree is shown in figure 4.

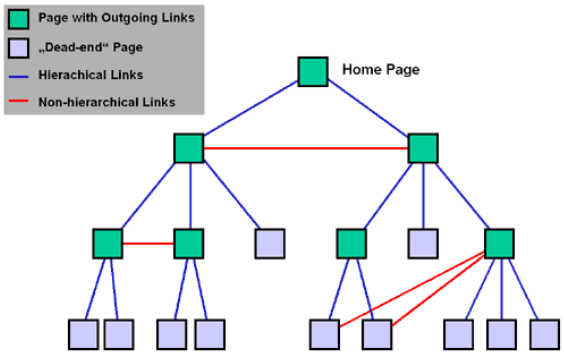

Fig 3. Tree Graph for a Website

A tree graph is constructed for a website by considering various hyperlinks in the website. Each sub tree of the graph represents a web page which has further hyperlinks to the next web pages and leaf node represents a web page which does not have further links to any web pages. In tree graph, at each level all web pages that do not have further links are represented with one leaf node at that level and a sub tree at each level consists of links to the web pages to the next level. The cyclomatic complexity is calculated using equation (3) and it should not exceed 10 according to Mc. Cab.

web_site_complexity = (e-n+d+1)/n (3)

Where e = number of web page links
n = number of nodes in the graph
d = number of leaf nodes in the graph

4. CONCLUSION

Website reorganizes facilitate user to improve navigability, this paper focuses the broad areas of web site reorganization and link analysis on the basis of web logs and user session and data mining techniques applied on web data, which enables user to reach target in fewer clicks. This survey is beneficial for web developer to understand different aspect of website, for researcher to improve more in website and for commercial organization. Website reorganization is imp aspects as now days; it is vast source of information. From clustering comparison we can conclude that time taken to build model is less in k-means so k-means in fastest than any other algorithm and as web structure mining needs to decrease waiting time it is good to use.

REFERENCES:

[1] Robert Cooley, Bamshad Mobasher, and Jaideep Srivastava" Data Preparation for Mining World Wide Web Browsing Patterns".

[2] Mike Perkowitz and Oren Etzioni "Towards adaptive Web sites: Conceptual framework and case study" Artificial Intelligence 118 (2000) 245–275, 1999.

[3] Joy Shalom Sona, Asha Ambhaikar "Reconciling the Website Structure to Improve the Web Navigation Efficiency" June 2012.

[4] C.C. Lin, "Optimal Web Site Reorganization Considering Information Overload and Search Depth," European J. Operational Research.

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

[6] 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.

[7] Min Chen and young U. Ryu "Facilitating Effective User Navigation through Website Structure Improvement" IEEE KDD vol no. 25. 2013.

[8] Yitong wang and Masaru Kitsuregawa "Link Based Clustering of Web Search Results" Institute of Industrial Science, The University of Tokyo.

[9] Bamshad Mobasher,Robert Cooley, Jaideep Srivastava "Creating Adaptive Web Sites Through Usage-Based Clustering of URLs".

[10] Bamshad Mobasher,Robert Cooley, Jaideep Srivastava"Automatic Personalization Based on Web Usage Mining".

[11] Bamshad Mobasher, Honghua Dai, Tao Luo, Miki Nakagawa" Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization"