International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 3 Issue: 6 4107 - 4111

eNavigate: Effective and Efficient User Web Navigation

Mr. Nitin B. Kasar
Department of Computer Engineering,
DGOI,FOE, Daund
Savitribai Phule Pune University,
Pune. India
nkumar.kasar@gmail.com

Prof. Amrit Priyadarshi
Department of Computer Engineering
DGOI, FOE, Daund
Savitribai Phule Pune University,
Pune, India
amritpriyadarshi@gmail.com

Abstract- Web Site is huge source of information. User requires different pages at the same time or same user may access different pages at different time. Web structure mining uses data from web usage and accordingly makes changes in structure of web site. Web site that semi- automatically maintains its organization and presentation by learning from visitors access patterns. Developer develop web site according to their own judgment of use, without considering users intension to use. So users suffers problem of searching in network. We propose a mathematical programming model to improve the user navigation on a website with minimum alterations to its existing structure. Test performed on publicly available real data sets shows that this model significantly improves the user navigation with very few changes as well as effectively solves the navigation problem. The completely reorganized new structure can be highly unpredictable, and cost of disoriented users after website structure changes remains unanalyzed. This approach is how to improve web site without introducing substantial changes. We use two metrics and use them to access the performance of the improved website using the real data set.

Keyword- *Web Navigation,* Static *and Dynamic Websites, Target pages, Website Optimization, Stay time, Mini Session*

****

I. INTRODUCTION

Web navigation and search are two most used ways for finding information on the World Wide Web. Web browser is the main tool used by most users to navigate websites. A general way to search a web is either start with the home page of the known web address or by searching web pages in search engine by interlinked web pages and then followed by hyperlinks as relevant to the search page until the need information is find out. Search facilities are might be provided by the individual websites to search particular web page within itself. As world wide web is fastest growing data as well as number of users till the 2009 there were 1.75 million netizens and the counts has increased by 40 percent by 2010 and till counting day by day. Theses growing number of users not only helps in more improvement in technology but also create a huge business opportunities for the online firms like Ecommerce and Educational fields. To satisfy the increasing demand of online users the firms are competing in improving more and more business. To help the increasing demand of users' website plays the important role for firms and hence firms are investing more and more money in development and maintenance of the websites. Poor website design result from gap between understanding of website structure of developers and the end users [1],[16]. Such differences result in miss organization of website contents hence user cannot easily find out the information on the website. This issue is difficult to handle because when creating a website, web user's preferences are not match up with the developers and developer my develop websites based on their own ideas. However, the website effectiveness can be measured by end users satisfaction and not by the website developers [17], and hence, Web Pages should be organized in a way that generally matches the user's model of how pages should be organized and not the developers. Therefore the problem of improving a website for

efficient and effective navigation for users with minimal Changes to the current structure is an important issue.

II. RELATD WORK

Survey on other work in this area has guided us to categories in five desired data for websites (1) Do not create an extra work for users (*e.g., filling out surveys, questionnaires or feedbacks*). If user are not getting any rewards and said questionnaires is only for improvement then visitors more easily neglect the site and my stay away rather than participate. If site would not be improved without feedback then it's difficult to retain visitors.

(2) Design the website which would be easier to use for every user, including first-time users, occasional users, etc the

changes or customizations are useful for repeating users, but do not benefit first-time users. It is also possible one user's customizations do not apply to other users, as there is no sharing or aggregation of information across multiple users. These both limitations can be overcome by Transformation. (3) *Webmaster work should be minimum*, although human-authored Meta information (e.g., XML annotations) may facilitate site adaptively, it should weigh the benefits of this additional information against the cost of authoring it. (4)T*he site's original design would be protected from destructive changes.* For safety, designs should be limited to *nondestructive transformations*: that are changes to the site which leave existing structure as it is. This approach may have permission to add links but not take out them, create pages but not remove them, add new structures but not change existing ones.

(5) *This approach should keep 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".

2.1 Re-organizing Link Model

We can assume websites as directed graph and each page as node and redirecting URL between pages as edges figure-1 has the website with seven pages and many links, links as an edges represent as 1 or 0. Let's say link Xij where source node "I" connect it with the destination node j.

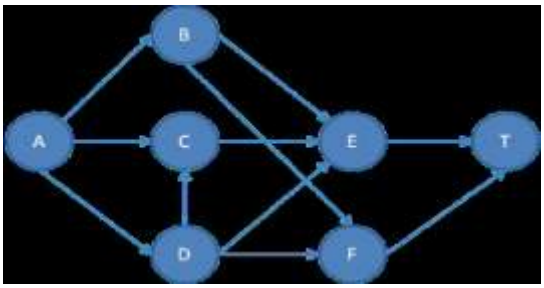

Figure 1: An Example

Consider, it would be forming cluster for which it have object are URL on which user is active and URL which satisfies threshold criteria for cluster of links between node i and j for which cluster can be represent as K_{ij}, in clustering it will use the similarity measures by farthest distance and use distance measures is Euclid, here it will have two parameter session S and Clicks C on which cluster is performed for most far value from mean.

2.2 An Adaptive Web paradigm

As World Wide Web is tremendously increasing, there come a demand of today for intelligent tools, designs and structures which can make simple navigation for users and feels happy experience. And also maximize the quality and completeness of the website experience. These provided tools and structures should provide good level of intelligence so that anyone can sense the environment, know and interpret the situations in order to make decisions and also to control actions. From paper's point of view, this is possible through the integration of techniques from multiple disciplines and research areas. One approach can be applicable like use of artificial intelligence [1], user modeling, graph theory, and information mining techniques to create websites and website browsing tools which are adaptive [2]. *Adaptive* refers to the ability of the website or tool to change its behavior or responses in reaction to the way it is used. An adaptive website should have the ability to recognize end users and actions to plan for the future design.

The behavior of users can be identified by web logs and which is the main source for continuously improve and for tuning sites for users. Adaptation may be done in the form of temporarily altering text, links or page format. New pages creation and removing or adding links is also possible which may be considered as permanent adaptation or change. Although adaptive web technologies belong properly as a subset of the field of adaptive hypermedia, the rapid growth of the World Wide Web has increased its importance and valuation to the point where it currently overshadows the larger field. This paper considerate issues and technologies specifically relating to the Web, mentioning those that belong to the larger field only for historical reasons or when there are useful abstractions or generalizations.

On a broadest sense we can define adaptive website is a website which changes depending on its way of use. Changes can take on many forms, may either be immediate (as in the case of recommendation systems) or gradual (as in the case of systems where suggest changes to a website administrator). An adaptive website technique is applicable and useful for smaller, closed-corpus websites, where the entire site is known well before in advance, or to the Web in general, where it is virtually impossible to know all the web pages even casually. In the latter case, we might describe it as an adaptive web browsing tool rather than an adaptive website.

2.3 A Hybrid Web Personalization Model:

The process of Web personalization generally can be divided into three phases: Data preparation and transformation, pattern discovery, and recommendation. In traditional collaborative filtering approaches, two phase out of above mentioned three are performed in real time namely those phases are the pattern discovery phases (e.g., neighborhood formation in the k-nearest-neighbor) as well as the recommendation phase. In contrast, personalization systems based on Web usage mining i.e. web logs, perform the pattern discovery (second phase) phase offline. Data preparation phase transforms raw web log files into click stream data that can be processed by data mining tasks. Different data mining techniques can be applied to the click

 ISSN: 2321-8169
4107 - 4111

stream or Web application data in the pattern discovery phase, such as clustering, association rule mining [3],[4], and sequential pattern discovery[5].

The recommendation engine considers the active user session in conjunction with the discovered patterns to provide user-specific content. The user-specific content can take the form of specified links or products, or user targated advertisements tailored to the user's perceived preferences as determined by the matching usage patterns. focus In this paper is specifically on association rule mining and sequential pattern discovery, and the suitability of the resulting patterns for personalization approach.

III. PROPOSED SYSTEM

Web Mining Technique: Discovering the model of underling link structure of web pages is the web structure mining, catalog them and generate information such as the similarity and relationship between them. Web Mining is to automatically discover and extract information from web documents. There are three areas of Web mining according to the usage of the Web data used as input in the data mining process, namely, Web Content Mining (WCM), Web Usage Mining (WUM) and Web Structure Mining (WSM).

Effective User Navigation with Minimal Changes: The web navigation design should be simple and consistent. The websites even having high end information would be ignored by users having difficulty in locating the targets. The complete reorganization could radically change the location of familiar items and hence the new website may disorient users.

Mathematical Programming Model: The model for improving the user navigation named as Mathematical Programming (MP) model, which is used for improving the user navigation with most possible minimal changes to the current structure of a website. This MP model not only successfully accomplishes the task but also generates the optimal solutions surprisingly fast. Mathematical Programming (MP) model is applied to clustering results to improve the user navigation on a website effectively by identifying the existing links to be improved and new links to be added. The experiments on synthetic data help to show that this model also extended very well. Mathematical formula for this MP model is shown below with the used notation details in table1.

$$\text{Minimize: } \sum_{(I,j)\in E} x_{i,j}\,[1 - \lambda_{ij}\,(1-\epsilon)] + m \sum_{i \in N_E} p_i$$

Subject to,

$$C^{S}_{kr} = \sum_{(I,j)\in E} x_{i,j}\; a^{s}_{ijkr}\; x_{ij}\,;\; r = 1,2,.L_p(k,s)$$

$$k = 1,2,\ldots..,L_m(S);\ \forall S \in T^R$$

$$C^{S1}_{11} + C^{S1}_{12} + C^{S1}_{21} \geq 1$$

$$C^{S2}_{11} + C^{S2}_{12} + C^{S2}_{21} + C^{S2}_{22} \geq 1$$

$$\sum_{J=1} x_{i,j}\,(1 - \lambda_{ij}) + W_i - P_i \leq 3\,;\ \forall_i \in N_E$$

Such that,

$$W_i = \sum_{j \in N} \lambda_{ij}$$

$$L_{ij} = \begin{cases} 1 & \text{establish link} \\ 0 & \text{discard} \end{cases}$$

Table 1 – summary of notation

Notation	*Definition*
S	***A mini session containing set of path traversed by user to locate one target page***
T	***The set of all identified mini sessions***
T^R	***The set of all relevant mini sessions***
N	***Set of all web pages***
λ_{ij}	***1 if page i has link to j zero otherwise***
E	***Set of candidate link which can be selected for improve navigation***
E^R	***Relevant candidate link***
N_E	***The set of source nodes of link in set E***
Wi	***Current out degree of page i***
Ci	***Out degree threshold for page i***
Pi	***Number of pages that exceeds out degree threshold in page i***
M	***Multiplier for penalty function***
Bj	***Path threshold for mini session for which page j is target page***

As a first step once identifies the drawbacks of website it can now form a reconstruction model which is supposed to improve the structure of the website instead of reorganizing it

 ISSN: 2321-8169
4107 - 4111

from scratch. The proposed approach shown in figure 2, will facilitate the user navigation on a website with minimal changes to its current structure. Like for example in a informational websites whose contents are static and relatively stable over time (called as static websites) such as universities, hospitals, etc. and this proposed model would not work effectively on dynamic websites because such sites contains runtime scripts which runs depending on the input attribute and hence not easy to improve.

In a directed graph *G* may have vertices and edges unless it is a null graph and in such graph there may be a incoming edge and may be some outgoing edges based on which in degree and out degree of vertices are being calculated and in the same manner a webpage will be having calculation of degrees based on which it can model the web structure. If a website is having too little web pages then calculating the in degree and out degree is considered to be undesirable and too many links which may exceed the threshold value of the allowable degrees may led to overloading of information and is being considered to be undesirable and in this study, approach considered out degree threshold value to be 10 and based on this threshold value it eliminate the excess links to improve website performance.

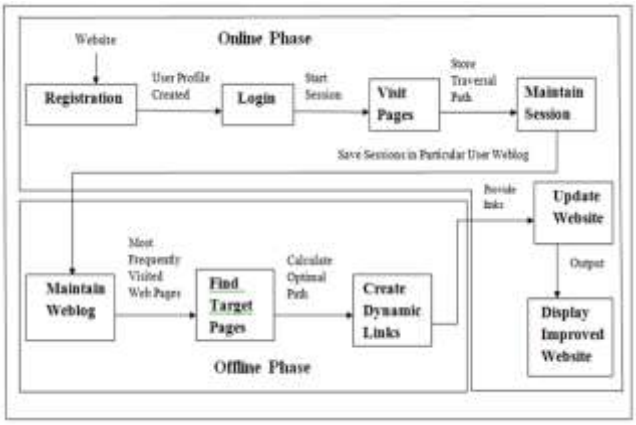

Figure 2. System Architectures

IV. OUTPUT DESIGN AND DATASETS

Two types of data sets can be used for analysis of the work as 1. Real Data sets: The real data set is collected from music machine company website http://machines.hyperreal.org. This is publicly available and also used in most number of literatures [12] [13]. Statistically it would have about 4 millions recorded requests in a time span of four month.

2. Artificial Data sets In addition to the real data set, synthetic/artificial data sets would be generated and considered for computational experiments to evaluate the scalability of this model with respect to the size of the website and the number of mini sessions. For this purpose, the artificial website structures and mini sessions would be generated to have similar statistical characteristics as the real data set. For instance, the average out degree for pages in the real website is 15, so the link structure for the artificial website would be generated in a way such that each page contained 15 links on average.

The output design would have the reorganized and improved link structure for websites with possible minimal changes in current structure.

VI. RESULT ANALYSIS

We demonstrated the results of the proposed system on a 10 page static website designed for engineering college(s). The users' navigation/session log maintained in log file is parsed to calculated page visits and corresponding stay times. Further we achieved set of target pages, set of backtracking pages and a set of candidate links to be added to efficiently improve the web structure. Then upon visiting the next web page we optimized the candidates' links by using mathematical programming and modeled the new links which should be added to the loading page. Table 2 shows the resulted precision and recall calculated on the data achieved from the experiment and corresponding to the values Precision Vs. Recall graph is shown in figure 3.

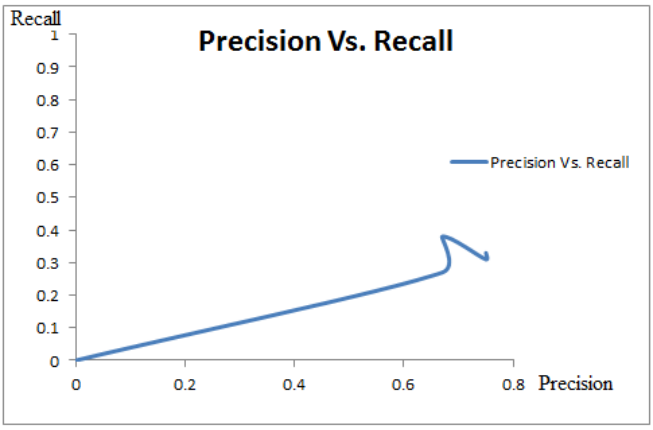

Figure 3. Precision Vs. Recall Graph

We found the accuracy of the proposed system in terms of precision as 0.77 (77%) while the availability in terms of recall as 0.39 (39%) approximately. Further we analyzed the root cause for lower recall and we found that the optimization strategy applied for calculating the final links to be added from set of candidate links decreases the recall values.

Web Page	#Target Pages	#Candidate Links	#New Links Added	#Correct Links	Recall	Precision
HomePage	4	12	4	3	0.33	0.75
Department	3	8	3	2	0.38	0.67
Admission	5	13	4	3	0.31	0.75
Academics	4	11	3	2	0.27	0.67
Hostel	2	5	2	2	0.40	1.0

Table 2. Result Analysis

V. CONCLUSION

This paper explains solution for effective website navigation by making minimal changes to current structure, hence resolved a critical issue in domain of web development for websites having stable contents over the long period of time.
This paper uses mathematical model which reorganize the current link structure of a website rather than redesigning the website itself time to time. We have applied and tested the proposed system on real websites. The result of application significantly improves user navigation with addition of very few new links, optimal solution quickly obtained with suggesting that model is very effective to websites in real world within very few seconds of time.

ACKNOWLEDGMENT

I am greatly thankful to Prof. Amrit Priyadarshi for his great effort of guidance and supervision to accomplish this fine work. I would like to thanks all my college staff and department for support and encouragement. Last1 but not least my friends and family, for their warm, kind encourages
and loves. To every person who gave me something too light along my pathway. I thanks for believing in me.

REFERENCES

[1] M. Perkowitz and O. Etzioni, "Towards Adaptive Web Sites: Conceptual Framework and Case Study," Artificial Intelligence, vol. 118, pp. 245-275, 2000.
[2] 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.
[3] 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
[4] R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large databases, in: Proc. ACM SIGMOD Conference on Management of Data, 1993, pp. 207–216.
[5] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, A. Verkamo, in: Fast Discovery of Association Rules, MIT Press, Cambridge, MA, 1996, pp. 307–328.
[6] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: Proc. 20th VLDB Conference, 1994.
[7] T. Anantharaman, M. Campbell, F. Hsu, Singular extensions: Adding selectivity to brute-force searching, Artificial Intelligence 43 (1) (1990) 99–109
[8] Altavista's Babel Fish translator. http://babelfish.altavista.com/.
[9] R. Agarwal, C. Aggarwal, and V. Prasad. A tree projection algorithm for generation of frequent itemsets. In Proceedings of the High PerformanceData Mining Workshop, Puerto Rico, April 1999.
[10] L. Ardissono, L. Console, and I. Torre. An adaptive system for the personalized access to news. *AI Communications*, 14(3):129–147, 2001.
[11] M. Balabanovic. An adaptive web page recommendation sservice. In *Proceedings of the 1st International Conference on Autonomous Agents*, pages 378–385, Marina del Rey, CA, USA, 1997
[12] R. Gupta, A. Bagchi, and S. Sarkar, "Improving Linkage of WebPages," INFORMS J. Computing, vol. 19, no. 1, pp. 127-136, 2007.
[13] 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.
[14] R. Agrawal and R. Srikant. Mining sequential patterns. In Proceedings of the International Conferenceon Data Engineering (ICDE'95), Taipei, Taiwan, March 1995.
[15] B. Berendt, B. Mobasher, M. Nakagawa, and M. Spiliopoulou. The impact of site structure and user environment on session reconstruction in web usage analysis. In Proceedings of the 4th WebKDD 2002 Workshop, at the ACM-SIGKDD Conference on Knowledge Discovery in Databases(KDD'2000), Edmonton, Canada, July 2002.
[16] 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.
[17] M. Perkowitz and O. Etzioni, "Towards Adaptive Web Sites:Conceptual Framework and Case Study," Artificial Intelligence,vol. 118, pp. 245-275, 2000.