RESEARCH ARTICLE

An Eccentric Methodology for Augmenting Web Navigation

Adhokshaj Haridas[1], Pragati Patil[2], Yogesh Bhute[3]

[1]Abha Gaikwad-Patil College of Engineering, RTMNU, India

[2]H.O.D, Department of Computer Science & Engineering, Abha Gaikwad-Patil College of Engineering, RTMNU, India

[3]Assistant Professor, Department of Computer Science & Engineering, Abha Gaikwad-Patil College of Engineering, RTMNU, India

[1] adhokshaj66@gmail.com; [2] pragatimit@gmail.com; [3] yog.bhute@gmail.com

Abstract— *In today's world websites are commonly used in almost every walk of life. The use of websites has certainly brought a huge exchange of important data via the websites, where websites being the platform for such exorbitant information exchange. Being a bridge for such a change there needs to be an inherent requirement to make the websites highly efficient and hence strengthening their performance. An attempt is made in this direction to improve the websites efficiency by following a mathematical model which treats the website as a graph and resolves its complexity assuming it as a graph optimization problem. Treating website as a graph has its own advantage as it enables us to handle the complex links present in a website in an efficient manner.*

Keywords—Web Structure Improvement; Web Usage Mining ; Website Navigation ; Web personalization

I. INTRODUCTION

Use of websites has become a part of life, today. From paying ones monthly expenses to buying products of desire or choice, whatever it is today can be attained through the websites. Websites are not only easy to work with but are a great platform to view all the information that we may want for different purposes. We can safely assume that in present day world the websites are the most favored and fairly convenient means of information exchange. WWW (World Wide Web) has certainly brought the world closer and has hence connected the lives of humans throughout the world. As we have learnt so far about the websites that they serve as a virtual bridge which has brought the most distant of the places on the earth closer to one another. This is the biggest revolution that has literally captured the imagination of the whole world. This is the advantage of having the information technology take the centre stage in the global perspective.

But the medium that we trust today more than a fellow human, if this medium of information exchange becomes a bit dodgy then there are serious repercussions on the productivity in general. Not only that the human life may be affected by it to a great extent. It is such an imperative need today, that it is necessary to have websites of the highest performance. This research is directed in a way, which leads to the destination of achieving the enhanced results that the users of the website anticipate.

This method of improving the web navigation within a website is one of its kind, as it has been based on a mathematical foundation. We have applied the concepts of graph optimization on the websites which are assumed to be graph to carry out the link optimization work. This method which is eccentric in its kind combines the concepts of web usage mining along with the mathematical model. So web usage mining plays a vitally integral role in fetching the activity data as and when required from the web server. The goal is to achieve an augmented standard of website for which the following work has been carried out.

II. RELATED WORK

As explained by Bamshad Mobasher[1] his work presents an overview of Web personalization process viewed as an application of data mining requiring support for all the phases of a typical data mining cycle. These phases include data collection and pre-processing, pattern discovery and evaluation, and finally applying the discovered knowledge in real-time to mediate between the user and the Web. This view of the personalization process provides added flexibility in leveraging multiple data sources and in effectively using the discovered models in an automatic personalization system.

Also with regards to the work done by Wen-long LIN et.al.[2] the work presents a novel website structure optimization model for more effective web navigation. First, web page group with low access efficiency is discovered by its support and its topology average distance; Then a measure degree, website topology interest, which can overall indicate the website access efficiency is proposed as guidance rule to optimize the website hyperlink structure; Finally, user's navigation are facilitated by optimizing website linkage structure that reduces the number of steps to locate their target web pages. Experiments result on a distance education website show that our approach is efficient and practical for adaptive website. Other than this the work done by Asem Omari et.al. [3] explains that the success of a commercial company depends greatly on the success of its website. A successful website is the well-designed website. The website is well-designed from the user's point of view if it contains services that satisfy user's needs, if the user navigation is simplified, and if he can reach his target page in a short time without the need to make any search or to guess where his target page could be found. On the other hand, from the website's owner's point of view, a website is well-designed if it participates in increasing the company's overall profit, if it increases the user's trust in the company and its products, and participates in making promotions for the company. Therefore, it is important to have a well designed website.

Apart from these methods another explained by Izzat Alsmadi et.al [4] describes evaluation of website characteristics can take several methods. Some of those methods depend on users while others depend on the websites themselves. There are many software tools and related websites that measure websites attributes such as vulnerability, performance, navigability, structure, etc. His work focuses on studying website structural and related metrics that can be used as indicators of the complexity of the websites. Websites structural metrics can be also used to predict maintainability requirements. Examples of some structural metrics evaluated in his study include: size, complexity, and speed of page loading. These are the few methods that have inspired us in a way to take forward our research in a specific direction .The knowledge that these methods imparted has certainly been crucial to us in formulating a completely different way of approach.
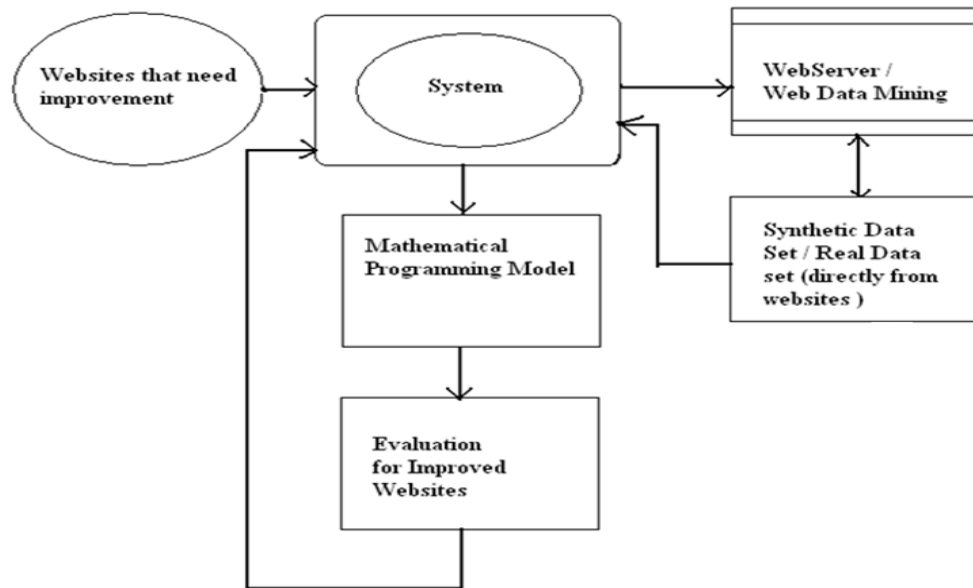
III. METHODOLOGY

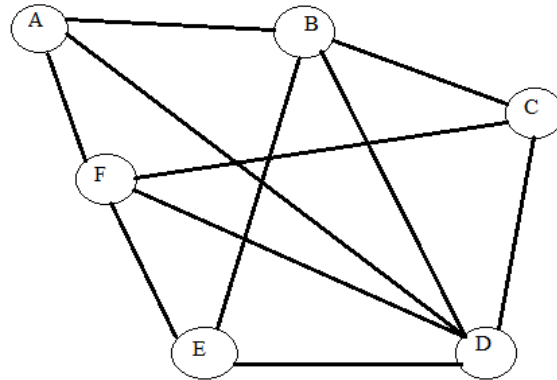*Fig.1. Diagrammatic representation of the system*

With regards to the approach that we follow to achieve the required implementation we focus on various activities. Starting with the first, we want to have a website as an object for experimentation. As it is not available we first develop a basic news website using java netbeans (jsp,swing,servlet). Once this is done we use MySQL to establish our database. Since we host our own website we use the apache tomcat an open source web-server and servlet container developed by the Apache Software Foundation (ASF). Tomcat implements the Java Servlet and the Java Server Pages (JSP) specifications from Sun Microsystems, and provides a "pure Java" HTTP web server environment for Java code to run in. In the simplest configuration Tomcat runs in a single operating system process. The process runs a Java virtual machine (JVM). Every single HTTP request from a browser to Tomcat is processed in the Tomcat process in a separate of thread.

Secondly we use the mathematical programming model. This is done to solve the complexities of the links that are already present within the website. These links are the old links i.e. they are developed by the web application developer without any emphasis on navigational efficiency of the website. The website with its original link structure is assumed as a graph. Since this website embodies more than a few navigational issues we assume it as a graph optimization problem. Our approach to tackle this link complexity is by addressing the problem by mathematical solution. We do the programming in java netbeans IDE to attack on the problem directly. We use the matrix method to solve the graph optimization problem. We organize the links that are being used within the matrix and then address each link relation.

Our system fetches the web usage data from the web-server. Not only that it also calculates the links priority depending upon users choice for each and every link. This we call as the out-degree of a link. Depending upon the out-degree of links we can personalize the links present in the website. Apart from that our system also has a mini session based link analysis. Mini session module calculates in seconds the amount of time the user was on a particular link. After that we apply the changes on the old links. We inform the user to refresh his websites page to apply the made changes. Once that is done we can identify the personalized links and also the navigational efficiency improved to a certain extent.

IV. IMPLEMENTATION

Fig. 2. Diagram representing a website as a graph

Above diagram represents a website as a graph. Based on the mathematical concept of graphs we solve the complexities of the links present in the website. In this method we treat the website as a graph and its connecting links as the edges of the graph. Apart from that the graph that represents the website is in itself an embodiment of complexities of internal links. This graph needs to be treated in a mathematical sense.

To begin with we have created a simple news website which comprises of few links. We as users of the website then have to use the website by clicking on the links of interest.

As we click on the links the web server gets notified and saves the web usage in its database, which we may require at a later stage.
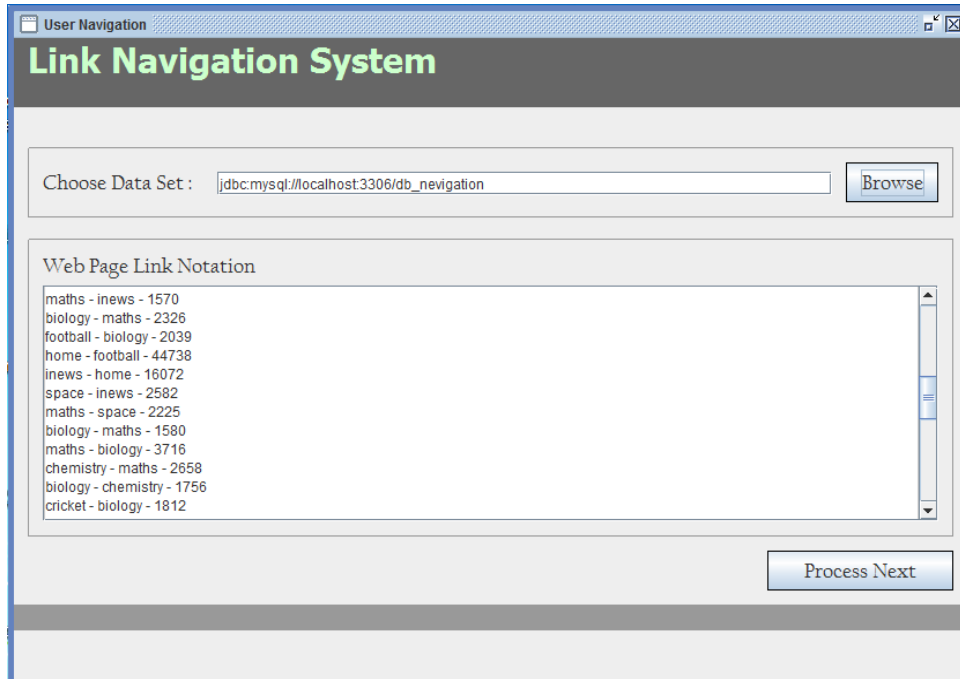
After that we program for a method that works around in solving the graph for its complexity.
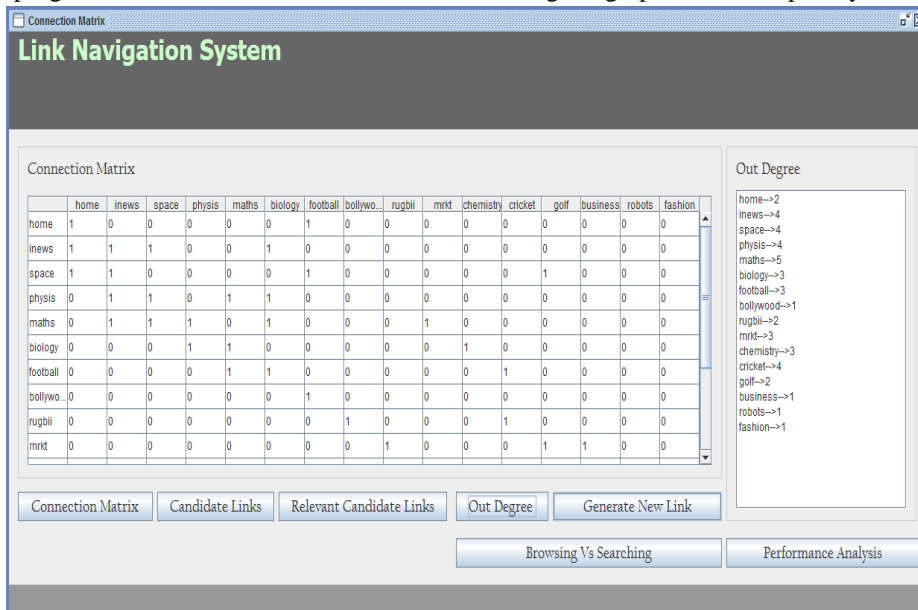
We then generate a connection matrix that holds the connectivity information of the various links that have been frequently used by the user. Once the connection matrix is resolved then we find the candidate links. Candidate links are all the links that are presently under the purview. Then once we get the list generated for such candidate links then we are interested in getting the relevant candidate links. The relevant candidate links are the links that we are interested in. Out degree is an aspect that will enable us to better search the relevant candidate links. Out degree is nothing but the number or the count that we maintain for every click that a relevant link gets from the user. From all the above activity we gather the information regarding the links that are present. Based upon this information we can find the mini session based report for a relevant link.
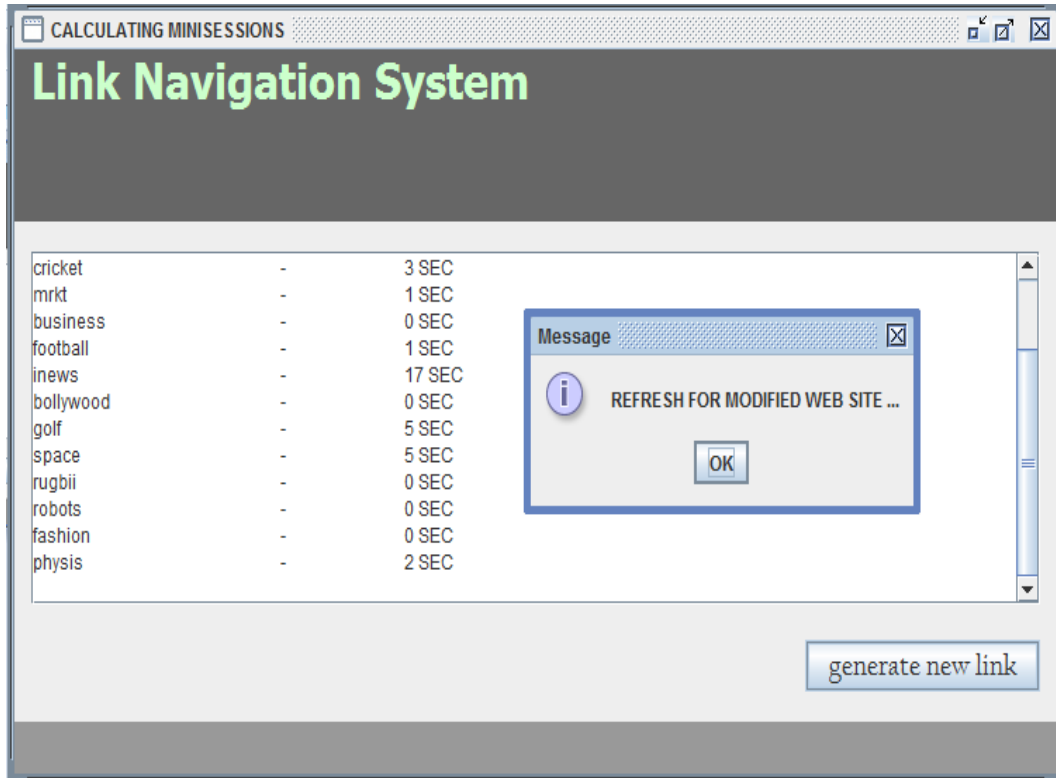
After this we generate a mini session based analysis of the links i.e. depending upon the time spent on a particular link by the user. Finally we apply all these to generate a set of efficient links that incorporate all the changes that we have made thus far. All the changes are applied on the same old links as we do not intend to disturb the old link structure of the website. So after applying changes on the old links we get renewed links that are comparatively shorter, but more efficient. Hence we improve the navigation of websites intra links.

V.   EVALUATION

As we implement all the different internal working modules of our system we get an affirmative response from the respective processes. To begin with when we host our website it successfully gets loaded ,apart from that all the links open as per our expectation .With regards to the connection between the website and the web server we do get a good synchronization as the database records the user activity. Similarly when we use our system to mine the database to fetch the required data it responds efficiently. When we use the other operations present in the system the processes beginning from connection matrix generation to the performance analysis work as per expectation and help in generating an efficient result. We can safely ascertain that both web usage mining and intra-web navigation is implemented together successfully. Also personalization of the links is done appropriately.
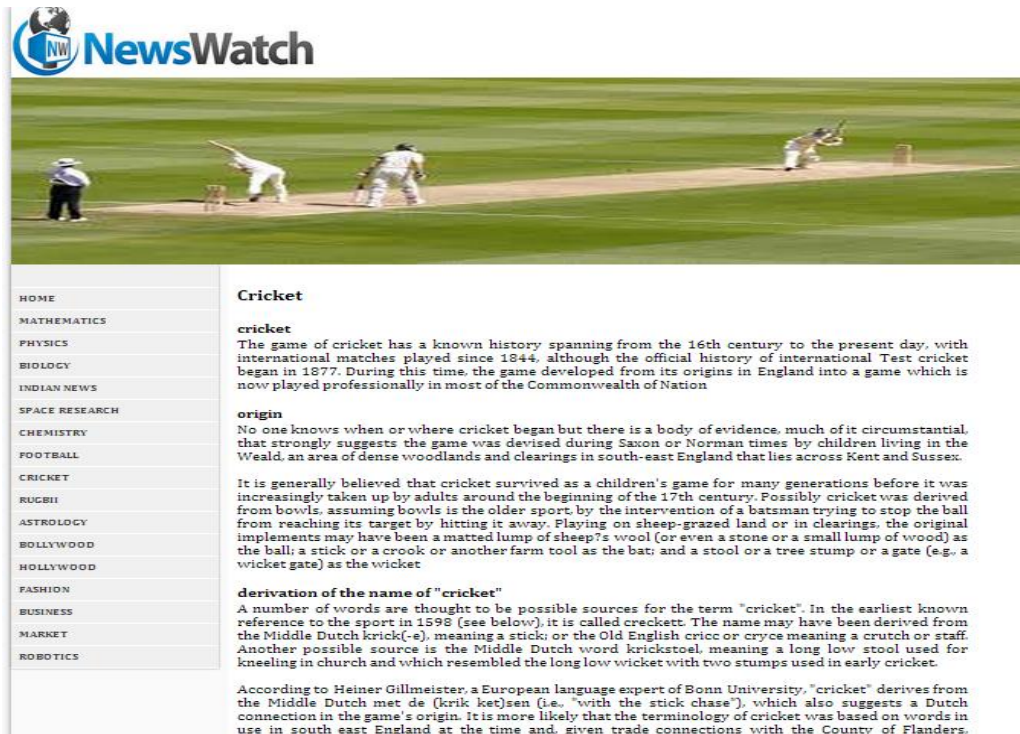
Fig.3. Website links before personalization

Fig.4. Website links after personalization

VI. CONCLUSIONS

To improve the navigational effectiveness of a website while minimizing changes to its original structure, is our central objective. Our system is particularly appropriate for informational websites whose contents are relatively stable over time. It improves a website rather than reorganizes it and hence is suitable for website link improvement. Personalization of website links is also successfully achieved by our system. Such a system which combines both the ideas that of the web structure improvement and web usage mining surely improves the user navigation.

REFERENCES

[1] 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

[2] Wen-long LIN et.al." A Novel Website Structure Optimization Model for More Effective Web Navigation", 2008 Workshop on Knowledge Discovery and Data Mining, IEEE Computer Society

[3] Asem Omari and Stefan Conrad," On the Usage of Data Mining to Support Website Designers to Have Better Designed Websites", Proceedings of the Advanced International Conference on Telecommunications and International Conference on Internet and Web Applications and Services (AICT/ICIW 2006), IEEE Computer Society

[4] Izzat Alsmadi et.al, "Web Structural Metrics Evaluation ", 2010 Developments in E-systems Engineering , IEEE Computer Society

[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.