

Recuperating Website Link Structure Using Fuzzy Relations between the Content and Web Pages

Raj Gaurang Tiwari
AZAD IET, Lucknow
UP, INDIA

Mohd. Husain
AZAD IET, Lucknow
UP, INDIA

Sandeep Gupta
VIET, G. B. Nagar
UP, India

Arun Pratap Srivastava
VIET, G. B. Nagar
UP, India

ABSTRACT

The link structure of website allows us to spread the link power of home page to the individual pages of the site. In this paper we define the content and web pages as two important and prominent factors in website navigation and restate the enhancement in the website navigation as making some useful changes in the link structure of the website based on the aforementioned factors. Then we suggest a new method for proposing the changes using fuzzy approach to optimize the website architecture. Applying the proposed method to a real case of Azad Institute of Engineering & Technology (AIET) Lucknow website, we discuss the results of the novel approach at the final section.

Keywords

Web content, Web navigation, Website system, Web usage mining, Website Link Structure.

1. INTRODUCTION

In current scenario website plays a significant role in success of an e-business and giving more intelligence to e-commerce sites is popularly recognized as one of the effective strategies that increases customer satisfaction because they react intelligently and can give a personalized response to each customer[3]. It is, then, no small wonder that most companies feel that they need at least some level of web presence today.

Moreover, web pages are hard to design in a systematic way. Web architecture, routine path, and page contents are often intuitively decided. These are some of the reasons that lead users to errors or inconvenient access when browsing a website, thereby bringing a negative impression to individuals or companies. In order to deal with this problem, identification of user intention and behavior becomes necessary and the concept of website usability is defined as “How well and how easily a user, without formal training, can interact with a website?” Better structure of web links takes visitors easier and sooner to their targets in a website and also it enhances website navigation. Many previous researches applied web mining techniques to analyze web logs to evaluate website link structure and usability. While some researchers analyze past user access patterns to discover common user access behavior in order to improve website structure, others may employ it to redesign websites or to discover user access patterns and to develop adaptive websites.

To analyze and capture visitors' opinions about website usability, different methods and metrics are proposed. While in some cases, web designs are measured according to their formatting, composition and different web topic categorization and also some scales focus on customer evaluations and different customer groups, many studies use operation research methods to evaluate website usability and enjoy OR potentials in formulating website link structure as mathematic relations. Also, the graph theory has recently attracted many attentions in website usability evaluation. On the other hand, recent developments in data and web mining technology can help enterprises determine problems in communication, and improve their tactics in response to customers. Some of the useful knowledge captured from customers, can be the information regarding the website structure and design, these groups of information from clients and their usage behavior are described as web usage mining which is a sub set of web mining literature. This kind of information about the website layout can help the designers to find the clients preferences and draw backs of his/her design based up on and so to improve the website structure in order to simplify the navigation for the clients. This paper presents the website as a link structure between some nodes of information which are the web pages and illustrates them using the notion of directed graphs. By using such a description about the website, improving the navigation in website can be paraphrased as making some useful changes in the link structure of the website in order to enhance the navigation possibilities for clients.

Web mining taxonomy[2] can be presented as:

Web Content Mining (WCM): is consisted of mining the multimedia documents, involving text, hypertext, images, audio and video information. This deals with the extraction of concept hierarchies/relations from the web and their automatic categorization.

Web Structure Mining (WSM): is composed of mining the inter-document links, provided as a graph of links in a site or between sites. For example, in Google a page is important if important pages point to it. WSM pertains to mining the structure of hyperlinks within the web itself .

Web Usage Mining (WUM): is made of mining the data generated by the users' interactions with the web. This includes trend analysis, and web access association/sequential pattern analysis. While the previous techniques utilize the real or primary data on the web, usage mining mines secondary data

generated by the users' interaction with the web. This category of web mining is the most recent method in personalizing web page.

During the next section we will state the research problem precisely. The third section goes to the characteristics of the utilized fuzzy method and its relative concepts. The fourth part of the paper will discuss a real case of website of Azad Institute of Engineering & Technology (AIET), Lucknow in order to illustrate the functionality of the proposed fuzzy method and finally, the last section brings up some conclusions on the novel contribution of the paper.

2. PROBLEM STATEMENT

Website structure plays an integral role in the success of the firms' marketing strategies and is exactly the problem we are going to deal with through this paper discussing a case study of website of Azad Institute of Engineering & Technology (AIET), Lucknow. "How can the website navigation be enhanced mostly using the data gathered from visitors?" states the research question in this paper. As the visitors are the best ones who can judge the navigation process of a website, we have applied visitors' surveys as the instruments for gathering data about the navigation status of the AIET website. What mostly complicate the research problem are the factors to survey the visitors on along with the necessary methodology for analyzing the data. Finding appropriate solutions for these complexities about the AIET website shapes the main and novel contribution of this paper.

3. WEB CONTENT, WEB PAGES AND THEIR PROPOSED FUZZY RELATIONS

3.1 Web Contents and Web Pages

Each web site is designed in order to present some highlighted content through its web pages. The website content topics are the ingredients which shape the discussed subjects through its web pages[14] and upon which the designers decide about the number of web pages and the link structure which shape the web site architecture. So, the strength of relations between the website content topics and also the intensity of dependency between each web page and each content topic are the main building blocks of the website architecture[1]. It can be paraphrased that each existing link between two distinct web pages is the result of two web pages covering the same content or two web pages covering different content topics which are largely related and dependent to each other. Based on such reasoning, the link structure of any website can be enhanced based on the existing relations between the content topics presented through the website and also the intensity of the relations between each content topic and each web page. This paper proposes a system which can enhance the website architecture based on two types of mentioned relations. Dealing with this problem, understanding the visitor intentions and preferences becomes highly necessary. Web users have different purposes and intentions when browsing a website. By recognizing the visitor preferences, the system can help designers understand visitors' usage behavior and can suggest ways to build more effective websites.

Each designer can extract the main content topics which have been covered through the web site. The intensity and strength of aforementioned relations in a website based on the user opinion can guide the designers in how to construct the link structure inside a web site system and can be used as a criterion in order to classify the web pages in to categories of: authoritative and hub web pages. While authoritative web pages usually contain the most reliable and detailed contents about a specific topic, hub documents contain many links to authoritative documents without going in to details of any special topic[14], presenting the agenda of web site contents, the home page can be a good example of hub web pages. On the other hand the final web pages which present intended information and services for each content, can be categorized as authoritative pages. The authoritative web pages usually have at least one strong relation to the content topics while the hub pages usually don't have any special strong relation with any content.

3.2 Proposed Fuzzy Relations between the Content Topics and the Web Pages

Presenting mathematically, the strength of discussed relations between the content topics and similarly the intensity of relations between the content topics and the web pages are not instinctively crisp. That's why we have used fuzzy theory in order to model such relations mathematically. Any two content topics can have different strength of relation which can be expressed linguistically as "weak", "medium" or "strong". On the other hand the intensity of relation between a content topic and a web page follows the same rule and can not be expressed in terms of crisp values, this intensity can be described similarly through the linguistic values of "weak", "medium" or "strong", too. This brings up the notion of membership value for each relation, which determines the existence intensity of each relation. The fuzzy relations can be illustrated as below:

$$\left\{ \begin{array}{l} \mu_{ij} \\ C_i \rightarrow C_j \\ \\ \mu'_{ij} \\ W_k \rightarrow C_j \end{array} \right. \quad (i \neq j)$$

Where C_i stands for the i th content and W_k represents the k th web page in the website structure, also μ_{ij} shows the membership degree for the relation between the content topics and presents the degree of dependency between them. Similarly the μ'_{ij} represents the membership value of the relation between the web page W_k and the content C_i , which is the degree of dependency between the content topic and the web page. The higher the membership degree, the more the web page covers the content through its subjects.

4. EXPERIMENTAL EVALUATION OF FUZZY APPROACH ON AZAD INST. OF ENGG. & TECH. (AIET) WEBSITE

One of the main responsibilities of AIET is imparting quality education and training to the students in the chosen field of study, to enhance the knowledge, infuse confidence and sharpen their skills preparing to compete in today's globalization age and changing scenario. Having the hits rate of more than 100-150

person/day, AIET web site is playing the integral role of a portal for engineering and management students. Thus, large number of new web pages are designed and added to/substituted by the previous web pages in the web site structure on a monthly basis aiming to simplify the users' access to the intended information as much as possible. Having such a gigantic size, providing an effective way of navigation for such a link structure is a real woe for the team of web designers these days. During previous months, the website of AIET (<http://www.aiet.ac.in>) has been frequently changed to have better face to users. During this period of time, different homepages were developed. The homepage and link structure were dramatically changed again and again and imposed a considerable amount of cost on the department monetarily and time wise. But after changing the structure of the website, there was no "the best design" and so they just agreed on the current layout. In order to present a real example for the proposed fuzzy inference system, we consider the link structure of the AIET website which covers almost 5 content topics through 16 web pages. Figure-1 illustrates the link structure as a directed graph. The web page number 1 on top of the graph illustrates the current homepage.

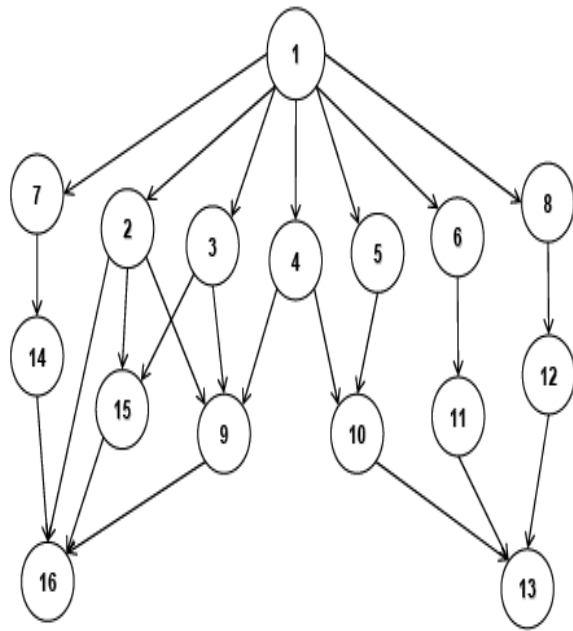


Figure-1: The directed graph of initial AIET website structure

In order to reach a performance measure for the navigation process in the AIET website, browser cookies were used to record the needed data. Since the content of cookies depends on the programmer, we wrote a cookie program to obtain the necessary information from visitors such as user IP, stay time at each page, browsing sequence, and so on. Then applying the time threshold of 10 seconds, which was the least time of stay for each user in any target page, we could distinguish the passing (hub) web pages from the target ones. This helped us measuring the average navigation time for a user since its entrance to the website till he/she reaches his/her first target page. This duration

was 29 seconds on average before applying the fuzzy method to the website. Measuring this KPI after applying the new link structure reached by the use of fuzzy method would help us to understand the degree of improvement in the website navigation. Two types of questionnaires needed to be applied in order to gather the necessary data about the AIET website. The first questionnaire which asked the visitor to comment on the strength of relation between each couple of content topics was presented to the user at his/her entrance to the web site. On the other hand, the second questionnaire which sought the visitor comments on the intensity of relation between each web page with each single content topic was presented to him/her during his/her visit from related web page. Through this process two types of questionnaires were filled by the visitors during his/her visit from the web site.

The visitors' qualitative answers were quantified using some defined fuzzy linguistic rules, so as to enable us to run the inferring process over the collected data. Figure 2 illustrates these fuzzy linguistic rules.

Running two visitors' surveys for 8 days, the surveys were responded by 60 visitors and based on the results we could reach to two different matrices of content topics relations and also the web pages and the content topics relations, denoting the inputs for the fuzzy minimal inference system [15]. Tables I and II present those matrices along with their membership values having used the linguistic rules (C_k stands for the k th content topic and W_i for the i th web page).

Using proposed linguistic rules and applying the minimal fuzzy inference engine, we calculated the membership value for each pair of web pages as the degree of intensity for their relation. Table III presents the resulted membership degree for each pair of web pages.

TABLE I

The relations' strength between the content topics with their fuzzy membership values

	C1	C2	C3	C4	C5
C1	S $\mu \geq 6$	S $\mu \geq 6$	M $3 > \mu > 6$	W $\mu \leq 3$	W $\mu \leq 3$
C2	S $\mu \geq 6$	S $\mu \geq 6$	S $\mu \geq 6$	W $\mu \leq 3$	W $\mu \leq 3$
C3	M $3 > \mu > 6$	S $\mu \geq 6$	S $\mu \geq 6$	M $3 > \mu > 6$	W $\mu \leq 3$
C4	W $\mu \leq 3$	W $\mu \leq 3$	M $3 > \mu > 6$	S $\mu \geq 6$	S $\mu \geq 6$
C5	W $\mu \leq 3$	W $\mu \leq 3$	W $\mu \leq 3$	S $\mu \geq 6$	S $\mu \geq 6$

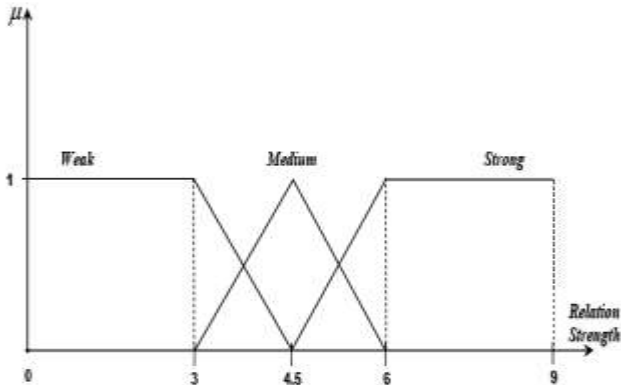


Figure-2 Illustration of the linguistic rules as fuzzy variables

Applying the linguistic rules presented before, the content of Table III can be changed to the linguistic variables demonstrated in Table IV.

The directed graph of Fig. 3 illustrates the link structure of AIET website in addition with the strength of its links, having resulted from the utilization of the proposed fuzzy method.

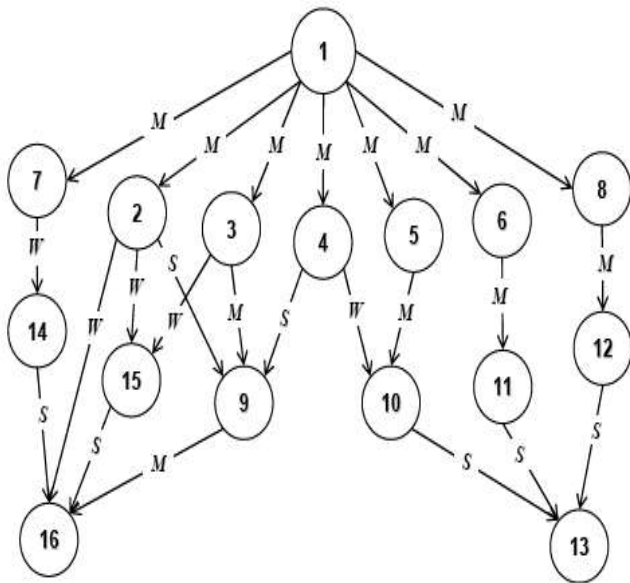


Figure-3: The directed graph of AIET web site structure along with its links' strength

As illustrated in Figure 3, it's obvious that some of the links which were considered in the AIET website structure are so weak that we could omit them from the architecture of the web site. In contrast, as presented in Table IV, there are some important links which can be added to the website link structure in order to better facilitate the navigation process through the web site and make the web site architecture more effective.

TABLE II

The relations between the content topics and the web pages with fuzzy membership value

	C1	C2	C3	C4	C5
W1	M 3>μ>6	M 3>μ>6	M 3>μ>6	M 3>μ>6	M 3>μ>6
W2	S μ≥6	S μ≥6	W5 μ≤3	W μ≤3	W μ≤3
W3	S μ≥6	W μ≤3	W5 μ≤3	W μ≤3	W μ≤3
W4	M 3>μ>6	S μ≥6	W5 μ≤3	W μ≤3	W μ≤3
W5	W μ≤3	S μ≥6	M 3>μ>6	W μ≤3	W μ≤3
W6	W μ≤3	S μ≥6	M 3>μ>6	W μ≤3	W μ≤3
W7	W μ≤3	M 3>μ>6	S μ≥6	W μ≤3	W μ≤3
W8	W μ≤3	M 3>μ>6	S μ≥6	W μ≤3	W μ≤3
W9	W μ≤3	W μ≤3	S μ≥6	W μ≤3	W μ≤3
W10	W μ≤3	W μ≤3	W μ≤3	S μ≥6	M 3>μ>6
W11	W μ≤3	W μ≤3	W μ≤3	S μ≥6	W μ≤3
W12	W μ≤3	W μ≤3	W μ≤3	S μ≥6	M 3>μ>6
W13	W μ≤3	W μ≤3	W μ≤3	M 3>μ>6	S μ≥6
W14	W μ≤3	W μ≤3	W μ≤3	W μ≤3	S μ≥6
W15	W μ≤3	W μ≤3	W μ≤3	S μ≥6	S μ≥6
W16	W μ≤3	W μ≤3	W μ≤3	M 3>μ>6	S μ≥6

In order to observe the efficiency of the proposed procedure we substituted all 5 weak-ranked links in the AIET website structure (Fig. 3) by 5 strong-ranked links from the Table IV which were not included in the structure. After seven days of trial for the new structure, the average navigation time for the visitors decreased to 20 seconds which demonstrates an enhancement of more than 30 percent in AIET website system. Fig. 4 illustrates the graph of the new structure for AIET site.

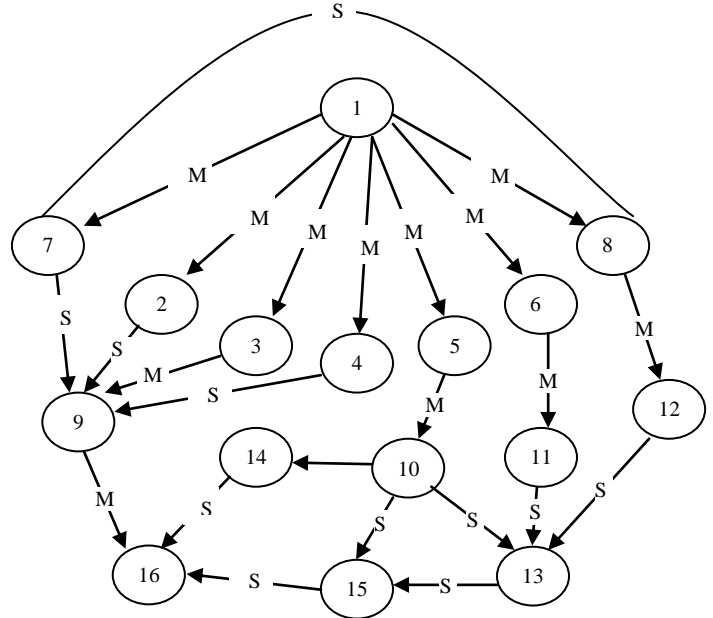


Figure-4: The directed graph of secondary AIET website architecture along with its links' strength

TABLE III
The fuzzy membership values for the relations between web pages

	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14	W15	W16	
W1		$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	
W2			$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	$\mu \leq 3$	$\mu \leq 3$	$\mu \leq 3$	$\mu \leq 3$	$\mu \leq 3$	$\mu \leq 3$	$\mu \leq 3$	$\mu \leq 3$	
W3				$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$\mu \leq 3$	$\mu \leq 3$	$\mu \leq 3$	$\mu \leq 3$	$\mu \leq 3$	$\mu \leq 3$	$\mu \leq 3$	
W4					$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	$\mu \leq 3$	$\mu \leq 3$	$\mu \leq 3$	$\mu \leq 3$	$\mu \leq 3$	$\mu \leq 3$	$\mu \leq 3$	
W5						$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$\mu \leq 3$	$3 > \mu > 6$	$3 > \mu > 6$	
W6							$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$\mu \leq 3$	$3 > \mu > 6$	$3 > \mu > 6$	
W7								$\mu \geq 6$	$\mu \geq 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$\mu \leq 3$	$3 > \mu > 6$	$3 > \mu > 6$	
W8									$\mu \geq 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$\mu \leq 3$	$3 > \mu > 6$	$3 > \mu > 6$	
W9										$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$3 > \mu > 6$	$\mu \leq 3$	$3 > \mu > 6$	$3 > \mu > 6$	
W10											$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	
W11												$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	
W12													$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	
W13														$\mu \geq 6$	$\mu \geq 6$	$\mu \geq 6$	
W14															$\mu \geq 6$	$\mu \geq 6$	
W15																$\mu \geq 6$	
W16																	$\mu \geq 6$

TABLE IV
The strength of relations between web pages

	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14	W15	W16	
W1		M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	
W2			S	S	S	S	S	S	S	W	W	W	W	W	W	W	
W3				S	S	S	M	M	M	W	W	W	W	W	W	W	
W4					S	S	S	S	S	W	W	W	W	W	W	W	
W5						S	S	S	S	M	M	M	M	W	M	M	
W6							S	S	S	M	M	M	M	W	M	M	
W7								S	S	M	M	M	M	W	M	M	
W8									S	M	M	M	M	W	M	M	
W9										M	M	M	M	W	M	M	
W10											S	S	S	S	S	S	
W11												S	S	S	S	S	
W12													S	S	S	S	
W13														S	S	S	
W14															S	S	
W15																S	
W16																	$\mu \geq 6$

5. CONCLUSIONS

Contents are the building blocks of the web pages in a website system and the degree of relation between pairs of content topics in addition to the intensity of relation between the content topics and each web page provide a sensible basis for deciding about the existence of a link between each pair of web pages. Using such a basis, a fuzzy method, applying minimal inference engine, was proposed to analyze the data gathered through two surveys from website visitors. The surveys collected the visitors' comments on

the intensity of relationship and dependency between the pairs of content topics in a website and also between each content topic and each webpage. In order to observe the proposed fuzzy method, a case study of AIET website was analyzed. Through the case study two aforesaid surveys were run and the gathered data were utilized by the proposed fuzzy method, the resulting link structure for the AIET website showed that our proposed method has an improvement of more than 30% in website structure efficacy, decreasing the time of navigation for website visitors from 29 seconds to 20 seconds on average.

In the majority cases, adding links increases the efficiency and efficacy of the website. However what determines which links to be added or substituted in a website structure can be confusing to the designers. Our proposed fuzzy approach can provide the designers with an appropriate method for ranking the links in a website which can excessively help designers to decide about which links to be omitted and which ones to be added to the website structure in order to enhance the visitors navigation effectiveness. It is obvious that applying such a fuzzy approach doesn't autonomously enhance the structure of the website, instead it can operate as a decision support system assisting the website designers to classify and rank the different links in their website architecture which can help them to take better decisions considering the existence and also the number of links in the website composition.

6. REFERENCES

- [1] Tiwari Raj Gaurang, Singh Hemant Kumar "Significance of State Cloning Concept in clustering clickstream" in the Proceedings of National Conference on Emerging Technologies in Computer Science (ETCS-2007) MIET, Meerut, India, Sep 23, 2007, pp – 227-233
- [2] Tiwari Raj Gaurang, Singh Hemant Kumar "Mining Interesting Knowledge and Pattern Discovery from Weblogs" in National Seminar Held at IJET, Bareilly, August 2007
- [3] Kim, W., Y.U. Song, and J.S. Hong, Web enabled expert systems using hyperlink-based inference. *Expert Systems with Applications*, 2004: pp. 1-13.
- [4] E.Rosen, D. and E. Purinton, Website design: Viewing the web as a cognitive landscape. *Journal of Business Research*, 2004(57): pp. 787-
- [5] Lee, J.-H. and W.-K. Shiu, An adaptive website system to improve efficiency with web mining techniques. *Advanced Engineering Informatics*, 2004. 18: pp. 129-142.
- [6] Benbunan-Fich, R., Using protocol analysis to evaluate the usability of a commercial web site. *Information & Management*, 2001(39): pp. 151- 156
- [7] Saudi, A. and M.H.A. Hijazi. Using similarity measures and association rule for web personalization. in *M2USIC*. October 2004. Malaysia. pp. 16-19.
- [8] Nakayama, T., H. Kato, and Y. Yamane. Discovering the gap between web site designers' expectations and users' behavior. in *The 9th Int'l World Wide Web Conference on Computer Networks*. May 2000. Amsterdam, Holland. pp. 811-822.
- [9] Arotaritei, D. and S. Mitra, Web mining: a survey in the fuzzy framework. *Fuzzy Sets and Systems*, 2004. 148: pp. 5-19.
- [10] Mitra, S. and H.L. Larsen, Special issue on web mining using soft computing. *Fuzzy Sets and Systems*, 2004. 148: pp. 1-3.
- [11] Vrazalic, L. Website usability in context: an activity theory based usability testing method. in *The national conference on Transformational Tools for 21st Century Minds*. 2003. pp. 41-47.
- [12] Blackmon, M.H., M. Kitajima, and P.G. Polson. Repairing usability problems identified by the cognitive walkthrough for the web. in *SIGCHI conference on Human factors in computing systems*. 2003. Florida, USA. pp. 497-504.
- [13] Huang, M.-H., *Web performance scale*, Information & Management, 2004.
- [14] Kim, K.-J. and S.-B. Cho, Personalized mining of web documents using link structures and fuzzy concept networks. *Applied Soft Computing*, 2005.
- [15] Gorzalczany, M.B., *Computational intelligence systems and applications: neuro-fuzzy and fuzzy neural synergisms*. 2002: Springer.
- [16] M. Husain, Q. Abbas, R. Gaurang Tiwari, "An Analysis Of Smart Data Fusion and Retrieval Strategies" accepted in proceedings of International Conference of Education, Research and Innovation (ICERI 2009) to be held in Madrid (Spain) on November 16-18, 2009