Effective User Navigation through Website using Dice Coefficient

Piyusha Dhalpe[1], Deeksha Bhardwaj[2]

[1,2]Department of Computer Engineering, GHRIET, Savitribai Phule, Pune University, Pune, India

Abstract: *In the website framework if the navigation is not efficient then user may worry, so it is very significant to enhance the user navigation. Performance of the websites relies on their information and functionality requirements of their users. The development of website is designed using needs of users. Thus it is a big task to build a web page with the efficient navigations. Hence unique method is suggested using customer navigations information to re-link the websites. Reorientation of whole web page may raise the complications for the acquainted user and it may irritate them. The restrictions of current system are it contributes all the applicant hyperlinks even if it is essential for individual user. Because of this the framework of web page enhancement has been improved. Thus we suggested the system which finds out the maximum hyperlinks which are necessary for the user. We suggested the system which is successfully signify the navigation of the websites. This system uses the new hyperlinks for helping the framework of websites. System used Dice's coefficient, we enhance the web page framework so that user can achieve to the focus on page quicker.*

Keywords: User navigation; website design; KNN; Dice's coefficient

1. Introduction

Web is used for letters, searching the details, information collecting, enjoyment, purchasing, and business. In this way, now a day's web is important part of our lifestyle. Nobody can think about lifestyle without internet. Web is generally used for organization purpose. Customer can do net financial, purchasing, reservation, billing from home without patiently waiting up long line. Usage of web for company purposed is extended. Thus, website for organization is essential to create business. At the point when developing a website, the routing of the website is essential in light of the fact that if client is not easily found desired details then benefit of the website is no usage. Navigation is best viewed as the system in which you discover and shift around the website. If client encounters problems in looking details then he results in the website and likes other website. So for getting business, efficient routing is essential. Organizations are more fascinated by efficient redirecting.

If there is a top great high quality of information yet the navigation is inadequate then that top great quality information is not beneficial. An important purpose of inadequate website settings is website is structured by web developer by developer's concerns and it might be not the same as clients' contemplations. More often than not there is not immediate conversation between web developer and customer so there is interaction gap among client and expert. So developer can't get appropriate concept regarding the client's wishes. In addition web professional makes the website by designer's objectives. Site adequacy depends on upon the satisfaction of the customers and not on how attractive website is structured. Properly methodized site means website should be structured out in such a way, to the point that comparison between its framework and clients wish of the framework is reduced. For efficient routing website should be designed by client's wishes with the objective that customer find out predicted outcome quickly. There are immeasurable customers, uses the web. Site furthermore has several web pages. The information is required for one customer may differ from other. At time several customers may implement the same website, and predicted webpages of each customer may differ. So it is not simple process to make highly efficient navigational website.

There are two techniques to improve website routing one way is Web customization and other is web modification. In web customization website is organized by considering individual clients or categories of details information and its actions on the redirecting. Also web modification indicates the website should be structured by considering all clients. In web transformation more most of the time website is structured in its framework. There are several drawbacks to change entire website. Completely change website may disorient nicely known clients. So our point is to improve website framework by least changes in its existing framework. This document is consisting further as: Section II speaks about related work analyzed until now. Section III provides implementation details, algorithm used and statistical design. It also contains trial installation. Section IV finishes with the conclusion and provides future work.

2. Related Work

For helping the user navigation with the knowledge mined from web server records and they can be usually classified in to web customization and web adjustment contacted. In this area we mentioned the relevant research done by the researchers for producing the efficient client the navigation system.

In [1], Ming-Syan Chen, Jong Soo Park and John p S. Yu suggested the remedy which find the novel information mining capability which engross traversal model in a allocated information offering atmosphere where records or things are connected together to accomplish entertaining accessibility. This remedy consist of two steps: in the initial step produced an algorithm for transforming the unique sequence of log information into a set of maximum ahead sources. In the second produced an algorithm for identifying

Paper ID: NOV153009

1353

the regular traversal styles i.e. sequence of huge sources or from the highest onward referrals obtain. These all two methods are developed for confirmative sequence of huge referrals from which one is based on the some of the hashing and pruning methods and other is enhanced with the choice of identifying the sequence in batch so as to decrease the number of database check out needed.

In [2], W.T. Fu and P. Pirolli signify a computational cognitive model which replicates how people search for information on the web. This design is termed as SNIF-ACT, which may be stands for scent based navigation and details foraging act in the structure. This method gives an account of how persons utilizes the information scent prompt, such as the text related with the web links, in order to make the navigation choices such as judging where to go upcoming on the website. In [3], Ms. Jissin Mary Kunjukutty, Ms. Apria, ME recommended the model which enhanced the web sites restructuring process. The system needs lowest site structure change. The method is helped by the high scalability. By utilizing the design based model navigability of the website was enhanced. This technique can be incorporated with web personalization problems.

In [4], Hung-Yu Kao, Shian-Hua Lin, Jan-Ming Ho and Ming-Syan Chen analyzed the issue of exploration the useful framework of information web websites which involve a large number of hyperlinked records. Writer explains the exposing design of information web websites as a set of index web pages and a set of a content web pages connected by this TOC web pages. On the reasons for web page link caused subject search criteria author suggested the procedure of entropy centered research for analyzing the entropy of core text messages and hyperlinks for removing the redundancy of the hyperlinked framework so that the complicated framework of the web page can be distilled.

In [5], Jose Borges and Mark Levene suggested a new technique for examining the capability of a varying length Markov design for outlining the customer the navigation period up to a given length. Although the summarization capability of a design is essential for allowing the recognition of customer the navigation patterns, the capability for creating forecasts is essential in order to estimate the up coming web link option of a customer after following a given pathway.

In [6], Chunyu, Yong Cao, Zaiqing Nie, Jie Zhou, and Ji-Rong Wen suggested a new framework WebNLP which allow bidirectional incorporation of web page framework knowing and written text knowing in a repetitive way. The writer used the system to the small business removal and Chinese removal and organization removal.

In [7], Jingyu Hou and Yanchun Zhang symbolizes the technique for discovering the appropriate web pages for a given website for this writer suggested two website link research based methods. One of the methods arrives from the prolonged cocitation research of the websites. One more is obtained from advantage of linear algebra concepts for exposing further connection among the websites for

determining appropriate web pages more perfectly and proficiently.

Min Chen and Young U. Ryu [8] have provided design which is suitable for websites whose materials are relatively constant over time that is informative sites. This design is targeted on helping the website so it is very helpful in keeping the websites. As this design procedures needless hyperlinks which impacts the efficiency of this design.

In [10], Scott Perkowitz, Oren Etzioni regarded problem of page indexing features and outperform the methods which based on recently clustering and conceptual clustering methods. The research outcomes of this strategy determine that top quality applicant catalog web pages can be produced instantly and this method outshine current methods.

In [15], Nancy De Marsico, Stefano Levialdi Implement the for determining a variety of websites which are supposed to be from identical group, for undertaking a comparative research of their feature. This technique is based for a two taken manufacturing of a type for the evaluation of that type of websites. In [16], Gek Woo Tan, Kwok Kee Wei examine behavior of customer Web page as a way for knowing Web page developing by utilizing a think loudly method exploration. Main theoretical efforts of this analysis are the shown of the circulation of intellectual procedures during the Web page surfing around and the organization of Web design measurements the significance and material indirect by Website content, its external type, and the framework and redirecting in regards to user show.
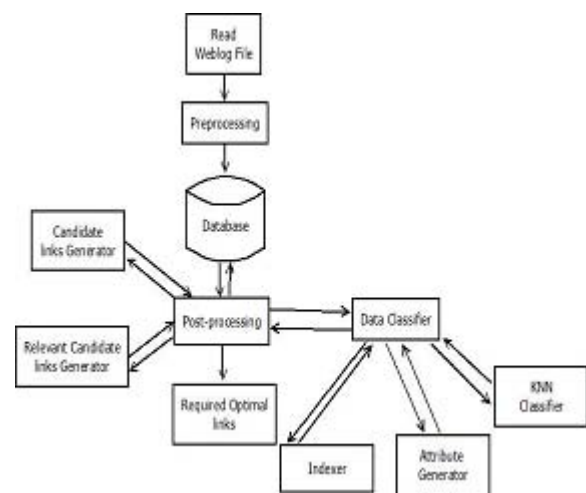
3. Implementation Details

Figure 1: System Architecture

Above diagram shows the system representation of the proposed system. Now we discussed the module by module description of the proposed system.

1) Read File:
This is the first step of the suggested system. In this component user study the dataset file from disk or storage space.

2) Preprocessing:
In preprocessing chosen dataset is study and needed information like the user id, time, date and page utilized by user are divided and saved in database.

3) Post Processing:
The information is study from the database and small sessions are produced. According to that small session, applicant hyperlinks are produced with the threshold principles, with threshold value1 and with threshold value 2.

4) Data Classifier:
By using dice's coefficient indexing of every hyperlink are done. According to the threshold value, the hyperlinks are marked as irregular hyperlinks and normal hyperlinks and lastly with the help of KNN maximum hyperlinks are produced.

A. Algorithm
Algorithm 1 Proposed System Algorithm
1: Website is represented as a directed graph.
2: Nodes representing pages and arc representing links.
3: n is the number of pages of the website. User may visit more than one page by accessing links during user session.
4: P_i = {P1, P2... Pm} be the set of m link accessed by user Ui where i= {1, 2... m}.
5: User Ui looking for the page Pm. Means Pm is the target page. Let U= {U1, U2... Un} be the set of N users.
6: C= {C1, C2... Cp} be set of p candidate links that need to redesign and re link.
7: We use Dices coefficient index. Dices coefficient index is used for checking similarity between two string set.
8: The aim of this paper is to identify links that are used to redesign.

B. Set Theory
S={s, e, X, Y}
Where
S= System
s= Distinct start of system
e= Distinct end of system
X= Set of Inputs
Y= Set of Outputs
1) Input Process
Let X is the input Where X={X1, X2. . . Xn} Such that their exist function
X1= {U, D, t, P} Where
U= {U1, ..., Un} is a set of users D= {YYYY-DD-MM} is date
t= {hour, minute, second} is time
P= {P1. . .Pn} is a pages visited by user
2) Create Session
E= {E1.....En}
Where E is a set of Session and E1.....En is number of Session
3) Target Pages
Ti = {T1 . . . , Tn} is set of target pages of ith user
4) Generate Candidate Links
Generate Candidate Links without threshold
Q= {fL1 . . . ..Ln}
Generate Candidate Links with threshold value 1 (path)
W= {Wl. . .Wn}

Generate Candidate Links with threshold value 2 (path)
I= {Il. . . In}
5) Existing Links
G= {G1....Gn}
G is set of the existing links and G1....Gn is number of existing link.
6) Generate New Links
K= G-Q
Where K is a set of new links and G is a set of existing links and Q is a set of Candidate Links
7) Dice coefficient
D (i) = (2|AB|) = (|A| + |B|)
Here in dice coefficient, we give the W1 that is link from the Candidate Links with threshold value 1 and K that is a set of new links as a input to the Dice coefficient to calculate index value. Take, threshold=0.20
For normal links 0.20¡ index value
N= {N1...Nn}
Where, N is a set of normal links and N1 ...Nn number of normal links.
For Anomaly Links, 0.20 > index value
A= {A1 ...An} Where, A is a set of anomaly links and A1...An number of anomaly links.
8) KNN Classifier
In KNN classifier, N that is the set of normal links and A that is the set of anomaly links take as an input and perform both training and testing process simultaneously.

C. Mathematical model for proposed work
For improving the user navigation in the websites the following equation are as follows:

Minimize $\sum i; j\varepsilon Exi; j [1 - \lambda ij (1 - \varepsilon)] + m\sum_{i\varepsilon Ne} P_i$ which is subject to

$$C_{kr}^S = \Sigma_{(i,j)\varepsilon E} a_{ijkr}^S X_{ij}; r = 1, 2.... L_p(K, S) . . . (1)$$

$$K = 1, 2, ... , L_m(S), \forall S\varepsilon T^R$$

$$\Sigma_{k=1}^{b_j} \Sigma_{r=1}^{L_p(k,S)} C_{kr}^S \geq 1; \forall S\varepsilon T^r, j = tgt(S) .... (2)$$

$$\Sigma_{j:(i,j)\varepsilon E} x_{ij}(1 - \lambda_{ij}) + W_i - p_i \leq C_i; \forall i\varepsilon N_E .... (3)$$

$$x_{ij}\varepsilon\{0,1\}, p_i\varepsilon\{0\} \cup Z^+, \forall(i,j)\varepsilon E, i\varepsilon N_E .... (4)$$

Table 1: Notation Table

Notation	Definition
C_{kr}^S	1 if in mini sessions S, a link from rth page in kth path to the target is selected; 0 otherwise.
p_i	The number of links i.e. exceed the outdegree threshold
C_i	The outdegree threshold for page i
a_{ijkr}^S	1 if the rth page in the kth path; otherwise 0
x_{ij}	1 if link from page i to j is selected otherwise 0
C	The outdegree threshold for page i
E	The set of candidate links which can be selected for improving user navigation.
T^R	The set of relevant mini sessions, i.e., mini sessions that need to be facilitated for given path threshold.
x_{ij}	1 if the link from page i to j is selected; 0 otherwise
λ_{ij}	1 if link i has a link to page j in the current structure; 0 otherwise
N_E	The set of source nodes of links in set E.
b_j	The path threshold for mini sessions in which page j is the target page.
tgt(S)	The target page of mini session S
m	Multiplier for the penalty term in the objective function.

4. Result and Discussion

A. DataSet

In the proposed system we utilized SQL yog for storing the data of user navigation.

B. Results

In this region we mentioned the outcome acquired for the user the navigation system. Here we evaluate our system on for direction threshold and enhanced hyperlinks. Path threshold indicates, user should reach the focus on page based on the threshold principles i.e. 1, 2 ...etc. Improved hyperlinks indicate the links in current websites which have to be modified for helping the redirecting.

We evaluate the outcome from the current system which is based on mathematical development. In the current system mathematical design is applied to enhance the web page routing which provides greater variety of hyperlinks. In the suggested system the Dice's coefficient is utilized to enhance the web page navigation which gives the less variety of hyperlinks as opposed to current system.

In the following table 1 improved link with time threshold 1 min. We display the outcome for both by using mathematical programming and by utilizing the Dice's coefficient.

Table 2: Improved link with time threshold 1 min

Sr no	Using Mathematical programming	Using Dice's Coefficient
1	1829	1627
2	679	590
3	267	218

The below graph describe the relation between path threshold numbers of improved links with time threshold 1 min. In x axis shows the path threshold and in y axis shows the number of improved links. In the following table 2 improved links with time threshold 2 min. We show the result for both by using mathematical programming and by using the Dices coefficient
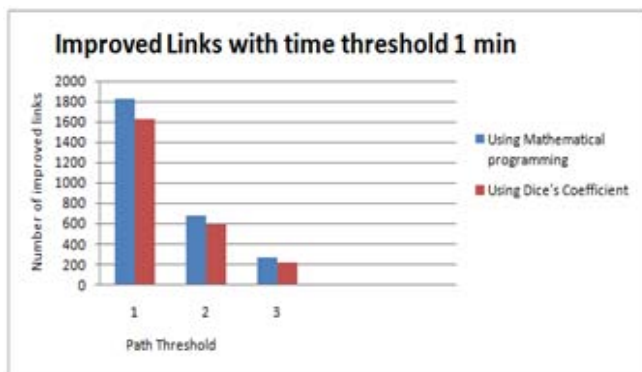

Figure 2: Graph of improved links with time threshold 1 min

Table 3: Improved link with time threshold 2 min

S No	Using Mathematical programming	Using Dice's Coefficient
1	1751	1540
2	703	604
3	313	217

The below graph represents the relation between path threshold numbers of improved links with time threshold 2 min. In x axis represents the path threshold and in y axis shows the number of improved links.
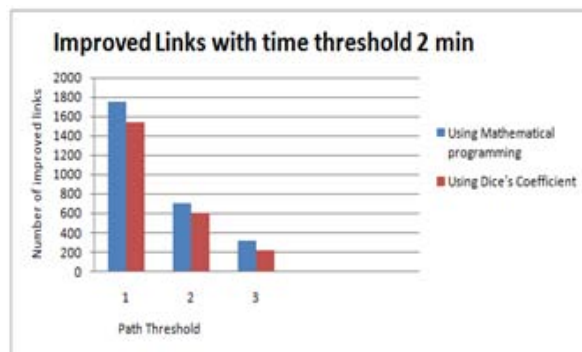

Figure 3: Graph of improved links with time threshold 2 min

5. Conclusion

A significant cause of inadequate web design is that the web designer thoughtful of how a web page should be organized can be significantly unique from those of the users. In the current system the methods which are applied are prolonged Cocitation and hidden linkage information algorithm is utilized for verdicting semantically appropriate web pages but these algorithms require number of variations in web page framework. In the current used mathematical development, this system has not produced effective maximum hyperlinks. For get over the restriction of the current system we suggested the system for helping the framework with the lowest variations in current framework of a websites. The suggested design enhances the web page rather than change the sites and hence it is beneficial for web page maintenance. We applied Dices coefficient for helping the framework of maximum link, and helping the framework of user navigation.

6. Acknowledgment

References

[1] Ming Syan Chen, Jong Soo Park and Philip S. Yu, "Efficient Data Mining for Path Traversal Patterns", IEEE Transactions on Knowledge and Data Engineering, Vol. 10, No. 2, March/April 1998 209.
[2] W.T. Fu and P. Pirolli, "SNIF-ACT: A Cognitive Model of User Navigation on the World Wide Web," Human Computer Interaction, vol. 22, pp. 355-412, 2007.

[3] Ms. Jissin Mary Kunjukutty, Ms. Apria, ME, Optimizing User Navigation with Pattern based Web Site Restructuring Scheme, International Journal of Innovative Research in Computer and Communication Engineering, Vol.2, Special Issue 1, March 2014.

[4] Hung-Yu Kao, Shian-Hua Lin, Jan-Ming Ho and Ming-Syan Chen, Mining Web Informative Structures and Contents Based on Entropy Analysis, IEEE Transactions on Knowledge and Data Engineering VOL. 16, NO. 1, JANUARY 2004.

[5] Jose Borges and Mark Levene, Evaluating Variable Length Markov Chain Models for Analysis of User Web Navigation Sessions, IEEE Transactions on Knowledge and Data Engineering, VOL. 19, NO. 4, APRIL 2007.

[6] Chunyu Yang, Yong Cao, Zaiqing Nie, Jie Zhou, and Ji-Rong Wen, "Closing the Loop in Webpage Understanding" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 5, MAY 2010.

[7] Jingyu Hou and Yanchun Zhang, "Effectively Finding Relevant Web Pages from Linkage Information", IEEE Transactions on Knowledge and Data Engineering, VOL. 15, NO. 4, July/August 2003.

[8] Min chen, Young U. Ryu "Facilitating Effective User Navigation Through Web Structure Improvement", IEEE Transactions On Knowledge And Data Engineering, Vol. 25, No. 3, March 2013.

[9] 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.

[10] Mike Perkowitz and Oren Etzioni, "Towards adaptive Web sites: Conceptual framework and case study", Artificial Intelligence 118 (2000) 245-275, 28 July 1999.

[11] Fang, Xiang, and Clyde W. Holsapple. "An empirical study of web site navigation structures' impacts on web site usability." Decision Support Systems43.2 (2007): 476-491.

[12] Ivory, Melody Y., and Marti A. Hearst. "Improving web site design." Internet Computing, IEEE 6.2 (2002): 56-63.

[13] Applen, J. D. "Tacit knowledge, knowledge management, and active user participation in Web site navigation." Professional Communication, IEEE Transactions on 45.4 (2002): 302-306.

[14] Srikant, Ramakrishna, and Yinghui Yang. "Mining web logs to improve website organization." Proceedings of the 10th international conference on World Wide Web. ACM, 2001.

[15] Maria De Marsico, Stefano Levialdi, "Evaluating web sites: exploiting user's expectations", Elsevier Received 19 February 2003; accepted 24 October 2003

[16] Gek Woo Tan, Kwok Kee Wei, "An empirical study of Web browsing behavior: Towards an effective Website design", Electronic Commerce Research and Applications 5 (2006) 261-271