AN APPROACH TO EVALUATE QUALITY OF WEBSITE STRUCTURE

Sreedhar G.[1], Chari A.A.[2], Venkata Ramana V.V.[3]

[1]Sr. Lecturer, Dept. of Computer Science, Rashtriya Sanskrit Vidyapeetha (Deemed University), Tirupati
[2]Professor, Dept. of OR & SQC, Rayalaseema University, Kurnool
[3]Scientist-E, National Informatics Centre, Hyderabad

Email: [1]gsrid74@yahoo.com, [2]chari_anand@yahoo.com, [3]ramana@gmail.com

Abstract

The primary goal of this paper is to identify the elements involved in quality assessment of website structure. Always quality of website structure is based on its navigability, average number of clicks and structural complexity. The web tool PowerMapper is used to establish the sitemap for the website, path length metric is used to evaluate average number of clicks to get desired web page and web site structural complexity is determined with cyclomatic complexity. The quality of web site structure is measured in 10point scale and scale is formulated by considering various levels of website tree structure. The status of web site structure is determined in 10 point scale and the value suggests the improvement of the site structure.

Key words: sitemap, path length, navigability, broken link error index, cyclomatic complexity, web site structure

I. INTRODUCTION

A website is a collection of web pages containing text, images, audio and video etc. Today, Web is not only an information resource but also it is becoming an automated tool in various applications. Due to the increasing popularity of web, one can be very cautious in designing the Website. Poor and careless web design [1] leads to hard ship to public utility and does not serve the purpose. Navigation places crucial role in the design of website structure because it determines the path to be traveled to reach a required web page. Normally a website structure resembles tree like structure starting from home page as root. In designing a website tree structure one has to concentrate in breadth and width sizes of tree. The home page of website must be designed in such a way that it should not be too much crowded with links and also it should not be too much empty. It was stated that number of links on each page should not exceed 20 links [4] and average number of clicks should not be more than 4 clicks to get a required page. Keeping view of this the developer must be very much careful in designing the website structure.

II. RELATED WORK

The quality of website can be assessed mainly in functionality and usability. The qualitative measures [5], [6] are used to achieve quality in functionality of website. The notion of usability is a key factor to interact a website. The efficiency of usability is depended on website structure. The structure of website [7] should be in such a way that user can easily interact website without any formal training. An effective web design [8] is one that makes it easier for users to navigate through the different pages on the site. The website structure [9] is represented by directed graph where each node represents a web page and edges represent link to corresponding web pages. It is already investigated that web link structure can also be used for page ranking [10] and web page classification [11]. These works stress on the navigational relationship among web pages. Our present work focus on discovering the various elements involved in website structure and thus we can concentrate on these elements to improve quality of a website structure.

III. METHODOLOGY

The procedure for the quality assessment of website structure involves four modules: creating a sitemap, computing path length metric, evaluating structural complexity of website, and finding broken link error index. All these modules are included in a web program.

A. Creating a sitemap:

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

B. 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 number 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.

$$\text{Path length} = \sum l_i \cdot m_i \qquad (1)$$

where l_i is the level number l, m_i is the number of nodes at level i.

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

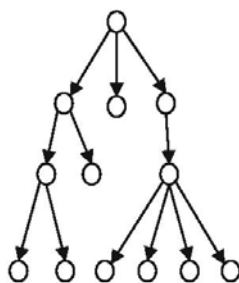where n is the number of nodes in the tree. An example tree is shown in figure 1.

Fig 1. A tree with 3 levels

Fig. 1. A tree with 3 levels

Level 0

Level 1

Level 2

Level 3

$$\text{Pathlength} = 0 \times 1 + 1 \times 3 + 2 \times 3 + 3 \times 6 = 27$$
$$\text{AvgClicks} = 27/13 = 2.07$$

C. 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 2.
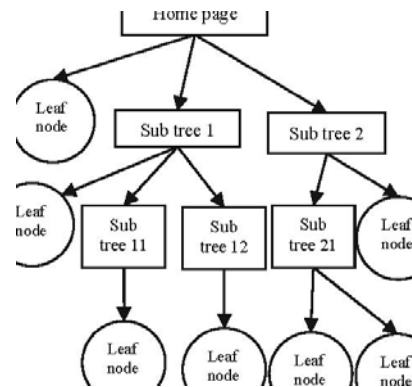

Fig 2. Tree graph for a website

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 represent a web page which do not have any further links to the 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 structural complexity is computed using cyclomatic complexity and it is calculated using equation (3). According McCabe value of Cyclomatic Complexity Metric (CC_{Metric}) should not exceed 10.

Cyclomatic Complexity Metric

$$(CC_{Metric}) = (en + d + 1)/n \qquad (3)$$

where e is the number of web page links, n is the number of nodes in the graph and d is the number of leaf nodes in the graph.

D. Broken link calculation:

In a website structure, navigation problem raised due to broken links. The broken link may involve at

various levels of web site structure. Broken links are identified in sitemap by the PowerMapper web tool. In a sitemap of website, broken links are counted at various levels and broken error index is calculated based on percentage of broken links involved in sitemap tree. The percentage of broken links [3] is calculated using equation (4).

Percentage of Broken links (PBL) = (number of broken links/number of web pages) $\times$ 100   (4)

IV. EVALUATION

The websites of more than 50 Indian universities are considered in the evaluation process. The web program accepts address of each university's website and generates sitemap using PowerMapper web tool. The sitemap consists of all web pages of the website and displays in hierarchical tree structure in various levels. The average number of clicks needed to access a web page is computed with path length metric using equation (1) and (2). The cyclomatic complexity value is computed for the web site structure using equation (3). The percentage of broken links is calculated at each level in the sitemap using equation (4). The quality of website structure for each university is evaluated in 10point scale. The 10point scale value for each university is based on organization of web pages in sitemap structure, cyclomatic complexity of website, average number of clicks and broken link error index. The 10 point scale value for sitemap structure is evaluated using SMP_{Metric} shown in equation (5).

$$SMP_{Metric} = \begin{cases} 10 & \text{if webpage is homepage and} \\ & wps = n \ \& \ wps \le 20 \\ 10 & \text{if } wps \le 20 \\ & wps \le 20 \\ 9 & \text{if } wps = 9 \text{ or } 21 \\ 8 & \text{if } = 8 \text{ or } 22 \\ 7 & \text{if } wps = 7 \text{ or } 23 \\ 6 & \text{if } wps = 6 \text{ or } 24 \\ 5 & \text{if } wps = 5 \text{ or } 25 \\ 4 & \text{if } wps = 4 \text{ or } 26 \\ 3 & \text{if } wps = 3 \text{ or } 27 \\ 2 & \text{if } wps = 2 \text{ or } 28 \\ 1 & \text{if } wps = 1 \text{ or } 29 \\ 0 & \text{otherwise} \end{cases}$$

where wps is the number of web page links in a sub tree of sitemap

The 10 point scale value for structural complexity of website is calculated using equation (6) and (7).

$$CC_{Metric} = \begin{cases} 10 & \text{if } CC \le 1 \\ 9 & \text{if } CC \le 2 \\ 8 & \text{if } CC \le 3 \\ 7 & \text{if } CC \le 4 \\ 6 & \text{if } CC \le 5 \\ 5 & \text{if } CC \le 6 \\ 4 & \text{if } CC \le 7 \\ 3 & \text{if } CC \le 8 \\ 2 & \text{if } CC \le 9 \\ 1 & \text{if } CC \le 10 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$$10_{PValue} \ \text{Avg} \ (MP_{Metric} + CC_{Metric}) \quad (7)$$

The Web Page Click (WPClick) value is determined using equation (8) and 10 point scale value is adjusted according to WP_{Click} value.

$$WP_{Click} = \begin{cases} 0.75 & \text{if AvgClicks} \le 2.5 \\ 0.5 & \text{if AvgClicks} \le 4 \\ 0.25 & \text{if AvgClicks} \le 5 \end{cases} \quad (8)$$

$$10_{PValue} = 10_{PValue} + WP_{click} \quad (9)$$

The Broken Link Error Index (BLE) is calculated to determine the errors in sitemap structure and it is determined using equation (10) and its value is deducted from the value of 10point scale using equation (11).

$$BLE = \begin{cases} 1 & \text{if PBL} \ge 10 \\ 0.75 & \text{if PBL} \ge 5 \\ 0.5 & \text{if PBL} \ge 2 \\ 0.25 & \text{if PBL} > 0 \end{cases}$$

$$10_{PValue} = 10_{PValue} - BLE \quad (11)$$

As an example, the sitemap of Aligarh Muslim university website is shown in the figure 3. The evaluation of value of 10 Point Scale value for Aligarh Muslim University website is shown in table 1.

Table 1: 10 Point Scale Value Calculation for Aligarh Muslim University

| Avg NoClicks	Path Length = 367			
	Avg No Clicks = 367/156 = 2.352564			
SMP_{Metric}	Level No	Sub tree in website structure	No. of web pages in sub tree	10-Point Scale Value
	1	2	20	10
		1	3	3
		2	3	3
		3	1	1
		4	1	1
		5	4	4
		6	3	3
		7	4	4
		8	7	7
		9	8	8
		10	3	3
		11	2	2
		12	1	1
		13	4	4
	2	14	16	10
		1	7	7
		2	1	1
		3	5	5
		4	4	4
		5	1	1
		6	1	1
		7	1	1
		8	1	1
		9	1	1
		10	2	2
		11	9	9
		12	11	10
		13	7	7
		14	13	10
	3	15	9	9
		1	1	1
	4	2	1	1
	10-point scale value = 4.21875			
CC_{Metric}	Cyclomatic Complexity Metric $(CC_{Metric}) = (e - n + d + 1)/n$			
	$= (156 - 20 + 28 + 1)/20 \ = 2.083333$			
	10-point scale Value = 8			
BLE	% of broken links		Broken link error index	
	5.769231		0.75	
10-point scale value for Aligarh Muslim University website = 5.4				
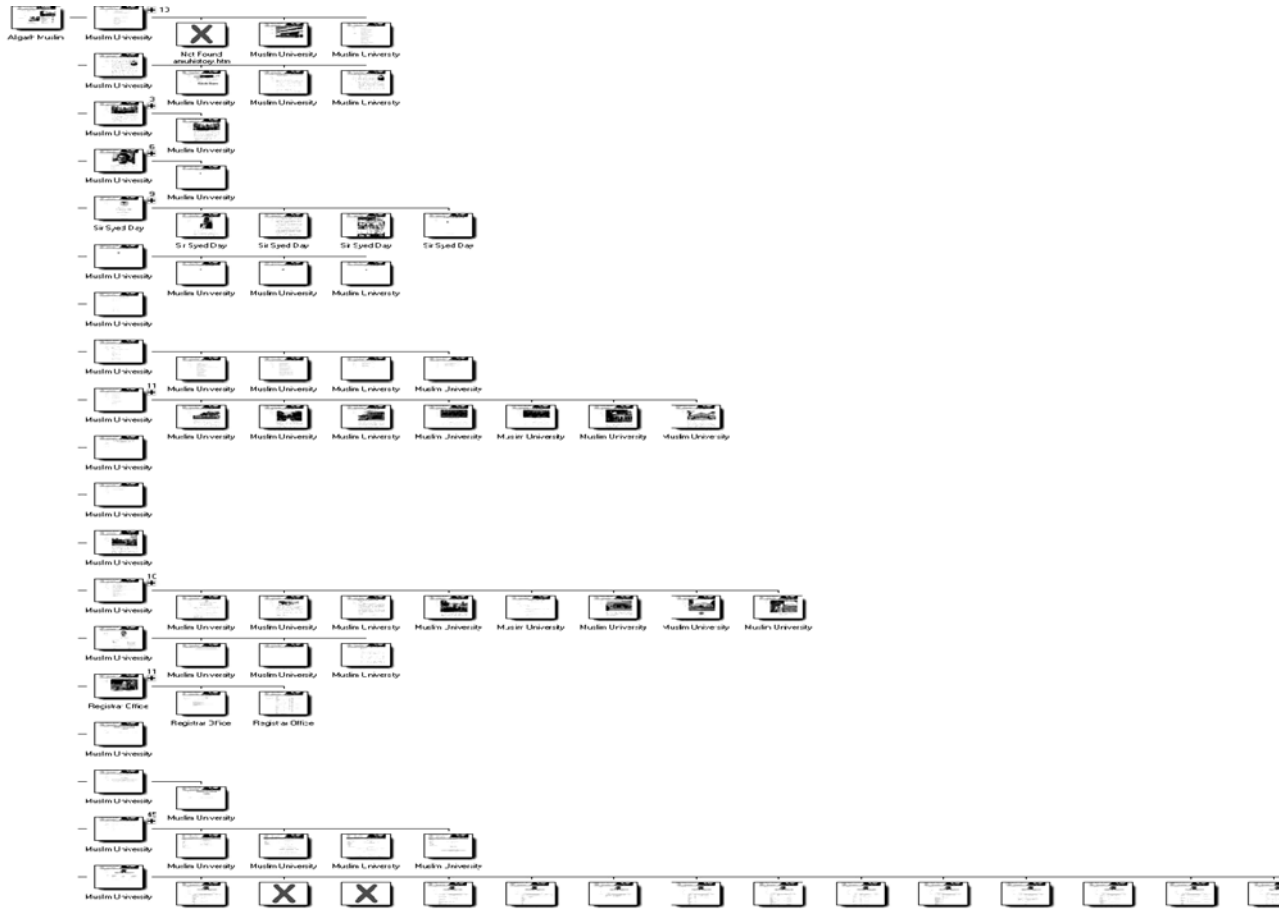
Fig 3. Sitemap of Aligarh Muslim University Website

Website Table 2: 10 Point Scale values for various universities' websites in India

S.No.	University Name	SMP_{Metric}	CC_{Metric}	10_{PValue}	WP_{click}	BLE	10-point scale value	Remarks
1	Aligargh Muslim University	4.21875	8	6.109375	0	0.75	5.4	Needs improvement
2	Assam University	2.555556	7	4.777778	0.25	0.5	4.5	Very Poor design
3	Dravidian University	5.666667	3	4.333333	0.25	0	4.6	Very Poor design
4	Bharathiar University	2.55	6	4.275	0.25	0.5	4	Very Poor design
5	Bharathidasan University	4.380952	8	6.190476	0.5	0.75	5.9	Needs improvement
6	Bhavnagar University	2.833333	8	5.416667	0.75	0.25	5.9	Needs improvement
7	Sri Venkateswara University	3.058824	9	6.029412		0.75	0. 6.8	Needs improvement
8	Rashtriya Sanskrit University	2.365834	6. 4.182927	0.75		1	3.9	Very Poor design
9	Madurai Kamaraj University	4.346154	5	4.673077	0.75	1	4.4	Very Poor design
10	Guru Jambeshwar University	3.352941	6	4.676471	0.75	0.5	4.9	Very Poor design

V. CONCLUSION

The main objective of the work is to evaluate quality of website structure and suggesting the importance of navigability of website. In this paper an attempt is made to find key components like sitemap evaluation, average number of clicks per page cyclomatic complexity and broken link error index which greatly influences quality of website structure. The web developer must concentrate these components in selection and design of website. We can further extend this work to identify other factors of website design in quality assessment which would further enable to improve the design as a part of the ideology of TQM which emphasizes the continous improvement of design aspect and promote excellence of web design.

REFERENCES

[1] Folders, Vincent and Michel Will, "Web Pages that suck: Learn Good Design by looking at Bad Design", SanFrancisco, CA, SYBEX.

[2] Yanlong Zhang, Hong Zhu and Sue Greenwood, "Website Complexity Metrics for Measuring Navigability", Proceedings of the fourth conference on quality software (QSIC'04), 0-7695-2207-6/04, IEEE.

[3] www.websiteoptimiztion.com

[4] Benjamin Yen, Paul Jen-Hwa Hu, May Wang, "Toward and analytical approach for effective Website design: A frame work for modeling, evaluation and enhancement", Electronic Commerce Research and Applications 6 (2007), 159-170.

[5] G. Sreedhar and A.A. Chari, "An experimental Study to Identify Qualitative Measures for Website Design", Global Journal of Computer Science and Technology, University of Wisconsin, USA, September, 2009, pp.12.

[6] G. Sreedhar, A.A. Chari and V.V. Venkata Ramana, "Evaluating Qualitative Measures for Effective Website De

[7] BenbunanFich, R, "Using Protocol Analysis to evaluate the usability of a Commercial Website, Information and Management, 39, 151163, 2001.

[8] Ali Azad, "Elements of Effective Web Page Design", Global Competitiveness, January, 2001.

[9] Zheng Chen, Shengping Liu, Liu Wenyin, Geguang Pu, WeiYing Ma, "Building a Web Thesaurus form Web Link Structure", SIGIR 2003, July 28 – August 1, 2003, Toronto, Canada.

[10] L. Page, S. Brin, R. Motwani and T. Winograd, "The PageRank Citation Ranking: Bring Order to the Web", Technical Report, Stanford University, 1998.

[11] E. Glover, K. Tsioutisiouliklis, S. Lawrence, D.Pennock, G. Flake, "Using Web Structure for Classifying and Describing Web Pages", in Proceedings of WWW2002, Hawaii, May 2002.

Mr. G. Sreedhar is pursuing Ph.D in Computer Science from S.K.U, Anantapur. He received M.Phil (Computer Science) from Alagappa University, Karaikudi, T.N, India and M.C.A from Sri Krishnadevaraya University, Anantapur, A.P, India in 2005 and 1998 respectively. He is working as a Sr. Lecturer in Dept. of Computer

Science at Rashtriya Sanskrit Vidyapeetha, Tirupati since 2001. He has over 10 years of teaching and research experience in the field of Computer Science.