International Journal of Advanced Trends in Computer Science and Engineering

Computing Quality of Structure of a Web-Site

B. Vishnu Priya[1], Dr. JKR Sastry[2]

[1]Koneru Lakshmaiah Education Foundation, Vaddeswaram, India, vishnupriya574@gmail.com

[2]Koneru Lakshmaiah Education Foundation, Vaddeswaram, India, drsastry@kluniversity.in

ABSTRACT

Information dissemination is the key. WEB sites play a vital role in this area. Most of the businesses are being conducted world over using websites. The effectiveness of the businesses conducted is dependent on the quality of web sites. The quality of a web site is dependent on many factors. The computation of the quality of a WEB site is rather complex. The quality of each factor computed individually, and then the total quality of a web site is to be computed. The quality of a web site is dependent on the quality of the structure using which the content displayed to the user.

In this paper, a data mining approach has been presented that determines all the hidden structures within the WEB site, and then a method is presented using which quality of the structure of the website computed. The method presented tested on a live website and weaknesses of the site found and forwarded the same to the site owners for their site correction.

Key words: Assessing quality, WEB sites, Quality Assessment Model, Mining Structures, WEB structures, quality of web structures

1. INTRODUCTION

The advents of the internet and the WEB have become sources heavily used for information exchange. People are using the WEB site looking for content of their requirements. There are no bounds to the content rendered on the WEB site. Any information required is available on the net these days. People are getting the required information in no time, and therefore the content posted on the WEB has become purposeful. Many forms of information posted on the WEB, which includes images, videos, audios, text, and the data presented using different widgets. The information is available in static, dynamic, and streaming mode. The style of living by the people has changed heavily, with WEB being the main resource used for everything. People use the content with the belief that the content hosted on the WEB is correct and accurate. People get dissatisfied when they come to know that the content hosted on the web is incorrect or inaccurate or lacks quality — people lose interest when one needs to do too much navigation of the website. People won't visit the web sites if the quality of the website is poor, leading to making the websites dormant.

Quality of a website is dependent on many factors such as completeness, look and feel, content, structure, multimedia, usability, navigation, etc. The quality of a web site is also related to reliability, maintainability, security, privacy, personalization, adequacy, safety, etc. Each feature of the WEB site must be assessed to determine quality. The overall quality of the web sites assessed considering the quality of all factors/features of the website. Each factor sometimes has to be decomposed further into the sub-factor, and the quality of each sub-factor needs assessing and combined to get the overall quality of the factor/feature of the WEB site. The overall quality of the website assessed combining the quality of individual factors.

One can use many methods for computing the quality of a sub-factor, which include subject, objective, or commutative assessments, while there is an element of prejudice in the subjective assessment, and sometimes there is incompleteness in the objective assessment of the quality of the websites. Computational assessments require proper empirical formals and quality models, which maps to discrete values representing the quality of the web sites. Some of the methods presented in the literature are purely based on the preferences of the Individuals or based on statistical measurements relating to response time, downtime, etc. which do not address the user concerns.

The way the quality of a web site is perceived differs from user to user. Programmers look at the quality of a web site from the perspectives of re-usability, maintainability, functionality, while actual users look at the quality of the WEB site based on look and feel, the accuracy of the content, navigability, personalization, etc. The complexity of assessing the quality of the web site is increasing with the increase in the size and depth of the web site. It becomes complicated if the navigational depth of web site increases drastically. The web site that implements eCommerce, animations, video streaming is quite complicated, and assessing the quality of such websites is quite complicated. The assessment of the quality of a website is complicated when the factors/features of the web sites are either interconnected or inter-leaved. It is necessary to select a suitable method for assessing the quality of the website. The methods used greatly vary, which may be subjective, objective or quantitative. Building a composite model of assessment of the quality of a web site is complicated when several methods of different types used.

Every user has his/her expectations from the website they attempt to surf. If the expectations met, then the users are satisfied and have a good rating of the site. One must find the quality factors that are relevant concerning the expectations of the user. The suitable and matching quality factors would become a base for computing the quality of the Website concerned. Assent methods suitable for computing the quality of a factor must then be selected

Many factors that match the user requirements considered and the quality of the same assessed. As said earlier there are many kinds of factors that exist, some of which will be relevant for a specific website. The factors generally considered include multi-media, content, usability, navigation, structure, etc. In this paper, the way the quality of the factor "Structure" is computed using a mining algorithm.

2. LITERATURE SURVEY

Several authors considered different factors and the methods required to assess the quality of those factors. Miss. Kausar Fiaz Khawaja1 et al., [1] have considered the factors that include appearance, usability, security, and privacy for computing the quality of websites. They have defined each factor and have shown the way the quality of the factors computed. They have defined the factor "usability" as a factor that is related to the learnability of the user. The factor expresses the user's experience while using the web site.

The factor that relates to appearances is concerned with the look and feel of the web site. The look and feel of the WEB site expresses the quality of presentation and appealing of the web site. The factor "adequacy" relates to the sufficiency and usefulness of the content hosted on the web site. Miss. Kausar Fiaz Khawaja et al., [1] have presented methods that use the observations of the user for computing the quality of the web sites. The observations of the users are collected while the user is surfing the website.

A different set of factors have been used by Vijay Kumar Mantri et al. [2], which include usability, safety, and flexibility for computing the quality of the web sites. They have presented that the factor "usability" is concerned with efficiency and effectiveness with which users can interact with the WEB sites and the extent to which the users are satisfied. They have also presented that the factor "safety" related to the extent to which the interaction of a user with the web site revealed. A 100% quality site will not allow revealing of any of the interaction of a user with the web site. The ability to add modifies, and delete the functionality of the web site dynamically without affecting the ongoing system is concerned with flexibility. PoDQA (Portal data quality assessment) – a tool is used by the authors for assessing the quality of the three factors that they have considered.

Vassilis S. Moustakis et al., [3] have considered several factors for computing the quality of the web site that includes Design, Content, Navigation, Uniqueness, Multimedia, and Appearance. They have defined that information presented to the end-user regarded as "content." The quality of content represented the extent of specialization and generalization, reliability, and completeness of the information. The factor "Navigation" is about the way the users can move around the web site, proper functioning of the links, and completeness of connectivity of the links.

The structure of a web site is related to the way the content presented, and the speed with which the website accessed. Use of multimedia, 2D/3D graphics, the display of objects used for presenting the content, the color selection, and composition, etc. related to the looking and feels of the web site are the factors considered by them. The quality of a web site lies in the uniqueness of the site easily differentiated from others. They have not, however, proposed any method used for computing the quality of the web site, especially concerning the structure of the WEB site.

Andrina Graniü et al., [4] considered the factor "portability" considered as one of the most important factors. Portability is the ability to move the web site from one platform to the other. But this kind of requirement rarely occurs and therefore is not a factor considered as a part of quality assessment of the websites.

The quality factors Privacy, Security, adequacy of information, Usability and appearance of the websites have been considered by Tanya Singh et al. [5], Usability is related to learnability and user experience gained while surfing the WEB site, Keeping the confidentiality of the content as per the user preference is called privacy. The quality of the web site can be measured based on the extent of preserving the privacy of the user. Securing is all about protecting the data while in transit from client to server. No unintended users shall have access to the content transmitted from the client to the server and vice versa. The quality of the web site, in this case, is the extent to which the content protected by making available the content to the intended users only. Making available data that meets the requirements of the end-users without any loose ends is called the adequacy of data. Use relevant and appropriate colors, graphics, and sequences that are related to the "Appearance" quality factor. The authors, however, have not presented any methods that compute the quality of the WEB site, considering any of the factors that they have discussed.

The factors that include Reliability, portability, usability, maintainability, functionality, and efficiency are considered by Anusha et al., [6] for computing the quality of the website. The factor reliability measures the consistency of making available the same content when queries surfed many times. Reliability measured as the probability that the same page presents the same content, in the same way, every time the content requested. Every website built with functions are application-specific or the functions relating to privacy and security. The more functionality covering every aspect of a website more will be the quality of the website.

They have defined the factor usability as ease of using the website and also as the learnability of an end-user. The factor maintainability defined as simplicity with which changes can be made to the website, especially when the website is in running state. The factor maintainability includes other sub-factors, which include changeability, and testability. They have introduced another quality factor called "Analysability," which means the ability to read the content while tracing the navigational paths. They have also introduced another factor, "Stability," which implies that any number of user requests to the same content served an inappropriate way. The factor "Testability" relates to the ability to undertake to test without affecting the actual running of the WEB site. The authors while explained various quality factors had not proposed any scientific methods using which the quality of those factors measured in quantitative terms.

Different factors, which include Design, organization, and user-friendliness, are considered by Filippo Ricca et al., [7] in addition to the factor "Content" for computing the quality of the websites. Organization of web sites related to the identification of web pages and the way the pages interlinked in a hierarchical structure. The issue of easy navigation is to be kept in view while designing the structure of the web site. The user preference for the content are taken into account while designing the structure of the websites. The "structure" of a website designed in such a way that the content presented as per the preferences of the user. They, however, have not recommended any model used for computing the quality of a website considering the factors that they have considered.

The ease with which a user plays with a website reflects the quality of a website. The playfulness of a web site depends on the way the content presented on the web site. No provision whatsoever provided in any of the quality frameworks provided in the literature. Not even quality computation methods proposed for computing quality of playfulness designed into a website. Saleh Alwahaishi et al., [8] have presented the need for considering the playfulness of the website and the need for computing the quality of the website.

General criteria presented by Layla Hasan and Emad Abuelrub [9] express the way the quality of a website computed irrespective of the kind of content presented on the website. They expressed that all quality factors such as user-friendliness, design, organization, content, etc. included in the general criteria. The web designers can design the web site considering all quality factors, comprehensive indicators, and checklists, which ultimately leads to quality service offered to the end-users.

The growth of the websites increasing rapidly, and therefore there is a need to evaluate the quality of the websites so that the websites with no quality discarded and only web sites of high quality persist. Kavindra Kumar Singh et al., [10] have evaluated the quality of the WEB sites using an objective method called WebQEM (Web Quality Evaluation Method). They have opined that assessing the quality of a website based on human experience and expertise is subjective. They have presented a strategy using which the quality of a web site computed in quantitative terms. The method proposed by them deals with identification characteristics and sub-characteristics and also attributes of those characteristics used for computing the quality of the websites.

Heavy interaction is taking place among users using social and community sites. For effective communication among the members, the quality of these sites must be high. Long-Sheng Chen et al. [11] have attempted to define the factors that determine the quality of theses social sites. They have used a feature selection technique to find the characteristics of the WEB site that influence the quality of the websites.

The usability of a website can be computed using the total number of pages, word count, body text percentage, size of the text in bytes, link count, etc. Naw Lay Wah et al., [12] has presented a Support vector method for finding the good and bad web pages and proceeded to define a method using which the quality of the website in quantitative measure computed.

Several approaches presented that lead to implementing personalized user interface [13], Enhancing the performance of search engines [14][15], and selection of the user interface dynamically for displaying mined results[16] that all lead to improvement of the quality of the web sites.

An architectural approach proposed by Sastry et al.[17] is useful for assessing the quality of a WEB site from a different perspective that includes usability of the WEB sites, quality of content, quality from the perspective of completeness, [18][19][20]

3. COMPARATIVE ANALYSIS OF MODELS USED FOR COMPUTING THE QUALITY OF WEB SITE

Comparative analysis of methods proposed in the literature for computing the quality of the web sites shown in Table 1. From Table 1 It can be shown that none have proposed a method for computing the quality of a web site considering "structure factor."

4. ASSESSING THE QUALITY OF A WEB SITE BASED ON STRUCTURE

A web site is considered as a set of structures with the pages hierarchically linked. A single page may contain multiple hierarchies, as shown in Figure 1. From the Hierarchical structures, the complexity of the structures in terms of its depth, number of structures emanating from the same page, the disconnected of the structures, etc. can be computed.

A structure recognized as in its existence either at Micro Level or Macro Level. An element at a particular position in the hierarchy could be a structure itself. The element, when

expanded, will lead to another structure. A set of structures existing in a web site are enumerated by expanding all the elements existing in the Macro level structure. Each of the structures can be analyzed to assess the disconnected, depth, extent of connectedness, etc. by using a quality assessment model. A mining technique used for mining all the structures existing in the web site so that structures can be further analyzed to find the overall quality of a website.
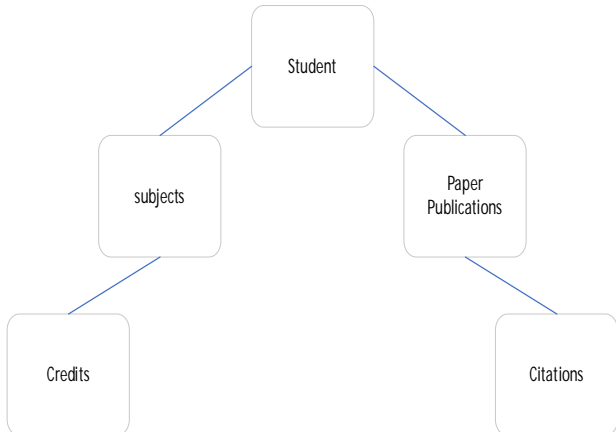

Figure 1 : Hierarchical structure of a typical website

Algorithm for mining the strictures existing in a web site

Step-1

Scan the WEB pages and find the webpage, which is a kind of a root of a structure. Generally, the root of a website is "home.html," etc., as none of the other pages traversed from this page.

Step-2

The resource files that comprise a website contained in the subdirectories of the root directory fetched in the first instance. All the sub-directories are traversed to count the total number of the resources file. The resource files are classified based on the type of extension, such as ".ASP", "PHP," etc. for each type of file, the number of files existing counted, as shown in table 2.

Table 2: Counting number of files with a specific extension

Serial Number	Type of extension	Number of extension files
1.	.php	5
2.	.html	6
3.	.asp	4
4.	.dhtml	3
5.	.jsp	2

Step-3

The details shown in table-1 are stored in a double dimensional array Type [j, k}. The first cell containing the

type, and the second cell contains the number of files of a specific type.

Step-4

For each row in Type [j,k], find the names of the resource files and its related URL and store them in Table   3 containing the serial number of the Type, name of the resource file, and the URL of the resources files.

Table 3: Tracing the URLs of the resource files

Type serial	Name of the resource file	URL of the resources file
2	home.html	Root/html/home.html
2	studentParticluars.html	Root/html/studentPartculalrs.html
2	login.html	Root/html/login.html

The URL in table 3 stored in a vector URL (i)

Step-5

For each of the URL carry structure mining using a standard algorithm and the sequences, fetched are stored in a tree-like structure as shown in Table 4

Table 4: Mined Tree structures

URL	Tree Number	Tree size in number of nodes	Tree structure
Root/html/home.html	1	3	{home, login}, {login, reset}, {home, main}, {home, student Particulars}
Root/html/studentPartculalrs.html	1	1	
Root/html/login.html	1	1	

Step-6

For each tree structure shown in table 4, compute depth, number of edges, size of connectedness and number of disconnected elements, number of elementary nodes and compute the quality of each sub-factor using a cognitive model as shown in quality assessment matrices as shown in Table 5, Table 6, Table 7 and Table 8

Table 5: Quality assessment based on depth (Q_{di})

Depth	Depth > =5	Depth = 4	Depth = 3	Depth <=2
Quality	0.25	0.50	0.75	1.00

Table 6 : Quality assessment based on Number of edges (Q_{ei})

Edges	*Edges >=7*	*Edges = 6*	*Edges = 5*	*Edges <=4*
Quality	0.25	0.50	0.75	1.0

Table 7: Quality assessment based on Connectedness (Q_{ci})

Pages	Connected ness >=8	Connected ness= 7	Connected ness =6	Connected ness <=5
Quality	0.25	0.50	0.75	1.0

Table 8 "Quality assessment based on disconnected Connectedness (Q_{di})

Pages	Dis - Connected ness >=8	Dis- Connected ness= 7	Dis- Connected ness =6	Dis- Connected ness <=5
Quality	0.25	0.50	0.75	1.0

Step-7

Compute average as the quality assessment takes place as one traces from one tree to another**.**

$$Q(do) = (\sum_1^n Qdi)/n \qquad [1]$$
$$Q(eo) = (\sum_1^n Qei)/n \qquad [2]$$
$$Q(co) = (\sum_1^n Qci)/n \qquad [3]$$
$$Q(do) = (\sum_1^n Qdi)/n \qquad [4]$$

Step-8

Compute overall quality of the web site considering the factor "structure" using the equation (5)

$$(Q_s)= Q_d * Q_e * Q_c * Q_d \qquad [5]$$

5. CONCLUSION

The number of websites used is increasing from day to day, especially in the social domains. People around the world are depending on the websites for each and everything. People quite carried away in case of wrong information posted on the websites. People get frustrated if one has to spend a strenuous effort to seek information from the websites. The dependability of the people on websites is increasing day in and after. Every website is different, considering the features and the purpose for which the websites built.

It is a complicated process to determine the factors that influence the quality of the websites. Computing the quality of each factor is important for computing the overall quality of the entire website. Every factor may involve many sub-factors and attributes of those sub-factors. Computing the quality of a factor using either subjective or objective methods will lead to inaccurate results. There are many methods proposed in literature which all attempt to compute the quality of web site as a whole not giving much emphasis to the individual factor. Quantitative methods provide accurate quality measurements.

Many factors exist that influence the quality of the website. The factor "structure" is the key to any website. The quality of the structure has a great influence on the overall quality of the entire website. No attempt has been made by anybody that helps to compute the quality of the "Structure" factor. The method based on structure mining presented in this paper helps to compute the quality of the factor most accurately.

REFERENCES

[1] Miss. Kausar Fiaz Khawaja1, Dr. Rahat Hussain Bokhari2, Exploring the Factors Associated With Quality of Website Department of Technology Management, International Islamic University, Islamabad, Pakistan, Vol.10, Issue 14 November 2010, Pg37-45.

[2] VijaykumarMantri, Dr. S. Kalaimagal, Dr N.Srinivasu, An Introspection of Web Portals Quality Evaluation, Vol.5, No.9, DOI:10.15693/ijaist/2016.v5i9.33-38.

[3] Vassilis S. Moustakis1,2, Charalambos Litos1, Andreas Dalivigas1, and Loukas Tsironis, Website quality assessment criteria, Proceedings of the Ninth International Conference on Information Quality, 2004

[4] Andrina Graniü, Ivica Mitroviü* and Nikola Maranguniü, Usability Evaluation of Web Portals, Proceedings of the ITI 2008 30th Int. Conf. on Information Technology Interfaces, June 23-26, 2008
https://doi.org/10.1109/ITI.2008.4588448

[5] Tanya Singh, Sachin Malik, Darothi Sarkar, E-Commerce Website Quality Assessment based on Usability, Department of Computer Science & Engineering Amity University Uttar Pradesh Noida, India Pg101-105

[6] R.Anusha, A Study on Website Quality Models, Department of Information Systems Management, M.O.P Vaishnav College For Women(Autonomous), Chennai, Volume 4, Issue 12, December 2014, Pg:1-5.

[7] Filippo Ricca and Paolo Tonella, Analysis and Testing of Web Applications, Centro per la Ricerca Scientifica e Tecnologica, I-38050 Povo (Trento), Italy,

[8] Saleh Alwahaishi1 and Václav Snášel2, Assessing the LCC Websites Quality, © Springer-Verlag Berlin Heidelberg 2010, F. Zavoral et al. (Eds.): NDT 2010, S I, CCIS 87,2010, pp. 556–565
https://doi.org/10.1007/978-3-642-14292-5_57

[9] Layla Hasan1 and Emad Abuelrub, Assessing the Quality of Web Sites, Applied Computing and Informatics, Volume 9, Issue 1, January 2011, Pages 11–29

https://doi.org/10.1016/j.aci.2009.03.001

[10] Kavindra Kumar Singh, Dr. Praveen Kumar, Jitendra Mathur, Implementation of a Model for Websites Quality Evaluation – DU Website, International Journal of Innovations & Advancement in Computer Science IJIACS ISSN 2347 – 8616 Volume 3, Issue 1 2014

[11] Long-Sheng Chen1, and Pao-Chung Chang Identifying Crucial Website Quality Factors of Virtual Communities, Proceedings of the International Multi-Conference of Engineers and computer scientists, Vol 1, IMECS 2010

[12] Naw Lay Wah, An Improved Approach for Web Page Quality Assessment, IEEE Student Conference on Research and Development,2011
https://doi.org/10.1109/SCOReD.2011.6148757

[13] Jammalamadaka, S.B. Duvvuri, B.K.K. Jammalamadaka, K.R.S. Priyanka, J.H., Automating WEB interface in relation to user behavior, Advances in Intelligent Systems and Computing, 2019, 815, pp. 91-102
https://doi.org/10.1007/978-981-13-1580-0_9

[14] Sastry, J.K.R. Chittibomma, C.S. Alla, T.M.R, Enhancing the performance of search engines based heap-based data file and hash-based indexing file, International Journal of Engineering and Technology(UAE), 2018, 7, pp. 372-375
https://doi.org/10.14419/ijet.v7i2.7.10722

[15] Sastry, J.K.R. Sri Harsha Vamsi, M. Srinivas, R. Yeshwanth, G., Optimizing performance of search engines based on user behavior, International Journal of Engineering and Technology (UAE), 2018, 7, pp. 359-362
https://doi.org/10.14419/ijet.v7i2.7.10715

[16] Sastry, J.K.R. Jyothsna Sai Sree, M. Mani Dedeepya, T. Kamesh, D.B.K., On selection of a user interface dynamically for displaying data mined results, ARPN Journal of Engineering and Applied Sciences, 2017, 12(11), pp. 3561-3572

[17] Kamesh D.B.K., Sasi Bhanu J., Sastry J.K.R, An architectural approach for assessing the quality of web sites, ARPN Journal of Engineering and Applied Sciences, 2018, 13(15), pp, 4503-451

[18] Sastry, J.K.R. Sreenidhi, N. Sasidhar, K, Quantifying quality of WEB-site based on usability, International Journal of Engineering and Technology(UAE), 2018, 7, pp. 320-322
https://doi.org/10.14419/ijet.v7i2.7.10606

[19] Venkata Raghavarao, Y. Sasidhar, K. Sastry, J.K.R., Chandra Prakash, V. Quantifying quality of WEB sites based on content, International Journal of Engineering and Technology(UAE), 2018, 7(2), pp. 138-141

https://doi.org/10.14419/ijet.v7i2.7.10280

[20] Bhanu, J.S. Kamesh, D.B.K. Sastry, J.K.R., Assessing completeness of a WEB site from Quality Perspective, International Journal of Electrical and Computer Engineering, 2019, 9(6), pp. 5596-5603

Table 1: Comparative analysis of methods used for computing the quality of the websites

Factor for assessment	Sastry	Miss. Kausar Fiaz Khawaja	Vijay Kumar Manthri	Tanya Singh	Anus ha	Filippo Ricca,	Saleh Alwahaishi	Kavindra Kumar Singh	Long-Sheng Chen	Naw Lay Wah
Overall Identification		YES		YES	YES	YES	YES			
Overall Framework	Yes									
Overall computation								Quantitat ive	Feature selection	Support Vector Machin es
Completeness	Quantitative									
Content	Quantitative									
Navigation										
Structure	Quantitative									
Multimedia										
Uniqueness										
Usability	Quantitative assessment	Identification	Quantit ative							
Adequacy		identification								
Privacy										
Security										
User-friendliness										
Look and feel										