International Journal of Engineering & Technology, 7 (2.7) (2018) 138-141

SPC

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper



Quantifying quality of WEB sites based on content

Y Venkata Raghavarao [1]*, K. Sasidhar [2], JKR Sastry [3], V Chandra Prakash[4]

[1] *Department of computer Science and Engineering, KLEF University, Vaddeswaram, Andhra Pradesh, India*
[2] *Department of computer Science and Engineering, KLEF University, Vaddeswaram, Andhra Pradesh, India*
[3] *Department of computer Science and Engineering, KLEF University, Vaddeswaram, Andhra Pradesh, India*
[4]*Department of computer Science and Engineering, KLEF University, Vaddeswaram, Andhra Pradesh, India*
**Corresponding author E-mail: venkataraghav@kluniversity.in*

Abstract

Information dissemination is taking place extensively through WEB and use of internet. The quality and reliability of the information hosted on the WEB is questionable. Many factors are to be considered for assessing the quality of the WEB sites. The Information hosted on the WEB would become valuable only when top quality of the same is maintained. Each quality factor can have many dimensions. There should be a mechanism of computing the quality of a web site quantitatively so that quality of a web site can be realistically measured. Any subjective or objective way of measuring quality is questionable and can be prejudiced at times. Every factor as such must be measured the entire quality of a web site must be measured considering all the factors. One can develop a norm for quality of a factor and any deviation from the norm needs to be rectified and controlled.
Among other factors quality of the content hosted on the WEB plays vital role. In this paper the computational method which can be used for computing the quality of a web site is presented considering all the dimensions of content related quality factor.

Keywords: Quality, Assessment; Computational Methods; Framework; WEB content; WEB sites

1. Introduction

The information exchange is predominantly taking place due to advent of Internet and WEB technologies. Most of the content these days is hosted on the WEB making it easy for the people have the access to the information in least time and negligible cost. The information hosted on the WEB is of different forms which include text, images, audio and videos. Information is also being available in streamed mode. The Web sites are being used for doing e-commerce including marketing. The way the people live is changing drastically with the content being available on the web changing. How the quality of the content hosted on the WEB needs to be assessed before taking the content hosted on web for granted. The attitude of the people requiring the information is changing heavily. The customers won't be satisfied if the WEB site is not of quality. The WEB site loses its existence and never be surfed if the quality of information hosted on the WEB is erroneous and some time has no relevance or connectedness. The quality of a website can be assessed using some quality factors such as Usability, content Reliability, Flexibility, Functionality, Portability, Maintainability, Privacy, Security, Adequacy of Information, Safety, Content, Navigation and etc. The quality of website can be assessed in many ways. The assessment can also be done using evaluation tools. Many organizations are developing their information in terms of Videos, audios, graphics, text, tables, forms, animations, Logos, dynamic content, mouse over effects, graphics etc. While some web sites are static in nature some other sites are dynamic.

Many methods existing in literature for computing the quality of the WEB site are either subjective or objective. There us prejudice involved in the subjective assessment and there is lack of completeness in the objective assessment of quality of WEB sites. Some methods are biased due to individual preferences or some purely based statistical measurements relating down time, response time etc. The quality of a web site is based on the persons who look at the WEB sites. Programmers look at the WEB site characteristics that include security, functionality and maintainability whereas the users using the WEB site look at the quality of the WEB site in terms of usability, credibility, efficiency and credibility.

Every web site has certain number of quality characteristics. The characteristics must be evaluated to assess the quality of the WEB site and also to find the extent to which the user needs are being met. Assessing the quality of the web site is quite complicated especially when the web sites are complex. The WEB sites that are related to e-commences, museums, sites that deal with animations are actually quite complicated. Even the process that is used for evaluating the quality of a web site is more complex. Most of the quality characteristics of web sites is also are interleaved and inter connected through complex logic that relates the Quality characteristics of the web site. Some of the quality factors can only be computed subjectively and some other parameters are to be evaluated objectively and some other parameters are computed quantitatively. A composite model comprising all aspects of computation of quality is required. Connectivity between subjective, objective and quantitative is required for the development of proper quality assessment model.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The expectations of a site must first be determined. Quality parameters that are representative of the expectations must be determined and must form into a base or a quality assessment model using which the overall quality of the web site can be assessed. The metrics that measures the quality characteristic must be determined and the quantitative technique that can be used for measuring the quality characteristic must be determined.

The quality assessment methods vary from factor to factor and the number of features that must be considered shall also vary from factor to factor. Each feature and the combined effect of all the features must be measured through most suitable computational methods. In this paper the features and computational methods related to the quality factor "Content" is presented.

2. Literature Survey

Miss. Kausar Fiaz Khawaja1 et al., [1] have dealt with factors that include Usability, Privacy Security, and Adequacy of information and Appearance which can be used for evaluating the quality of a WEB site. Usability is the ease of use and learn ability of a human-made object such as a tool or device. It describes the quality of user experience across websites. Appearance means the visibility of the WEB site which includes appealing, polished and professional presentation. Adequacy of Information means Putting sufficient and useful information in the website. A method has been presented paper that helps computing the quality based on the observations made while the WEB site is in use.

Vijay kumar Mantri et al. [2], have presented that quality of a WEB site can be computed using the factors such as Usability, Safety and Flexibility. Usability of a website must be effective, efficient and satisfactory. Safety of a Website must ensure that no interaction of a user with the WEB site could be ever revealed. Flexibility is the ability to add / modify / remove functionality affecting the WEB pages without damaging the functioning of current ongoing system. The authors have used a tool called Portal Data Quality Assessment Tool (PoDQA) using a quality model called Portal Data Quality Model (SPDQM).

Vassilis S. Moustakis et al., [3] have used the quality factors which include Content, Navigation, Structure and Design, Appearance, Multimedia, Uniqueness. Content is the information conveyed to the end user through a user interface. Content reflects quality, completeness, degree of specialization or generalization and reliability of information included in the website. Navigation reflects the support provided to the user when moving in and around the site. Elements of navigation include easiness of moving around, easiness in understanding site structure, and availability and validity of links. Structure and Design incorporates aspects that affect order of presentation, speed and browser. Appearance and Multimedia captures aspects that relate to site's "look and feel" with special emphasis on the state of the art graphics and multimedia artefacts. Uniqueness refers to user's perception that the site carries something that makes it different in a world full of sites. A computational method called AHP (Analytical Hierarchical Process) has been used for computing quality of a web site.

Andrina Graniü et al., [4] have presented the quality of a WEB site from the point of portability. Portability means the ability to move the website from one host platform to another and the platform that runs the site will work on the new host.

Tanya Singh et al. [5], have used quality factors that include Usability, Privacy and Security, Adequacy of information and Appearance. Usability is the ease of use and learn ability of a human-made object such as a tool or device. It describes the quality of user experience across websites. Privacy is all about revealing information to those users who are identified by the owner of the user. Only selected users are provided with the information of those data elements intended to be shared by the owner of the data. Security is all about preserving the interaction of the user with the WEB site. Adequacy of information is related to making available complete data without any loose ends that suffice the actual requirements of the end user. Appearance of a WEB site is all about displaying the content in most understandable by using colours. Graphics, sequences etc.

Anusha et. al., [6] have considered Factors such as Portability, Reliability, Functionality, Usability, Maintainability, and Efficiency to asses quality of a WEB site. Portability means the ability to move the website from one host to another and the platform that runs the site will work on the new host. Reliability means that a WEB site shall reflect the same information any number of times it is sought in the same context. It is the probability that the intended page will be available and presented to the user. Website must be Free from errors. Functionality of a website includes accuracy, security, suitability and etc. Usability is the ease of use and learn ability of a human-made object such as a tool or device Maintainability includes analysability, changeability, stability, Testability. Maintainability implies the simplicity with which changes can be made to the WEB site while the WEB site is up and being used for other. Analysability includes the readability of the content and the ability to intemperate the same including the tracing of navigational paths. Stability is a feature that dictates that the same content is displayed any number of times a user visits a WEB page in the same context. Testability of a WEB site includes all those features using which the proper working of the WEB site is tested while the WEB site is up and running.

Filippo Ricca et al., [7] have considered Content, Design, Organization and user friendliness as the quality factors that must be considered in evaluating the quality of a WEB site. Organization of a Web site includes the identification of WEB pages and the way the WEB pages are linked hierarchically. The linking of the WEB pages is done in such way that is easy to navigate. The WEB pages must be simple and user friendly in the sense, the content shall be presented to the user as per preferences of the user.

Saleh Alwahaishi et. al., [8] have considered playfulness, and Level of representation of the content as the most important factors that should be considered for evaluating the Quality of a WEB site. Most of the presentation on the basement of quality framework has neither provided a framework or appropriate computational methods using which quality of a WEB site can be computed.

Layla Hasan and Emad Abuelrub [9] have proposed a general criterion for evaluating the quality of any website regardless of the type of service that it offers. They have contended that the dimensions of quality criteria that include content, design, organization and user-friendliness. These dimensions together with their comprehensive indicators and checklist can be used by web designers and developers to create quality websites to improve the electronic service and then the image of any organization on the Internet.

Kavindra Kumar Singh et al., [10] have expressed that the rapid growth of web applications increases the need to evaluate web applications quantitatively. WebQEM (Web Quality Evaluation Method) have been used for objectively trying to evaluate the web applications. Weighing a web attribute has been proved to be subjective and mostly dependent on judgements made by the experts. The authors have presented quantitative evaluation strategy to access the quality of WEB sites and Applications. The methodology proposed by them is useful to systematically assess characteristics, sub-characteristics and attributes that influence product quality. They have presented models, methods, procedures, principles and for assessing quality of a WEB site.

Long-Sheng Chen et al., [11] have presented that heavy interaction is taking place among the members especially through social WEB sites. It has become very important to consider the quality of WEB sites. The authors have attempted to define the quality factors of virtual communities and then proceeded to identify key factors by attracting new members by using feature selection technique

Naw Lay Wah et al., [12] have presented that the WEB sites must be evaluated and measured for quality. He has presented several

metrics that are quite related to usability associated with good design elements such as word count, total pages, size in bytes, body text percentage, average link text count and others. He has presented the computation of WEB site quality on basis of 16 factors. He has used support vectors to predict good and bad web pages. A quantitative analysis of WEB page attributes has been presented.

3. Investigations and Findings

Content is hosted on the WEB in terms of text, data, images, videos, 2D / 3D graphics, graphs etc. The quality of the content in basic form can be computed in terms of hierarchical representation of the content and the connectedness of the content elements. Quality of Images, Videos, 2D / 3D graphic shall come under the category of the quality factor called "Multimedia"

When content is displayed, there should be connectedness between various elements displayed to the end user. The text must be connected, the data displayed must be connected and text displayed must read with the data displayed properly.

The computation of depth of connectedness can be understood to be having three components which include connectedness between the text, connectedness between data and the connecters between the text and data.

Computation of connectedness is more complex as the text and data is spread across several pages and developing relations ships among several text and data elements that are spread across several pages is complex. While it is possible to develop connectedness within a given single page, the overall connectedness can be computed as the summation of the connectedness of all the pages.

CP = Connectedness within a page

CW = Connectedness within a site = /n

CP = Connectedness of text elements (CPtext) + Connectedness of data elements (CPdata) + Connectedness of text and data elements (CPtdata)

CPtext =

CPdata =

CPtdata =

Given a set of html pages one can parse the files and get count of text and data token available in the recourse files related to the Web site. To find the connected between the Text, Data and Text and Data three repositories are to be maintained by using an interactive Tool. The repositories to be built are shown in Table 1, Table 2 and table 3.

Table 1: Text relationships

Serial number	From text word / String	Two Text word / String
1.	Registration Number	Student Name
2.	Registration Number	Student Age
3.	Registration Number	Student Gender
4.	Registration Number	Department name
5.	Registration Number	Program name
6.	Registration Number	Program Year

Table 2: Data relation ships

Serial number	From text word / String	Two Text word / String
1.	Regst	SName
2.	Regst	SAge
3.	Regst	PNumber
4.	Regst	SGender
5.	Regst	Dname
6.	Regst	Pname
7.	Regst	PYear

Table 3 Text and Data relation ships

Serial number	Data	Text
1.	Regst	Registration Number
2.	SName	Student Name
3.	SAge	Student Age
4.	PNumber	Phone Number
5.	SGender	Student Gender
6.	Dname	Department name
7.	Pname	Program name
8.	PYear	Program Year

By scanning through the HTML files all the text, data, Text-Data pairs listed in the repositories are compared with the parsed content existing in the files and the number of text, data, text-data elements that have been matched are counted. Using these counts the connectedness parameters can be computed.

An existing web site has been processed and the connectedness counts have been computed which are shown in the Table 4.

Table 4: WEB connectedness computations

Serial Number	Computing parameter	Count
1.	Text count	23003
2.	Number of connected text counts	21001
3.	% of Connected text Counts	0.91
4.	Data count	2409
5.	Number of connected data counts	2109
6.	% of Connected data Counts	0.87
7.	Text-data count	25412
8.	Number of connected text-data counts	18099
9.	% of Connected text-data Counts	0.71

CP = (0.91 + 0.87 + 0.71)/3= 0.83

The quality of connectedness can be computed over a scale of 1 to 4 as shown in the table 5. Connectedness value being 0.83, the processed web site can be categorised as High

Table 5: Computation of Quality of connectedness

Quality Factor	Poor	Average	Good	High
Quality Weight	0.00	0.01 - 0.25	0.26 - 0.45	0.46 - 1.00

Hierarchical depth used for presentation of the content is one the most important issue as the user who reads the content hierarchically find it difficult to collaborate. The more the depth of the content the more complex the content is for understanding.

Based on the repositories shown in Tables 1-3 connectedness of the WEB site can be generated and also can be computed using the resource files of the WEB sites. Table 6 shows some of the possible Hierarchies

Table 6: Connected Hierarchies

Hierarchy Number	Text/data	Text/data	Text/data	Text/data	Text/data
1	RegstNo	Student Name	Sname	Student Age	SAge
	Phone Number	pnumber	gender	Sgender	Department name
	dname	Program name	Pname	Program Year	Pyear

Number of possible hierarchies can be known from the repositories. The hierarchies that have been encapsulated into the web sites can be counted based on the match of the text, data and text-data and developing the connected among the matched content. The quality of the content can be assessed based on the % of generated hierarchies in comparison with possible hierarchies.

CHE = Connected hierarchies exiting in the WEB site = 89

CHP = Possible number of hierarchies = 72

CH% = Percentage of existing hierarchies = CHE / CHP = 72/89=0.80

The quality of the WEB site can be computed by comparing the actual with expected and grading on a 4 point scale as shown in the table 7.

Table 7: Computation of Quality of connectedness

Quality Factor	Poor	Average	Good	High
Quality Weight	0.00	0.01 - 0.25	0.26 - 0.45	0.46 - 1.00

From table 7it can be seen that the quality of the WEB site can graded as HIGH as the hierarchy count is computed s 0.80

TQC = Total Quality of the quality Factor "Content"

TQC = CP + CH% = (0.83 + 0.80) / 2 = 0.815

4. Conclusion

Every web site must be meticulously designed and developed for the use of information hosted on the WEB site. Most of the world is dependent on the quality of information hosted on the WEB site. The dependability on a web site for information can be expressed in terms of the quality of the WEB site. Many factors are to be considered for computing the quality of a WEB site. Content is one such factors. Every factor must be expressed using the computational expressions so that the factor is expressed in quantitative terms. Every factor may involve in subsidiary factors existing in hierarchical fashion. The computation of the quality must be expressed considering computations of the entire subsidiary factors. The computation of the quality of content hosted on the web can be expressed in terms of connectedness and hierarchical presentation of the same.

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 Tsironis1 , 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.

[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.

[9] Layla Hasan1 and Emad Abuelrub, Assessing the Quality of Web Sites,"www.sciencedirect.com/science/journal/22108327", January 2011, Pages 11–29.

[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.