

Original Article

Web Accessibility Quality Index Varies with the Selection of Evaluation Tools

Abhay R. Palaskar¹, Raju R. Gondkar², Surekha R. Gondkar³

¹Scrum Master, BNFS, TCS Amsterdam, Netherlands

²Department of Postgraduate Studies, CMR Institute of Technology, Bangalore, India

³BMS Institute of Technology and Management, Bangalore, India

Received Date: 23 September 2021

Revised Date: 25 October 2021

Accepted Date: 07 November 2021

Abstract — This article on web accessibility illustrates a study on the accessibility quality index of government services websites related to 10 countries after its implementation of web-accessible guidelines. This assessment is performed to analyze the website's compliance with the latest WCAG 2.0 guidelines and ARIA rules published by the World Wide Web Consortium (W3C). The intent of this research is to determine whether the implemented guidelines are reflecting and usable as per norms or not. Research also includes quality index differences that are stats how different tools are analyzing the same website and differ on quality issues. The WCAG 2.0 guidelines support web accessibility implementation, and we are referring to the Website Accessibility Conformance Evaluation Methodology (WCAG-EM) to refer to how easily disabled people can access websites. It is observed that the majority of the websites do not achieve an acceptable level of compliance. From this research, we conclude that most of the evaluation tools are different in providing exact accessibility issues present on a specific website, and this caters to final quality index differentiation. This research was specifically performed on government services-related websites because the majority of disabled people are using government services for day-to-day operations such as banking, education, policies, registrations, taxes. With multiple manual testing iterations, it was found that several non-compliance errors are present on these websites. Changing the defect count of every tool on the same website on two different occasions results in more work planning for developers; with this observation, there is a significant amount of work that needs to be done in this area where tools need to be more robust w.r.t accessibility issues.

Keywords — Disability, Evaluation, Quality, WCAG 2.0, ARIA, Accessibility, Web accessibility.

I. INTRODUCTION

Web accessibility is getting more specific to addressing the disabilities with each newer version and slowly becoming a necessity for all industries with online functions. Nowadays, when most of the IT companies are moving towards the agile way of development as compared

to the traditional waterfall development methodology, developers on the agile team prefer more robust and automated tools and processes to test the web accessibility issues that were implemented as per WCAG 2.0 guidelines. Even if these tools are highly intelligent, I making issues uncover, there is a significant amount of limitations w.r.t to its functionality [1]. It has been observed that many of the low priority issues have been given more priority and importance as compared to the highly critical features, so the validations need to be handled in a sensible manner while exposing the irregularities.

I have been working on software testing, both manual and automated related to accessibility, for the last 12 years for multinational clients and their e-Commerce services. Testing with different types of frameworks and working with people is always on priority. The important point here to make is no validation tool can test all accessibility success criteria, which makes this entire implementation more complex. The Software Quality Index is an attempt to relate software development practices to the quality of the final project. As additional forms are returned, the results will be updated. While working on software testing projects, the quality index matters a lot in order to decide whether the final product is a success or not, and there could be numerous elements attached to that such as revenue, defect count for production, client-vendor understanding of success criteria and future work.

The software quality index plays a vital role while releasing the feature to production or to end customers, and this applies to any of the business or institution, or services. Here we have defined the quality index of accessibility features based on certain important criteria mentioned as part of WCAG 2.0 guidelines for e.g., Must have criteria for the specific type of disabilities. There are specifically 6 types of disabilities categorized as part of web accessibility guidelines, and those are 1. Auditory 2. Cognitive 3. Neurological, 4. Physical 5. Speech 6. Visual. All the web accessibility guidelines are developed to address issues faced by the above categorized disabled users while accessing the web. The poor-quality development or missing guidelines from its implementation will have a



significant impact on how these people will use the web. [2] There is a study that suggests that the majority of disabled people are not able to use government aids or policies due to the way they must use the web before the implementation of accessibility. Our research is not to focus on guidelines of uncovering accessibility issues on the specific website but to measure application quality index after the implementation of accessibility standards on specific websites.[3]

There is a study that covers the performance, quality, and accessibility of government websites [8] which suggests that there is a significant increase in the number of government services related websites, but their success is largely dependent upon how the accessibility features are implemented and how functional it is to use. Most of the websites are failed with perceivable, operable, understandable, and robust principles, which are listed under the WCAG 2.0, and none of the websites reaches AA conformance with the guidelines.[4]

Quality of any software application is the most important factor in its implementation as it decides on how usable the application will be and if there is any need for modification or change of the way how it was developed, need of another development methodology of do we need to change how we test the application, the number of questions can be raised over before we call it a success on the quality index. Here we are testing government services-related websites where the majority of the disabled users are using the website for critical issues or policies. Implementation of successful accessible guidelines should be the first priority for development.[5]

II. SOFTWARE QUALITY INDEX

A. What is the mean by Application Quality Index?

The term "Quality" is often used in a very vague way. It really meant different things to different people, even in the same context. Some technical people with an IT background may say the application quality is having no defects in the application, whereas the other set of people may define it as something delivered on time, within budget, meet a perfect set of requirements, and can be maintained. Such a difference of opinion on the definition.

Let's see what the standard definition of the quality as per the ISO/IEC 8402 standard is-"The totality of features and characteristics of a product or a service that bear on its ability to satisfy stated or implied needs". Basically, it is impossible to measure the quality of the product that is functioning well today and that suddenly may go down tomorrow, and this breakdown may not be listed under the circumstances where the quality was given 100%, that is why we measure Quality Index instead of quality itself. The Quality Index (QI) is a measure of quality.[6]

- Customer satisfaction is the most important thing when we measure the quality of the product, so based on the Quality Index (QI), one can understand how satisfied customers are with this product.

- It should be easy for the management to go with one number and then drill down from that number to the root cause.
- The QI numbers with upward or downward trends provide continuous feedback, which is absolutely required for quality control. It is also easy to monitor the progress.

Here is the list of characteristics mentioned in the quality index; we will be using 2 out of 9 quality characteristics for our study of tools that evaluate the accessibility of websites.

- Completeness of Accessibility features – It comes with certain sub-characteristics such as suitability, accuracy, data integrity, and interoperability. The definition of completeness speaks about having all the required information.
- Conciseness – Having no excel of information.
- Consistency – The entire set of the product should follow the consistent approach, uniform notation, symbols, and terminology itself.
- Portability – The application or the product should be able to adapt to a different set of environments. It comes with two subcategories as adaptability and Installs ability.
- Maintainability - It deals with the ability of the product to extend with new or changing requirements.

It comes with three subcategories as Analyzability, Changeability, stability.

- Testability – This refers to the ability of the application to get tested on the requirements against the acceptance criteria and should be able to do performance evaluation.
- Usability – It deals with the efforts that require to use the product or application itself. Usability has three different subcategories as Understandability, Learnability, and Operability.
- Reliability – Application or product's ability to perform satisfactorily under different conditions. It has two subcategories as Fault tolerance and Recoverability.
- Efficiency – Efficiency is the ability of an application or product to perform its function with required latency, i.e., response time and with resources. It has two subcategories Latency time which is the response or processing time, and Resources used.

Security – Last and very important factor in measuring quality is how secure the application or product is, the security is the ability to protect data against unauthorized access and to block any malicious interference with the application or product operations. [7]

III. EVALUATION OF QUALITY INDEX

Evaluation of the Quality Index is a complex task relative to web accessibility implementation on related web applications. There are two basic modes of evaluating the Quality Index.

- Metric based approach
- Queries based approach

We are using Metric based approach for this research on how the quality index gets differs on selective tools, and that leads to major future decisions on the project. Each characteristic and sub-characteristic are converted into meaningful questions with parameters and ratings based on issues and defects reported by tools for a specific set of pages scanned for accessibility issues. We are selecting two characteristics based out of 9 to calculate the Quality index of 10 government services-related websites on which we can see the accessibility features are implemented. [8]

- a. less than 100 issues= 5 point
- b. More than 100 and less than 200 issues = 4 points
- c. More than 200 issues and less than 300 = 3 points
- d. More than 200 issues and less than 300 = 2 points
- e. More than 200 issues and less than 300 = 1 point

Once we finish the rating, then we need to calculate the score of each character based on the sub-characteristics score to define the Quality Index.

$$C(j) = \frac{\sum_{i=N(j)} S(i)}{N(j)} \quad (1)$$

Where,

C(j) = Rating received for quality characterizes

S(i) = Rating received for Sub-characteristics

N(j) = Number of sub-characteristics

Once we get the value for C(j), then we need to calculate the Quality index (QI) of our web application for the specific tool.

$$QI = \frac{\sum_{j=n} C(j)}{n} * 100 \quad (2)$$

Were,

QI = Quality Index

N = Number of characteristics

Based on the final value of the Quality Index, we can assign that specific value to the tool, and the web application will go under test with a different tool to prove the difference in a quality index.

Notable Assumptions,

- a. QI is a number, and it will be more meaningful after repeated tests with time to define the trend.
- b. This number cannot be final and can be changed based on the industry, projects, development tools,

web applications, time of testing (due to downtime of a web application)

- c. QI cannot be the same as other similar applications or teams working on the same project; it differs.

IV. SELECTION OF TOOLS TO EVALUATE

A. Web accessibility guidelines

We have a stable WCAG 2.0 version in place to refer to, and the latest WCAG 2.1 version is just released. For the study, we have used WCAG 2.0 version with all the mentioned guidelines. It has 12 guidelines that are part of 4 principles 1. Perceivable 2. Operable 3. Understandable 4. Robust.

The perceivable speaks about the information present on the website w.r.t user interface of a web application as the user can see and feel it while using it. The Operable, whereas mostly associated with operating UI parts of an application or website. The third, i.e., understandable, is associated with how easy and clear the information is present for some to understand it; it deals with Operable and Perceivable as both. The final principle, Robust, is associated with content and platform. It should be adequately deciphered irrespective of the platform it is being used on. For all the 12 mentioned guidelines, WCAG 2.0 has rules available which will put the application under test into one of the 3 certification buckets or levels: A, AA, and AAA. The levels are here to determine the conformance of web applications with guidelines.

B. Existing tools for Web Accessibility evaluation

There are now more than a hundred tools available to test web accessibility, but none of them confirms to test all the rules and guidelines at once. This study is based on the difference in Quality index with each tool on the same application. These tools break down the HTML codes of web applications or websites to automatically test the rules by using different techniques. By looking at issues, developers can understand the severity of the defect on the number of parameters to decide the priority to fix it for the client or stakeholder. Looking at the guidelines, it is not an easy job to fix the issue at just one or to entirely develop a new feature based on web accessibility standards. It takes a great amount of study and a deep understanding of accessibility rules and guidelines.

a) Wave

Wave is the most popular tool in the market to test web accessibility elements which were created by Web Aim. It gives a detailed visual description of the web application with an exact page that is being evaluated with yellow notices highlighting the mistakes. On the other hand, Warnings on the right side highlights the issues listed as per guidelines. With the latest version on the wave tool, now we can see accessibility errors, contract errors which are crucial for vision-related disabilities, alerts and features, structural elements, and ARIA-related issues. Wave is a great graphical representation of issues with accessibility. It can also show more details on exact issues, references of where the issues are exactly located, structural representation of HTML elements and issues, in detail contrast related issues making the vision-related issues

more exposed to get it fixed. We will be using wave to get our Quality index for government services-related websites. We have selected the Wave tool for the study because of its robust nature and simple issue reporting techniques.[9]

b) AChecker

AChecker is used to evaluate HTML content for accessibility problems by entering the location of a web page, uploading an html file, or by pasting the complete HTML source code from a Web page. AChecker produces a report of all accessibility problems for your selected guidelines. A Checker identifies 3 types of problems: Known problems: These are problems that have been identified with certainty as accessibility barriers. You must modify your page to fix these problems; Likely problems: These are problems that have been identified as probable barriers but require a human to make a decision. You will likely need to modify your page to fix these problems; Potential problems: These are problems that AChecker cannot identify, that require a human decision. You may have to modify your page for these problems, but in many cases, you will just need to confirm that the problem described is not present.[10]

c) Powermapper or SortSite

SortSite is a one-click website testing tool used by federal agencies, Fortune 100 corporations, and independent consultancies. The tool is available as a desktop application for Mac or Windows and is also available as a web application. Platforms: Macintosh and Windows. It can test inside and outside of a firewall, including intranets and development sites. One click is all it takes to analyze an entire website. Each page is checked against more than 1300 standards-based checkpoints.[11]

Accessibility - check WCAG and Section 508 guidelines against many file types: find flashing GIFs, untagged PDFs

Broken Links - check for broken links and spelling errors, Compatibility - check for HTML, script, and image formats that don't work in common browsers

Search Engine Optimization - check Google and Bing webmaster guidelines, Web Standards - validate HTML and CSS

Usability - check against Usability.gov guidelines

So basically, our study will be using the above 3 best available tools from the entire pool of automated tools available both for free and commercial use to create our Quality Index (QI) score for selected websites. This study is not related to the effectiveness of the tools we are using but getting the score only.

V. METHOD AND MATERIAL (DATA)

Web Accessibility is an inclusive practice of removing the issues that come in between using the Web in a normal way by disabled people. Irrespective of their disability, they

should be able to use the internet as any other common person. These guidelines are very important for developers to create accessible websites which are robust, operable, perusable, and understandable by everyone. This collection of data and website evaluation has two parts to it. Selection of websites based on their accessibility implementation flag, manually we are checking if any or few of the accessibility components are implemented or not, if we see that website is started implementing WCAG 2.0 guidelines then it does qualify for the study, our intention is not to validate the website based on its accessibility implementation status but to derive Quality index (QI) with a different set of tools. We will be using the Quality index formula to get different sets of values for selected websites.[12]

For the analysis here, there are two tables listed with :

1. The first table shows the selected websites for the study and evaluation with 3 accessibility check tools
2. The second table shows the Quality index (QI) calculated through a series of steps.

Table 1. Government websites

Country	Selected Websites for Evaluation	
	Website URL	Purpose
Ireland	http://www.gov.ie/	Ireland
Japan	http://www.japan.go.jp/	Government of Japan.
Germany	https://www.bundesregierung.de/	Government of Germany.
USA	https://www.usa.gov/	Government of the USA
Finland	http://valtioneuvosto.fi/	Government of Finland.
Netherlands	https://www.government.nl/	Government of the Netherlands.
Sweden	http://www.government.se/	Government of Sweden.
Israel	http://www.president.gov.il/	Government of Israel.
UK	https://www.gov.uk/	Government of the UK
Denmark	http://denmark.dk/es	Government of Denmark.

As a next step, we have evaluated each of the above websites through three different web accessibility tools, 1. Wave 2. AChecker and 3. Powermapper for the list of accessibility issues related to WCAG 2.0 guidelines. All

the above tools support WCAG 2.0 guidelines to check existing accessibility issues. The Metric based approach is a specific set of characteristics and sub-characteristics we have selected for the study to get the Quality index(QI) score.

For example, of the detailed calculation here, we can select one website as <http://www.gov.uk>, which is the government of UK website with accessibility standards implemented. Now when we have tested this website on the 3 different tools here, we get the score from the characteristics and its sub-characteristics.[13]

Here for the first Characteristic, we have checked the Completeness of accessibility features.

Here is the score as per out point system from three different tools.

- a. Wave – 132 issues – Gets 4 points
- b. AChecker–238 issues – Gets 3 points
- c. Powermapper – 322 issue – Gets 2 points

Here for the study, we are considering 1 point if the subcategory is not available; for the Completeness category, we do not have any subcategory, so the 1 point for each.

Here is the final score calculated based upon the Quality Index (QI) formula.[14]

$$\begin{aligned} \text{QI through Wave} &= 4/1 * 100 = 400 \\ \text{QI through AChecker} &= 300 \\ \text{QI through Powermapper} &= 200 \end{aligned}$$

With further studies on this, we will be adding more characteristics and sub characteristics to go deeper in calculating the Quality Index. The below table shows the difference in how Quality Index differs based on how and what tool we select for evaluation.

Table 2. Quality Index Calculation

Country	Selected Websites for Evaluation		
	Wave QI	AChecker QI	Powermapper QI
Ireland	400	300	200
Japan	400	200	300
Germany	100	300	200
USA	200	100	300
Finland	300	400	100
Netherlands	200	200	300
Sweden	400	300	100
Israel	100	200	300
UK	200	300	100
Denmark	300	100	200

The above table shows the difference of Quality Index (QI) based on selection of tools, the Quality Index (QI) is very important factor while making decision to load the feature to production or to launch it for the end customer, this key value needs to be always on top. Here in this study, we

focused mainly on websites with government-related information where more people need accessibility features.[15]

VI. CONCLUSION

We have systematically followed the methods and calculations to the Quality Index (QI) as the final value for the 10 websites with 3 different sets of tools, and the analysis shows that all the tools differ in their own implementation and provide different results for the same website. We see a need for more robust tools here. We understand that the method of tools development differs and the way of interpreting the WCAG and ARIA rules, but when it comes to the reliability of any website to understand whether that specific website is compliant or not, then we see the difference here. In the next part of the study, we are adding more categories and subcategories to the Quality Index (QI) calculation and the trend looks the same.

REFERENCES

- [1] International Telecommunications Union(ITU), Measuring the Information Society 2014: Report 2014, Retrieved from <http://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis2014.aspx>,(2014).
- [2] R.J. Freeman and P. Loo, Web 2.0 and E-Government at the Municipal Level, Proceedings of the 2009 World Congress on Privacy, Security, Trust and the Management of e-Business, (2009) 70-78.
- [3] S. Lujan Mora, Web Accessibility Among the Countries of the European Union: A Comparative Study, Actual Problems of ComputerScience,1(3) (2013) 18-27.
- [4] C.Boldyreff, Determination and evaluation of Web accessibility, Enabling Technologies: Infrastructure for Collaborative Enterprises, Eleventh IEEE International Workshops, (2002) 35-40.
- [5] D.Ortner, L.U.A.Inst.Integriert Studieren and K.Miesenberger, Improving Web accessibility by providing higher education facilities for Web designers and Web developers following the design for all approach, 16thInternational Workshop on Database and Expert Systems Applications, (2005) 866-870.
- [6] W3C,W3CAG2.0Guidelines,12112008.[Online].Available:<https://www.w3.org/TR/WCAG20/>. [Lastaccess:02022016].
- [7] http://www.projectperfect.com.au/info_quality_index.php
- [8] W. Yaokumah, S. Brown and R. Amponsah, Accessibility, Quality and Performance of Government Portals and Ministry Websites: A View Using Diagnostic Tools, 2015 Annual Global Online Conference on Information and Computer Technology (GOICT), Louisville, KY, (2015) 46-50.
- [9] Y.Akgül,Quality evaluation of E-government websites of Turkey, 2016 11th Iberian Conference on Information Systems and Technologies (CISTI), Gran Canaria., (2016) 1-7.
- [10] M.Scholl, How Usable and Clear are the Websites of European Capitals from the Point of View of German Students Adopting an E-Government Perspective? , 2016 49th Hawaii International Conference on System Sciences (HICSS),Koloa,HI, (2016) 2893-2902.
- [11] Techniques for WCAG 2.0, <http://www.w3.org/TR/WCAG20-TECHS/> (retrieved July 12, 2016)
- [12] Classifications of Accessibility Evaluation Tools, <http://webaim.org/articles/tools/> (retrieved June 22, 2016).
- [13] Website Accessibility Conformance Evaluation Methodology (WCAG-EM) 1.0, <https://www.w3.org/TR/2014/NOTE-WCAG-EM-20140710/> (retrieved July 2, 2016)
- [14] P. Giulianelli, C. Pons and C. Gonzalez, Mobile government web sites analysis, 7th Colombian, Computing Congress (CCC), Colombia, 202 (2012)11-12.
- [15] S. Luján-Mora, R. Navarete, and M. Peñaflor, eGovernment and Web Accessibility in South America, First International Conference, Quito, (2014) 77-82.