

NISTIR 8010

Challenges and Benefits of a Methodology for Scoring Web Content Accessibility Guidelines (WCAG) 2.0 Conformance

Frederick Boland
Elizabeth Fong

<http://dx.doi.org/10.6028/NIST.IR.8010>

NISTIR 8010

Challenges and Benefits of a Methodology for Scoring Web Content Accessibility Guidelines (WCAG) 2.0 Conformance

Frederick Boland
Elizabeth Fong
*Software and Systems Division
Information Technology Laboratory*

This publication is available free of charge from:
<http://dx.doi.org/10.6028/NIST.IR.8010>

September 2014



U.S. Department of Commerce
Penny Pritzker, Secretary

National Institute of Standards and Technology
Patrick D. Gallagher, Under Secretary of Commerce for Standards and Technology and Director

Abstract:

The World Wide Web (W3C) Web Accessibility Initiative (WAI) has developed comprehensive guidance on promoting web accessibility. As part of the WAI, the Web Content Accessibility Guidelines (WCAG) 2.0 specifies success criteria for evaluating the conformance of web content to the Guidelines. In this paper, we explore a methodology to score adherence of a web page to WCAG 2.0 conformance requirements. The paper summarizes our results (in terms of challenges and benefits) and provides a detailed description of our research on a sample web page. Finally, the paper lists some issues for further research derived from our approach. The hypothesis we used for our research was that the methodology is a feasible, practical, and accurate way to score WCAG 2.0 conformance. Our conclusion after completing our research is that the methodology, while feasible in theory, in practice is too cumbersome, too subjective, and not sufficiently reliable to be useful for real-world WCAG 2.0 conformance testing, but that the methodology may still inform the thinking of web accessibility evaluators in the future in a general sense.

Keywords:

Conformance; measure; score; structured assurance case; web accessibility; Web Content Accessibility Guidelines (WCAG).

Disclaimer:

Certain commercial entities, equipment, or materials may be identified in this document in order to describe and experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

Challenges and Benefits of a Methodology for Scoring web Content Accessibility

Guidelines (WCAG) 2.0 Conformance

Frederick E. Boland and Elizabeth Fong

1.0 Background

1.1 Motivation

The goal of web accessibility is to make web pages usable by people of all abilities and disabilities. There are social and business reasons to promote web accessibility, and it is recognized as a basic human right. It involves components of assistive technology, user agents, content design, tools, users, and developers.

It is important to measure web accessibility as objectively as possible because the law in many countries requires web sites to be accessible, and because enabling people with disabilities to use the Web is the right thing to do. However, web accessibility is difficult to measure objectively because accessibility means different things to different people. It also depends on the context of use, the user's environment, and other factors. Consequently, it is difficult to numerically "score" web accessibility in an objective sense.

The current state of practice of evaluating web accessibility involves a review of a web page to find issues, which are then reported in a list along with a brief description of those issues. The scoring is a listing by a reviewer as to which requirements pass and which ones do not. Practically, evaluators need more detail, objectivity, and rigor in their reviews. Developers, evaluators, and other stakeholders need to know where they stand in terms of their web pages being accessible, so that they can improve. Thus scoring web accessibility is a desirable goal, although it is controversial in practice, because it needs to be done consistently. In our paper, we explore applying a methodology to numerically score the content of web pages in terms of web accessibility that attempts to provide what evaluators and developers need.

A secondary purpose of the paper is to attempt to provide a framework for expressing information about accessibility findings in the context of an evaluation. Documenting findings (in terms of accessibility support) is critical for the progression of web accessibility evaluation as a discipline. However, currently, there is a lack of shareable information on particular accessibility findings.

1.2 WCAG 2.0

The World Wide Web Consortium (W3C) has developed a Web Content Accessibility Guidelines (WCAG) 2.0 Recommendation [1] (hereafter WCAG), which lists requirements (called success criteria) for web page content to conform in order to improve web accessibility.

The goal for WCAG is to provide a single, shared standard for web content accessibility that meets various needs. The scope of our methodology is WCAG conformance.

WCAG is decomposed hierarchically into four principles as follows:

- (1) perceivable – content must be presentable in ways users can perceive,
- (2) operable – users must be able to operate the interface,
- (3) understandable – users must be able to understand information/user interface operation, and
- (4) robust – content can be interpreted by a wide variety of technologies.

Figure 1 illustrates this decomposition.

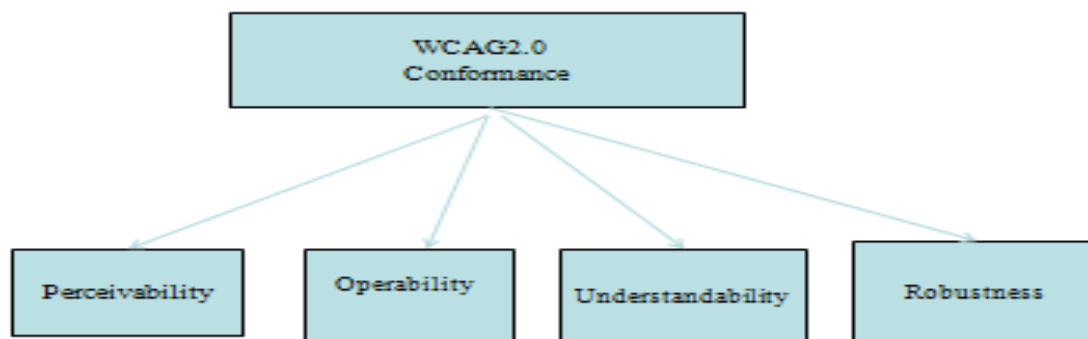


Figure 1 – WCAG Decomposition Into Principles

Each principle is further hierarchically decomposed into a list of guidelines applying to that principle. Continuing in the hierarchy, under each guideline there is a list of success criteria (testable, technology-neutral requirements). Success criteria describe what needs to be achieved for WCAG conformance. There are three “tiers” of success criteria: A, AA, and AAA. Success criteria in tier A are the easiest to meet, AA are harder to meet; and AAA are the hardest to meet.

Under each success criterion there are lists of techniques which describe possible ways of meeting that success criterion, as well as possible ways that the success criterion fails to be met. There may be other evidence at this level that illustrates how an associated success criterion is met. Examples of other evidence include evaluation tool reports and user testing results with assistive technology. [2].

It should be emphasized that in WCAG a web page either conforms or it doesn't; our methodology maintains fidelity with all aspects of WCAG. However, the methodology purports to give additional guidance to evaluators in practice to move towards the goal of full WCAG conformance.

2.0 Our Methodology

2.1 Introduction

Our research seeks to answer the question: is our methodology a feasible, practical, and accurate way to score WCAG conformance of a web page?

We designed our methodology to be consistent with WCAG conformance requirements. We based the methodology on the Structured Assurance Case technique [2]. We developed equations to score WCAG conformance for a given web page. These equations are in the form of summations of terms. Each equation at one level feeds into the next equation at the next higher level, in a hierarchical fashion. We only considered success criteria at tier A for brevity.

An important contribution we made was to use a weighting mechanism in the equations as appropriate for each equation. Thus, evaluators may assign different degrees of importance to different terms in the equations of this hierarchy, according to their particular environmental, purpose, or context considerations. These weights are computed by evaluator responses to appropriate and relevant questions in questionnaires.

In most of the equations, the summation of terms is normalized, meaning that the resultant value goes from 0 to 1. This provides uniformity to the assessment. Also, the modular approach of leveling used keeps only relevant detail at each level, so that if a change has to be made, it is only made at one place.

2.2 WCAG Conformance Equation

The WCAG conformance equation is given by

$$WC = 1/4 (P + O + U + R)$$

where WC is WCAG conformance score, P is perceiveability score, O is operability score, U is understandability score, and R is robustness score. The 0.25 is the normalization factor, since there are four principles in WCAG. The values of P, O, U, and R are computed as follows:

The equation for perceiveability score is given by

$$P = (1/9) (SC1 + SC2 + SC3 + SC4 + SC5 + SC6 + SC7 + SC8 + SC9)$$

Where P is the perceiveability score, and SC1..SC9 are scores from the nine tier A success criteria from WCAG under perceiveability. The (1/9) is the normalizing factor.

The equation for operability score is given by

$$O = (1/9) (SC10 + SC11 + SC12 + SC13 + SC14 + SC15 + SC16 + SC17 + SC18)$$

Where O is the operability score, and SC10..SC8 are scores from the nine (different) tier A success criteria from WCAG under operability. The (1/9) is the normalizing factor.

The equation for understandability is given by

$$U = (1/6) (SC19 + SC20 + SC21 + SC22 + SC23 + SC24)$$

Where U is the understandability score, and SC19..SC24 are scores from the five (different) tier A success criteria from WCAG under understandability. The (1/5) is the normalizing factor.

The equation for robustness is given by

$$R = (1/2) (SC25 + SC26)$$

Where R is the robustness score, and SC25 and SC26 are scores from the two (different) tier A success criteria from WCAG under robustness. The (1/2) is the normalizing factor.

2.3 Success Criterion Equations

For each of the success criteria SC1..SC26, we developed an equation giving a SCn score as follows:

$$SCn = wgt1 ES-1 +- wgt2 ES-2 +- \dots +- wgtm ES-m$$

Where n is a number from 1 to 26, and ES-i are evidence strength values (positive or negative), each representing a piece of evidence contributing to or detracting from a score for “meeting” SCn (may be considered as an “argument” – this term is borrowed from [2]). The number of terms on the right is unbounded, to allow an unlimited number of evidence pieces; however, if $SCn > 1$, we set $SCn = 1$, and if $SCn < 0$, we set $SCn = 0$. In this equation wgt1..m represents the weights assigned to these evidence strength values 1..m; m starts at 1 and has no fixed upper limit. We compute the evidence strength values as follows:

For each evidence strength value ES-1..ES-m, we developed an equation giving an ES-m score as follows:

$$ES-1 = (1/5) (val1) (S1 + S2 + S3 + S4 + S5)$$

Where S1..S5 represent scores for a piece of evidence based upon five factors: reproducibility, objective measurability, relevance to the associated success criterion, not subject to compromise or tampering, and degree of uncertainty or error. These scores were computed from evaluator responses to (different) relevant and appropriate questions in a questionnaire. In the equation val1 is an intrinsic evidence value, and 1/5 is a normalizing factor. We realized we needed to include an intrinsic value for the evidence because just describing the evidence in terms of its factors wasn't complete; this value is computed from evaluator responses to relevant and appropriate questions in a questionnaire.

2.4 Test-Driving Our Methodology

Once we developed our methodology, we exercised the methodology on a sample web page (hereafter called “page”). We chose a web page [3] that was known to have basic accessibility issues that had some vetting through the W3C process, and were illustrative. We reviewed both the annotated source code listing and the summary report associated with this page. Evaluators A and B were postulated to answer the questions posed for appropriate levels in the hierarchy. We worked from the bottom moving up in the hierarchy (opposite of equation design progression) to compute a WCAG conformance score for the page. We could estimate from the summary report the upper and lower bounds of a possible score for the page.

First, we gathered and organized evidence from the page, and for each piece of evidence from the page, computed evidence strength from Section 2.3 (Task 1). Next, we took those values and computed success criteria scores (Task 2). Then, we took the values from Task 2 and computed scores for each principle for the page as described in Sections 2.2 (Task 3). Finally, we took the scores for each principle and computed a WCAG conformance score for the page. The final score we arrived at was 0.608. This validated our assumption that the sample page used for the test drive was known to have accessibility problems and not a conformant page. A full description of the work is available from the authors.

3.0 Challenges and Benefits

3.1 Challenges

It took too much time and documentation to compute the WCAG conformance score for the page. Table 1 indicates the time (in hours) and the documentation (in number of printed pages) for Tasks 1 through 4 and the total for the page. Because we were moving up the hierarchy, the numbers decrease moving down the table. In practice, evaluations need to be done quickly and with a minimum of documentation. The total values given are excessive for a page of low complexity (it is a static page with several hundred lines of source code). Therefore our methodology is not practical.

Task #	Time in Hours	Number of Printed Pages
1	20	6
2	10	3
3	6	2
4	4	0.5
Total	30	11.5

Table 1: Estimated Time/Printed Pages for Our Tasks

Furthermore, subjectivity is possible in the evaluator responses to questions in the questionnaires, so our methodology is not entirely accurate (although our score for the page is within upper and lower bounds for a score for such a page).

3.2 Benefits

This methodology is feasible for scoring a web page, given enough time and resources. It is also possible to automate the scoring by building a software tool. The methodology provides a flexible structured framework for expressing what goes into computing such a score (besides just a list). The structure is flexible in the sense that evaluators are free to modify or add their own questions in any of the questionnaires or modify or add to evidence strength factors without changing the structure of the methodology. For the first time, evaluators have a consistent way of representing their findings that is shareable (which should help in alleviating evaluator objectivity concerns if evaluators use this framework for indicating their findings).

In a modeling sense, we can simulate the effect of changing one piece of evidence on the overall WCAG conformance score quickly (which promotes traceability and rigor – important for auditing purposes). It is possible to track changes consistently in a given web page over time, so stakeholders can know where they stand and can hopefully improve the accessibility of their pages. When evaluators perform reviews of web pages in practice, they can mentally consider the tenets of this methodology to improve the quality of their reviews. For example, they can ask, “How would this be considered evidence in an argument for meeting a particular success criterion?”.

Our methodology provides a means to score individual accessibility issues; thus, a developer could use this scoring to decide which issues to tackle first. This aspect could help solve a critical need in practice which has been identified to us.

4.0 Issues for Further Research

In the course of designing and applying this methodology, we identified some issues for further research.

- (1) Should multiple occurrences of an accessibility problem on the same web page be treated as one piece of evidence or separate pieces of evidence?
- (2) Is there a way to calculate intrinsic evidence value that would maximize its impact on meeting a given success criterion?
- (3) Should a given web page be assumed conformant first or non-conformant first (what is the baseline)?
- (4) Should thresholds or categories be used for scores (to reduce the effect of the granularity of a score)?

This publication is available free of charge from <http://dx.doi.org/10.6028/NIST.IR.8010>

- (5) Should statistics be used to incorporate responses from multiple evaluators on the same web page to reduce subjectivity and how should evaluator qualifications be addressed in the methodology?
- (6) How would our methodology handle dynamic or third-party web content?
- (7) Would it be better to multiply the terms rather than sum them in the equations where appropriate?
- (8) Is there a way to objectively identify missing evidence (or how much evidence is enough) in terms of meeting a success criterion?

5.0 Summary and Conclusion

We described a methodology to attempt to score WCAG conformance. We tested this methodology on a sample web page with known issues. Our findings were that our methodology, although feasible in theory, was not practical and had accuracy concerns. We surmised that we could influence web accessibility evaluators in the future and provide tracking of web pages in development. Finally, we identified issues for further research.

6.0 Acknowledgements

The authors would like to recognize the contributions of Dr. Paul E. Black, Dr. Sharon Laskowski, and Terri Hunter of NIST in providing valuable and insightful comments on the manuscript.

7.0 References

[1] W3C Web Content Accessibility Guidelines (WCAG) 2.0 Specification link:

<http://www.w3.org/TR/2008/REC-WCAG20-20081211/> ,

[2] Kelly, T., “An Introduction to Structured Argumentation Within Assurance Cases”, presentation on August 25, 2009

[3] Inaccessible Home Page Report Before and After Demonstration (BAD) link:

<http://www.w3.org/WAI/demos/bad/before/reports/home.html>