

Page Sample Size in Web Accessibility Testing: How Many Pages Is Enough?

Eric Velleman

Bartimeus Accessibility Foundation
Christiaan Krammlaan 2
3571 AX Utrecht
*313 0239 8270

evelleman@accessibility.nl

Thea van der Geest

University of Twente
P.O. Box 217
7500AE Enschede
*316 1350 1263

T.M.vanderGeest@utwente.nl

ABSTRACT

Various countries and organizations use a different sampling approach and sample size of web pages in accessibility conformance tests. We are conducting a systematic analysis to determine how many pages is enough for testing whether a website is compliant with standard accessibility guidelines. This poster reports the work-in-progress. Data collection has been completed and we have started the analysis to determine how many pages is enough for specified reliability levels.

Keywords

Accessibility testing, WCAG guidelines, page sample size, measuring web accessibility, conformance

1. HOW MANY PAGES?

In international web accessibility measurement practice, we see large differences in the number of pages that are put to the test for conformance claims. UWEM [1] suggests a page sample size of 30-50 pages. In Germany, the recommended test practice is to evaluate 3-8 pages, in France 5-20 and in the Netherlands 50 or more. Brajnic [2] argues that the page sampling approach and the page sample size can lead to big differences in accuracy and reliability of the measurement and hence the validity of the conformance claim. In line with our work on the costs and benefits of accessibility measurement [3, 4], we wondered: How many pages is enough? This poster reports the work-in-progress.

2. APPROACH

2.1 Evaluated websites

Sixty websites of national and local governments, banks and other organizations were evaluated for conformance to WCAG 1.0 priority 1 guidelines. Of the sixty website, a number was evaluated only for priority 1, the others also for the full WCAG guidelines. The mean website size was 782 pages (smallest 8,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ACM Assets Conference '13, October, 2013, Bellevue, Washington, US.
Copyright 2013 ACM 1-58113-000-0/00/0010 ...\$15.00.

largest over 4000 pages). In total over 47.000 pages were available for inspection. From the total of evaluated websites, we chose the websites that were only evaluated for priority 1 and not for additional guidelines.

2.2 Page sampling approach

Both UWEM 1.2 [1] and the Working Draft of the W3C Evaluation Methodology (WCAG-EM) [5] propose to combine a specific set of core web pages (ad hoc sampling, [2, p. 6]) and a random sample in a test. WCAG-EM proposes a core sample of common web pages, web pages with distinct common functionality, specific web page types and web pages with distinct web technologies. WCAG-EM indicates that a selected web page could have any number of these features.

Our core set consisted of 13 specific pages as described in UWEM 1.2, like home page, login page, sitemap, a complete process or transaction, a page with video or a form, etc. (Block 1).

In addition, we randomly sampled 4 blocks of 10 webpages if available (Block 2 – 5). Hence the page sample size varied from 8 pages (a complete, very small site) to 53 pages (five blocks). Because in an 8 page website, all the guidelines violations can be found in the first block, we have chosen to only use websites that have a full sample of five blocks.

2.3 Measuring accessibility

The pages in samples were inspected for WCAG 1.0 Priority 1 compliance by one of five experienced accessibility inspectors of the accredited web evaluation agency Accessibility Foundation in the Netherlands. The testing procedures followed ISO 17020 for inspection. Once a specific (unique) guideline violation was marked, it was not registered in all successive tested pages. The evaluators started with inspecting Block 1 (specific core pages) and then inspected 4 x 10 randomly selected pages, marking guideline violations that were not registered before (unique, new problems). The evaluators also registered time spent per block of pages.

3. FIRST RESULTS

Figure 1 (below) shows the percentage of the total number of unique accessibility problems (guideline violations) in a site that was found by inspecting the successive blocks. As none of the

accessibility evaluation guidelines suggests to test more than 50 pages of any given site, we use the total number of unique problems found when testing 53 pages ($13 + 4 \times 10$) as the 100% reference score of all unique accessibility problems in a specific site. The mean is calculated over the sites that were tested for priority 1 guidelines only and that have a full sample of five blocks.

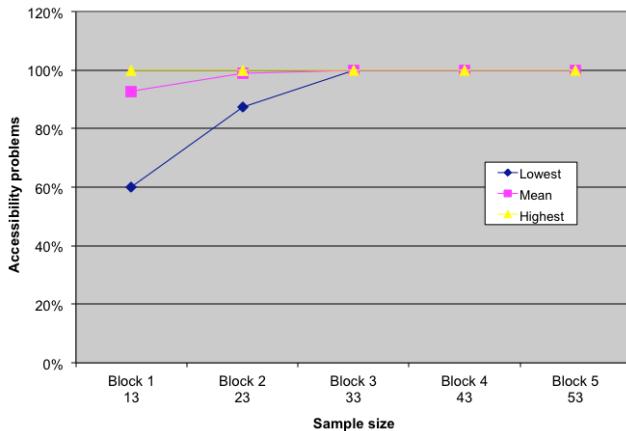


Figure 1: Percentage of unique accessibility problems identified per block

A quick first analysis of the data showed the following.

1. A core set of 13 specifically selected pages (Block 1) reveals a mean of 93 % of all unique accessibility problems (guideline violations) in a website, if we assume that inspecting 53 pages will reveal all problems.
2. The variation between websites and the variation in the yield of testing the first block is large. In one specific website, only 60% of all unique accessibility problems occurred in the Block 1 sample.
3. A mean sample of 13 specific and 10 random pages is enough to find 99% of the unique accessibility problems in a website.
4. For 68% of the websites, no new guideline violations were found after the first sample block of 13 pages (Block 1). For 92 percent of the websites, no new guideline violations were found after Block 2 (23 pages).

4. WHAT'S NEXT?

We will continue our analyses, focusing on the following set of issues.

1. Less is more! What if we distinguish among the selected pages in Block 1? What percentage of unique guideline violations is found when inspecting just one

page, the home page? What happens if we limit our sample to three or five specific pages?

2. How sure are we? We plan to calculate confidence intervals for the various additional blocks and for selections within Block 1. Also for evaluations of more than just priority 1.
3. Agreement among evaluators. A subset of the sixty sites has been inspected by two independent evaluators. Does it make a difference who is the inspector?
4. Type of accessibility problem. What type of problems are easiest to find in a page samples of various sizes?
5. Sample size and site characteristics: We have classified the sixty tested sites for their size (total number of pages) and their complexity (three levels). Do all sites require the same page sample size, or can the optimal sample size be related to site characteristics like size and complexity?
6. Cost-benefit analysis: The inspectors have been keeping time during inspection. From their records we can make an analysis of costs (in terms of time needed) against benefits (in terms of additional unique web accessibility problems that are identified).

We hope to report our additional analyses during the poster presentation at Assets 2013.

5. ACKNOWLEDGMENTS

This study could not have been conducted without the generous support of the Bartiméus Institute in the Netherlands, of the experienced testers of the Accessibility Foundation, and in particular the invaluable work of Wilco Fiers.

6. REFERENCES

- [1] UWEM, *Unified Web Evaluation Methodology* version 1.2. Retrieved from: http://www.wabcluster.org/uwem1_2/, 28 June 2013.
- [2] Brajnik, G. 2007. *Automatic testing, page sampling and measuring web accessibility*. Retrieved from <http://www.dim3.uniud.it/~giorgio/papers/csun08.pdf> , 28 June 2013.
- [3] Geest, Thea van der, Velleman, Eric & Houtepen, Martijn (2011) *Cost-benefit analysis of implementing web standards in private organizations*. Enschede: Universiteit Twente.
- [4] Velleman, Eric, & Geest, Thea van der, (2011) *Business Case Study Costs and Benefits of Implementation of Dutch Webrichtlijnen* Enschede: Universiteit Twente.
- [5] WCAG-EM, *Website Accessibility Conformance Evaluation Methodology (WCAG-EM) 1.0*. Working draft 26 February 2013. Retrieved from <http://www.w3.org/TR/WCAG-EM/>, 28 June 2013.