

The BenToWeb XHTML 1.0 Test Suite for the Web Content Accessibility Guidelines 2.0

Christophe Strobbe, Sandor Herramhof, Evangelos Vlachogiannis, Johannes Koch
and Carlos A. Velasco

Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B3001 Heverlee-Leuven
(Belgium)

christophe.strobbe@esat.kuleuven.be

University of Linz "integriert studieren - integrated study" (i3s3), Altenbergerstrasse 69, A-
4040 Linz (Austria)

sandor.herramhof@jku.at

University of the Aegean, Voulgaraktonou 30, GR11472 Exarchia, Athens (Greece)
evlach@aegean.gr

Fraunhofer-Institut für Angewandte Informationstechnik (FIT), Schloss Birlinghoven,
D53757 Sankt Augustin (Germany)

{Johannes.Koch,Carlos.Velasco}@fit.fraunhofer.de

Abstract. This paper presents a detailed description of the work carried out under the umbrella of the EU-funded project BenToWeb to develop a complete XHTML 1.0 Test Suite in regard to conformance with the Web Content Accessibility Guidelines 2.0 from the W3C. This initial work covered the Working Draft version of June 2005. A thorough evaluation involving end users is carried out at the moment of writing this paper.

Introduction

The EC-funded project BenToWeb (Benchmarking Tools and Methods for the Web¹) has a goal to develop test suites for the forthcoming Web Content Accessibility Guidelines 2.0. The development of these test suites serves a dual purpose: first, they support the WAI Working Groups in the development of support documents, such as technology-specific techniques, for WCAG 2.0; second, the test suite can be used to benchmark accessibility evaluation and repair tools (ERT). WCAG 1.0 has general techniques (core techniques), HTML techniques and CSS techniques; WCAG 2.0 will probably have general techniques, HTML techniques, CSS techniques, client-side scripting techniques and server-side techniques. The BenToWeb test suites will not cover all of these technologies, but at least XHTML 1.0 + CSS 2.0.

¹ <http://www.bentoweb.org/>

Structure of the Test Suite

In this context, a test suite is not a set of tests than can be used to validate web content, but a set of test files with accompanying metadata both for human and machine consumption. The test suite is a collection of “test cases”, where a test case consists of one or more XHTML files that implement or fail a requirement specified by a WCAG 2.0 success criterion, and an accompanying metadata file. The metadata are recorded in an XML format specially created for this purpose: Test Case Description Language (TCDL) [3]. The metadata include a description of the test file or files, the purpose of the test, a link to the success criterion that the test case is meant to implement or fail, a statement on whether the test case passes or fails the success criterion, and sometimes also test scenarios for the purpose of validation (especially for end-user evaluation).

For each WCAG 2.0 success criterion, at least two test cases need to be created: at least one that fails and at least one that passes the success criterion. When the test suite is complete and validated, running the test files through an accessibility evaluation tool should then provide data on the completeness of the tool’s coverage of WCAG 2.0 and whether it generates false positives and negatives.

There can be several types of test cases. “Atomic test cases” address only one success criterion and use only a single XHTML file (supporting files such as images or CSS style sheets do not count in this context). However, some accessibility requirements apply to sets of files instead of single files: WCAG 2.0 contains success criteria about consistency of navigational elements, and about information on a user’s location in a web site. Test cases for these success criteria use multiple XHTML files and are called “compound test cases”. However, it is also possible to create “complex test cases”: test cases that implement or fail multiple success criteria. At the time of writing, the BenToWeb test suite contains no test cases that are identified as “complex”, but the Test Case Description Language supports this. Previous research has shown that the concept of “complex test cases” is new in the area of accessibility test suites [2].

Development of the Test Suite

The development process requires that each test case moves through several steps before it is finally accepted in the test suite. Each test case starts out as a draft and is then reviewed by another accessibility or HCI expert. If any issues are found, the test case is sent back to the test case author. It is also possible that the test case contains scenarios for end-user evaluation. A test case evaluation framework can select these test cases and present them to end users. (BenToWeb’s test case evaluation framework is discussed elsewhere [1].) After evaluation and when all data are definitive, the test case is finally “accepted” into the test suite. The Test Case Description Language contains metadata related to the status of a test case.

At the time when development of the test suite started, the Web Content Accessibility Guidelines Working Group (WCAG WG) had only defined a relatively small set of HTML techniques for WCAG 2.0. At that time, there was also an HTML Test

Suite² which defined tests and contained examples of files that passed or failed these tests. Since the techniques and the tests were still at an early stage, BenToWeb chose to map the test cases directly to success criteria instead of the existing techniques and tests. Test case authors were free to draw on any documentation of techniques or failures they could find, regardless whether the source was WCAG or not.

Current State of the Test Suite

At the time of writing, the test suite contains 481 test cases for the 67 success criteria in the 30 June Working Draft of WCAG 2.0, the current draft when the development of the test suite started. These test cases contain over 530 XHTML test files (or JSP files that generate XHTML), which often uses supporting files, such as JavaScript, CSS, GIF, JPEG, WMV (audio/video), WMA, WAV, MP3 and Java applets.

Some success criteria have only two test cases, while others have more than thirty. The variability in the number of test cases per success criterion is often related to the number of XHTML elements or attributes that can be used to pass or fail a success criterion: for example, this number is much higher for Guideline 1.3 Level 1 Success Criterion (“Structures within the content can be programmatically determined”) than for Guideline 3.1 Level 3 Success Criterion 3 (“A mechanism for finding the expanded form of acronyms and abbreviations is available”). Another reason is that some success criteria in the June 2005 draft of WCAG 2.0 were unfinished or open to interpretation. Unfinished success criteria included some success criteria for Guideline 1.4 (contrast between foreground and background), success criteria for Guideline 2.3 (flashing content), and the Guideline 4.2 success criterion that depends on Guideline 2.3. Success criteria that were open to interpretation included several success criteria for Guideline 3.2, where the phrases “programmatically determined”, “change of context” and “initiated only by user action” were either too strict or too loose for what was intended. These two types of ambiguity were handled in the Test Case Description Language by setting the expected evaluation result of the test files to “cannot tell” (instead of “pass” or “fail”). Some issues, especially about changes of context, were fed back to the WCAG Working Group through their mailing list. Some success criteria are not covered by the test suite, because they are not applicable to XHTML, because they depend on the definition of a baseline (which BenToWeb chose not to do), or even because no technique to implement the success criterion could be found. On the other hand, there are a few test cases for success criteria that existed in previous drafts of WCAG 2.0, but which were removed from the June 2005 draft: the much contested success criteria about well-formedness (for XML) and validity in Guideline 4.1. For XHTML, checking well-formedness and validity against a DTD is very straightforward, so test cases that check whether an accessibility evaluation tool supports validation are useful components in a test suite.

² <http://www.w3.org/WAI/GL/WCAG20/tests/>. The test suite still exists at the time of writing, but the WCAG WG is integrating tests into the techniques documents. The ERT and WCAG WGs might set up a joint task force for the development of tests for WCAG 2.0.

Future Work

At the time of writing, the WCAG Working Group has published a new draft of WCAG 2.0 (November 2005) and is preparing a Last Call Working Draft. The most important implication of this for the test suite — and one which is inherent in the mapping of success criteria to a working draft — is that the mapping of the test cases to success criteria will need to be updated to a stable version of WCAG 2.0 (Last Call Working Draft or, eventually, Recommendation). It also means that test cases will need to be created for success criteria that were added since the Working Draft of 30 June 2005. Moreover, some existing test cases will need to be reviewed because the related success criteria have changed. Other test cases will disappear as the related success criteria no longer exist.

Another important task is the creation of more test cases for all success criteria, in order to create a larger set from which subsets can be generated. This is necessary to avoid that developers of ERT “optimize” their software for the test suite, in order that they score high in benchmarks without matching improvements in their evaluation of real web sites. Creating a bigger set of test cases from which test suites can be generated would ensure that the actual test set used to benchmark an evaluation tool is not predictable.

Acknowledgements

This work has been undertaken in the framework of the project BenToWeb — IST-2-004275-STP — funded by the IST Programme of the European Commission.

References

1. Herramhof, S, Petrie, H, Strobbe, S, Vlachogiannis, E, Weimann, K, Weber, G, Velasco C A (2006). Test Case Management Tools for Accessibility Testing. Computers Helping People with Special Needs. 10th International Conference, ICCHP 2006, Linz, Austria, 12-14 July 2006, Proceedings.
2. Strobbe, C (2005). Test-suites' State of the Art and Quality Assurance Methods for W3C Recommendations. BenToWeb deliverable D 4.1. Available at: http://www.bentoweb.org/html/BenToWeb_D4.1.html.
3. Strobbe, C, Herramhof, S, Vlachogiannis, E, Velasco C A (2006). Test Case Description Language (TCDL): Test Case Metadata for Conformance Evaluation. Computers Helping People with Special Needs. 10th International Conference, ICCHP 2006, Linz, Austria, 12-14 July 2006, Proceedings.