

Testaro

Efficient Ensemble Testing for Web Accessibility

Jonathan Robert Pool

jonathan.pool@cvshealth.com

CVS Health

Woonsocket, Rhode Island, USA

ABSTRACT

As automated web accessibility testing tools become enriched with new and improved tests, it can be impractical to leverage those advances. Each tool offers unique benefits, but effectively using multiple tools would require integrating them into a uniform testing and reporting scheme. Such integration is complex, because tools vary in what they try to detect, what they actually detect, and how they classify, describe, and report defects. Consequently, testers typically use only one tool.

Testaro[7] is a novel open-source NPM package that checks compliance with about 920 rules defined by an ensemble of 9 tools: alfa, Axe, Continuum, Equal Access, HTML CodeSniffer, Nu Html Checker, QualWeb, Testaro, and WAVE.

Attendees at the demonstration will, within 5 minutes, create jobs for Testaro, run them, and generate unified reports documenting more accessibility issues than any single tool can discover.

CCS CONCEPTS

• **Software and its engineering** → *Empirical software validation*; **Software testing and debugging**; • **Human-centered computing** → *Accessibility design and evaluation methods*; **Accessibility technologies**; **Accessibility systems and tools**.

KEYWORDS

web accessibility, accessibility testing, test automation, test efficiency

ACM Reference Format:

Jonathan Robert Pool. 2023. Testaro: Efficient Ensemble Testing for Web Accessibility. In *The 25th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '23)*, October 22-25, 2023, New York, NY, USA. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3597638.3614505>

1 INTRODUCTION

Automated testing tools for web accessibility[21] have become abundant, cheap, and effective. Multiple open-source packages and free or inexpensive APIs make it practical to run automated linters during development, run automated tests after revisions, and repair the discovered defects before starting more expensive human testing.

Permission to make digital or hard copies of part or all 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. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

ASSETS '23, October 22-25, 2023, New York, NY, USA

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0220-4/23/10.

<https://doi.org/10.1145/3597638.3614505>

Testing with an ensemble of tools, rather than only one, would be beneficial. Accessibility testing tools differ in the issues they try to detect and actually detect, and in how they classify, describe, and report issues.[3][15] Each tool discovers issues missed by others and casts doubt on other tools' findings.

The same diversity that makes it rewarding to test with an ensemble of tools also makes it difficult to do so. Integrating multiple tools into a uniform testing practice can be forbiddingly complex. Facing impractical integration costs, testers commonly trust a single tool for comprehensive web accessibility testing.

Testaro[7], an open-source NPM package, is designed to make accessibility testing with an ensemble of tools more efficient, by:

- installing all the tools
- offering a uniform configuration language to select options for all the tools
- running all the tools in a single job
- extracting results from all the tools in a standard format
- outputting the results in a single JSON job report

The demonstration will permit attendees to test URLs of their choice with Testaro and inspect the results. Attendees will prepare jobs and post-process reports with a companion server package, Testilo[8], compressing the test-running workflow into about two minutes of human effort and three minutes of automated testing.

2 RELATED WORK

Previous research has generally found that accessibility testing tools substantially complement one another.[15] Researchers have concluded that using only one testing tool is inadequate. Stanton[18], for example, observed, "no one tool gave me a complete list of issues found by the others", and concluded, "you really do have to test in multiple tools."

Pool[15] obtained similar results testing 121 web pages with 9 tools. He found no tool redundant. Classifying the rules of those tools into 245 "issues", he reported that removing one tool from the set would cause at least 7 of the reported issues to be lost. Moreover, tools reporting the same issues sometimes reported very different counts of instances. For example, 8 of the tools discovered links without accessible names, but their instance counts ranged from 2 to 240.

An in-principle solution to the complexity of ensemble accessibility testing was proposed at IBM in 2005 by Englefield, Paddison, Tibbits, and Damani [3]: a single platform that would provide user interaction, result processing, and other shared services for tools, allowing tools to be nothing more than rule engines. That solution, however, has remained an ambitious idea. Tool makers have not cooperated to implement it, and the inventory of tools remains fragmented and incoherent.

Table 1: Tools integrated by Testaro

Code	Name	Creator
alfa	alfa[16]	Siteimprove
axe	axe-core[2]	Deque
continuum	Continuum[10]	Level Access
htmlcs	HTML CodeSniffer[17]	Squiz
ibm	Equal Access[4]	IBM
nuVal	Nu Html Checker[22]	W3C
qualWeb	QualWeb[5]	Universidade da Lisboa
testaro	Testaro[7]	Testaro
wave	WAVE[6]	WebAIM

Until now, no project integrating multiple accessibility testing tools under programmatic control with standardized reporting has been discovered. Pa11y[19], kayle[12], and AATT[9] integrate 2 tools: Axe and HTML CodeSniffer. Although a11yTools[1] integrates 5 tools and 13 single-issue tests, it runs only one tool or test at a time, and only under human control.

3 ARCHITECTURE

Testaro (in contrast with the Englefield *et al.* proposal) does not depend on cooperation from tool makers. It integrates existing tools as they are.

Testaro tests the way humans do. It launches web browsers, navigates to web pages, performs actions, checks whether the pages behave as expected, and notes the results. Hence, it runs on a Windows, MacOS, or Ubuntu workstation.

Testaro is an NPM package that performs its own tests and those of 8 other tools, of which one is a remote service and the others are installed dependencies. The tools integrated by Testaro are listed in Table 1. Among them, they check compliance with about 920 rules. Testaro uses Playwright[13] to launch and control Chromium, Webkit, and Firefox browsers.

4 PROCESS

A *job* is an object giving information and instructions to Testaro. The core of a job is its *acts*, an array of instructions to be executed. Version 18.0.0 of Testaro defines 19 act types, which include actions on a page, navigations among pages, and tool executions.

When an act tells Testaro to execute one of the 9 tools, the act can specify which rules of that tool the tool should test for, which of the 3 browser types the tool should use, how granular the output should be, and other options. Here is an example of an act, telling Testaro to make the alfa tool perform tests for two of its rules:

```
{
  type: 'test',
  which: 'alfa',
  what: 'Siteimprove alfa tool',
  rules: ['r25', 'r71']
}
```

As it performs a job, Testaro adds results to the acts. At the end of the job, Testaro adds whole-job data to the job and returns this elaborated job as a *report*.

5 EFFICIENCIES

Testaro is designed to streamline tool installation and configuration. It installs all the tools and provides a uniform configuration interface. The options made available by all the tools are documented in one location and selected in the job file with a uniform syntax.

Testaro simplifies the task of executing multiple tools. A single job file tells Testaro which tools to launch in which order, and Testaro runs them all. A job that includes all the tests of all the tools typically takes about 3 minutes. If that were not fast enough, execution could be further accelerated with job partitioning: installing Testaro on multiple workstations, having them perform complementary jobs in parallel, and combining their reports.

An instance of Testaro can be configured as an on-call agent. It polls a server for jobs. When the server replies by sending a job, Testaro performs it and sends the report to the server.

Finally, Testaro is designed to make the utilization of tool reports more efficient. For this purpose, Testaro translates the most common elements of native tool reports into standard results. Fully documented in the README.md file, the standard results uniformly present each tool's reports of violations of its rules, including what rule was violated, how serious the estimated impact is (on a 0-to-3 ordinal scale), what HTML element was involved, where on the page it appeared, and an excerpt from the HTML code.

Here is an example of an entry from a standard result:

```
{
  totals: [23, 11, 6, 8],
  instances: [
    {
      ruleID: 'image-no-alt',
      what: img element has no text alternative,
      count: 1,
      ordinalSeverity: 3,
      tagName: 'IMG',
      id: 'ocean-beach-sunset',
      location: {
        doc: 'dom',
        type: 'xpath',
        spec: '/html/body/div[4]/p[2]/img[1]'
      },
      excerpt: 
    },
    ...
  ]
}
```

In this example, a tool reported 23 instances of rule violations at severity 0, 11 at severity 1, etc. The first reported instance was an *img* element that violated a rule named *image-no-alt*.

Given the diverse ontologies of the tools, any standardization reflects some judgment. An example is the *ordinalSeverity* property, which interprets and combines the tools' various classifications of severity, priority, and certainty. Users are free to rely on the standardization performed by Testaro to simplify report consumption, but, if they want more control, they may extract data from original tool results, too.

Table 2: Targets

ID	Description	URL
w3c	World Wide Web Consortium	https://www.w3.org/
mozilla	Mozilla Foundation	https://foundation.mozilla.org/en
wikFnd	Wikimedia Foundation	https://www.wikimedia.org/
acm	Association for Computing Machinery	https://www.acm.org/

6 CUSTOMIZATION

Effective accessibility management requires checking conformity not only to industry standards such as the Web Content Accessibility Guidelines[20] (WCAG), but also to rules (brand standards, design systems, etc.) of one’s own organization. In a multi-tool integrator, each tool is potentially a platform for the creation of custom rules, and the set of tools is extensible. Users can customize Testaro by any of these methods:

- creating a tool and adding it as an installed dependency
- creating a tool and adding it as a remote service
- extending any of the tools, if it permits, by adding new rules to it

The Testaro tool contains a template for the creation of custom rules. Existing Testaro rules are typically defined in 10 to 30 lines of code. A custom rule would likely require a similar amount of code.

7 JOB PREPARATION

For routinized use of Testaro, job preparation can be partly automated. One package performing this function is Testilo[8]. The user can create files that answer the questions “What tests do you want to run?” and “What *targets* do you want to test?”. Testilo can convert those files to a job that Testaro will execute. Table 2 gives an example of data that might be in a target file.

8 REPORT ENHANCEMENT

A JSON report from Testaro narrows the gap between native tool reports and user-friendly reporting, but does not close that gap. A report contains standard results, but they are presented sequentially, tool by tool, and the result from each tool describes violations of that tool’s rules, not of universally defined norms. Users will often want to:

- map the tool rules onto a set of tool-agnostic issues
- gather the complaints of all tools about each issue into one place
- aggregate the issue reports into a total accessibility score
- export scores for use in dashboards or reports
- summarize the JSON report in a developer- or manager-friendly HTML document
- collect scores from reports on related targets into a comparative report

To perform such functions, users can create procedures and/or use Testilo. To interpret tool rules, Testilo offers a rule classifier that maps the approximately 920 tool rules onto about 260 *issues*. For example, two tools have rules prohibiting broken same-page links. One is AAA.2_4_1.G1,G123,G124.NoSuchID from htmlcs, and the other is link_internal_broken from wave. Testilo maps

both of these onto an internalLinkBroken issue and references WCAG Success Criterion 1.3.1 as the most relevant standard.

9 DEMONSTRATION

In the demonstration, a simple web service will ask each user for the URL of a web page to be tested. The service will use Testilo to create a job for Testaro. Testaro will perform the job and convert all the tools’ results into standard results. As the final step, Testilo will convert the Testaro report to a human-oriented web page.

10 FUTURE WORK

Some engineers using Testaro for accessibility testing have requested richer and more tailored issue reports, with better identification of instance locations, consolidation of duplicates, and resolution of tool disagreements.

Such improvements will require more work on instance identification. Some tools may locate instances by line number, others by XPath, others by CSS selector, others by bounding box, and others only with an HTML code excerpt. Further work could aim to determine when instances reported by various tools are the same and to supply images of, and links to, instances.

Empirical data from the use of Testaro may facilitate rule sequencing, pruning, and deprecation; bug reports to tool makers; and the training of machine learners in the prediction of accessibility issues.

Testaro welcomes contributions to improve functionality, reliability, and issue coverage.

11 CONCLUSION

As makers of testing tools innovate to narrow the gaps[14][11] between formal and practical accessibility, tools continue to complement each other. Accessibility testing with an ensemble of tools is complex but valuable, and it can be made practical even without the cooperation of tool makers.

ACKNOWLEDGMENTS

I acknowledge valuable editorial comments and research from Susan M. Colowick.

My opinions expressed herein are my own views and do not necessarily reflect the views of CVS Health, its affiliates, or any of my colleagues at CVS Health or its affiliates.

REFERENCES

- [1] Paul J. Adam. 2023. *a11yTools Extension for Safari macOS*. Paul J. Adam. Retrieved 2023-08-09 from <https://pauljadam.com/extension.html>
- [2] Deque. 2023. *axe-core*. Deque. <https://www.npmjs.com/package/axe-core>
- [3] Paul Englefield, Claire Paddison, Mark Tibbits, and Isha Damani. 2005. *A proposed architecture for integrating accessibility test tools*. IBM. Retrieved 2023-08-06

- from https://web.archive.org/web/20060822072819id_/http://researchweb.watson.ibm.com:80/journal/sj/443/englefield.pdf
- [4] IBM. 2023. *Equal Access Accessibility Checker*. IBM. <https://www.npmjs.com/package/accessibility-checker>
- [5] Informática, Universidade da Lisboa. 2023. *QualWeb*. Universidade da Lisboa. <https://www.npmjs.com/package/@qualweb/core>
- [6] Institute for Disability Research, Policy & Practice, Utah State University. 2023. *WAVE*. Utah State University. <https://wave.webaim.org/api/>
- [7] Jonathan Pool. 2023. *Testaro*. Jonathan Pool. <https://www.npmjs.com/package/testaro>
- [8] Jonathan Pool. 2023. *Testilo*. Jonathan Pool. <https://www.npmjs.com/package/testilo>
- [9] Nawaz Khan. 2023. *AATT*. PayPal. <https://www.npmjs.com/package/aatt>
- [10] Level Access. 2023. *Continuum Community Edition*. Level Access. <https://www.webaccessibility.com/tools/>
- [11] Manuel Matuzović. 2019. *Building the most inaccessible site possible with a perfect Lighthouse score*. Manuel Matuzovic. Retrieved 2023-06-19 from <https://www.matuzo.at/blog/building-the-most-inaccessible-site-possible-with-a-perfect-lighthouse-score/>
- [12] Jeff Mendez. 2023. *kayle*. Jeff Mendez. <https://www.npmjs.com/package/kayle>
- [13] Microsoft. 2023. *Playwright*. Microsoft Corporation. <https://playwright.dev>
- [14] Jakob Nielsen. 2005. *Accessibility Is Not Enough*. Nielsen Norman Group. Retrieved 2023-06-19 from <https://www.nngroup.com/articles/accessibility-is-not-enough/>
- [15] Jonathan Robert Pool. 2023. *Accessibility Metatesting: Comparing Nine Testing Tools*. CVS Health. Retrieved 2023-06-19 from <https://arxiv.org/abs/2304.07591>
- [16] Siteimprove. 2023. *alfa*. Siteimprove. <https://alfa.siteimprove.com/>
- [17] Squiz. 2023. *HTML CodeSniffer*. Squiz Labs. https://www.npmjs.com/package/html_codesniffer
- [18] Paul Stanton. 2018. *Which accessibility testing tool should you use?* Jadu. Retrieved 2023-06-19 from <https://medium.com/pulsar/which-accessibility-testing-tool-should-you-use-e5990e6ef0a>
- [19] Team Pa11y. 2023. *Pa11y*. Team Pa11y. <https://github.com/pa11y/pa11y#runners>
- [20] World Wide Web Consortium. 2018. *Web Content Accessibility Guidelines (WCAG) 2.1*. World Wide Web Consortium. Retrieved 2023-01-13 from <https://www.w3.org/TR/WCAG21/>
- [21] World Wide Web Consortium. 2023. *Introduction to Web Accessibility*. World Wide Web Consortium. Retrieved 2023-01-12 from <https://www.w3.org/WAI/fundamentals/accessibility-intro/>
- [22] World Wide Web Consortium. 2023. *Nu Html Checker*. World Wide Web Consortium. <https://github.com/validator/validator>

Received 23 June 2023; accepted 6 August 2023; revised 18 August 2023