Software Prototype for the Ensemble of Automated Accessibility Evaluation Tools

Peter Johnson[1](✉) and Mariana Lilley[2]

[1] Google Inc., London, UK
peterljohnson@google.com
[2] Department of Computer Science, University of Hertfordshire, Hertfordshire AL10 9AB, UK

Abstract. Web accessibility evaluation is concerned with assessing the extent to which web content meets accessibility guidelines. Web accessibility evaluation is typically conducted using manual inspection, user testing and automated testing. The process of automating aspects of accessibility evaluation is of interest to accessibility evaluation practitioners due to manual evaluations requiring substantial time and effort [1]. The use of multiple evaluation tools is recommended [9, 9]; however, aggregating and summarising the results from multiple tools can be challenging [1].

This paper presents a Python software prototype for the automatic ensemble of web accessibility evaluation tools. The software prototype performs website accessibility evaluations against the WCAG 2.1 AA guidelines by utilising a combination of four free and commercial evaluation tools. The results from the tools are aggregated and presented in a report for evaluation.

The tool enables practitioners to benefit from a coherent report of the findings of different accessibility conformance testing tools, without having to run each separately and then manually combine the results of the tests. Thus, it is envisaged that the tool will provide practitioners with reliable data about unmet accessibility guidelines in an efficient manner.

Keywords: Accessibility · Conformance testing · WCAG 2.1 · Inclusive design

1 Introduction

The web was designed to be accessible to all people regardless of their individual differences, use of hardware, software, language and location. Web accessibility means designing websites, tools and technologies to be inclusive of all users irrespective of their impairment (whether permanent, temporary or situational) so that everyone can perceive, understand, navigate and interact with the web and contribute to it.

Web accessibility evaluation means verifying that this is the case [15]. It should be noted that web accessibility goes beyond ethical and legal requirements; the relationship between user experience and accessibility is also well-documented in the literature (see, for example, [4] and [11]).

There are three main methodologies for conducting web accessibility evaluations: manual inspection, user testing and automated testing [1].

C. Stephanidis et al. (Eds.): HCII 2022, CCIS 1580, pp. 532–539, 2022.
https://doi.org/10.1007/978-3-031-06417-3_71

Manual inspections are conducted by expert evaluators who usually check a webpage against a checklist of evaluation criteria based on accessibility guidelines [6] and are often used during the design process [1].

User testing is, in principle, the most reliable accessibility evaluation approach as it typically involves expert evaluators observing a sample of representative users performing a set of carefully designed tasks [1]. User testing, however, can be slow and expensive [1], and results may not be reliable if the sample population does not accurately reflect the target user population [8].

Automated testing makes use of software tools running locally or online to parse the source code to identify unmet accessibility Success Criteria by executing a set of rules that are based on guidelines such as WCAG 2.1. There are several automated accessibility evaluation tools available and W3C maintain a comprehensive list of tools [16]. The automation of accessibility evaluations is of interest to practitioners within the field of web accessibility evaluation as automated testing significantly reduces the time, effort and thus cost to perform aspects of web accessibility evaluation [1]. Additional benefits of automated testing include [10]:

- More predictable resource requirements for evaluations such as time and cost;
- Greater consistency in detecting errors, and less prone to human error;
- Broader evaluation scope within resource constraints, for example, a tool can evaluate 100 pages which may not be possible with a manual evaluation;
- Easier for inexperienced testers to perform accessibility evaluations;
- Easier to enable accessibility guideline checks during development.

To gain a deeper understanding of the use of such tools in web accessibility evaluation, a literature survey was conducted in June 2021.

2 Literature Survey

The literature survey focused on papers published between January 2017 and June 2021, where the abstract contained the keywords ‘wcag’ and ‘tools’, and the language of publication was English.

A total of 123 papers were analysed, and the work reported in this paper focuses on the review of 50 papers where:

- An accessibility evaluation of website(s) was performed;
- Web accessibility evaluation tool(s) were used.

As can be seen from Table 1, the most frequently used web accessibility evaluation tools were WAVE [17], AChecker [3] and TAW [12]. From the 3 most frequently used tools, WAVE was the only tool that supports the current version of the WCAG guidelines, WCAG 2.1.

Research by Campoverde-Molina et al. [7] also found WAVE, AChecker and TAW to be the most frequently used tools for WCAG 2.0 web accessibility evaluations. The increased use of WAVE was also noted [7].

Table 1. Summary of most frequently used web accessibility evaluation tools.

Tool	Frequency	WCAG version	Conformance levels	License
WAVE	26	WCAG 2.1	A, AA	Free/Commercial
AChecker	19	WCAG 1.0, WCAG 2.0	A, AA, AAA	Free
TAW	11	WCAG 2.0	A, AA, AAA	Free
SortSite	4	WCAG 2.0, WCAG 2.1	A, AA, AAA	Commercial
Total Validator	3	WCAG 1.0, WCAG 2.1	A, AA, AAA	Free/Commercial
Access Monitor	3	WCAG 2.1	A, AA, AAA	Free
Axe	3	WCAG 2.0, WCAG 2.1	A, AA	Free/Commercial
EIII Page Checker	2	WCAG 2.0	A, AA	Free
SiteImprove	2	WCAG 2.1	A, AA, AAA	Commercial
Cynthia Says	2	WCAG 2.0	A, AA, AAA	Free

From the 50 selected papers, 35 (70%) only used tools to conduct their web accessibility evaluation. The remaining 15 (30%) used a combination of manual evaluation techniques and web accessibility evaluation tools. For the most part, manual evaluation techniques were used to analyse and collect the data from the tools used rather than used for identifying errors.

The rationale for employing web accessibility evaluation tools included detecting errors more consistently [10], greater objectivity [5, 5], producing comparable results to other forms of testing [13], and being less time and labour intensive [5].

Table 2. Summary of number of tools employed.

Number of tools employed	Frequency
1	24 (48%)
2	18 (36%)
3	7 (14%)
4	1 (2%)

An analysis of the number of tools used can be seen in Table 2 above. The analysis shows that the most frequently occurring number of evaluation tools used in these studies was 1, accounting for 48% of the studies.

The use of multiple web accessibility evaluation tools is recommended [9, 9], as different tools may produce different evaluation results for the same page [1, 1]. One of the reasons for this is that guidelines are implemented by different code implementations, algorithms and search/matching techniques [1].

Despite the predicted benefits of using multiple web accessibility evaluation tools, automated solutions to aggregate results from different tools are under-represented in the literature. To address this gap, a pilot study involving the design and implementation of a tool that aggregate results from multiple web accessibility evaluation tools was conducted and is presented next.

3 Pilot Study

The pilot study consisted of two phases. In the first phase, a software prototype was designed and developed. The second phase consisted of an empirical study with evaluators.

3.1 Software Prototype (First Phase)

A Python software prototype for the automatic ensemble of web accessibility evaluation tools was designed and developed (see Fig. 1). The aim of the prototype was to provide an automated approach to tool execution and results processing, so that the execution of multiple evaluations tools is simplified and results can be aggregated and easily displayed for analysis.

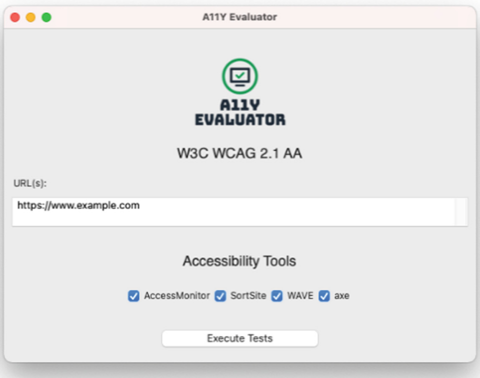

Fig. 1. Software prototype screenshot.

Abascal et al. [1] suggest 3 as the number of tools likely to yield useful results for the purposes of web accessibility evaluation, and 4 tools were selected for the pilot study to increase likelihood of useful results. The software prototype performs website accessibility evaluations against the WCAG 2.1 AA guidelines using WAVE, Sort-Site, AccessMonitor and Axe. The results from the tools are aggregated and presented in a report for evaluation.

The software tool ensembles the results from each of the web accessibility evaluation tools in an aggregated report that shows the unmet WCAG 2.1 Success Criteria. Table 3 below shows the results generated by the software prototype after evaluating the homepage of a UK Higher Education Institution (HEI) website.

Table 3. Results of unmet WCAG 2.1 Success Criteria from ensemble for UK HEI homepage. The results of each tool have been anonymized for the purpose of this paper.

Success criteria	Tool A	Tool B	Tool C	Tool D	Total
1.1.1 – Non-text Content	0	0	6	0	6
1.3.1 – Info and Relationships	0	0	2	1	3
1.4.11 – Non-text Contrast	0	5	0	0	5
1.4.3 – Contrast (Minimum)	1	0	25	1	27
1.4.4 – Resize text	15	0	0	0	15
2.4.4 – Link Purpose (In Context)	3	2	4	3	12
2.4.6 – Heading and Labels	0	0	2	0	2
2.4.7 – Focus Visible	0	5	0	0	5
3.3.2 – Labels and Instructions	0	0	2	0	2
4.1.1 – Parsing	14	0	0	1	15
4.1.2 – Name, Role, Value	3	3	0	7	13
Total	36	15	41	13	105

The software prototype also generates several graphs illustrating unmet WCAG 2.1 guidelines as illustrated in Fig. 2 below.

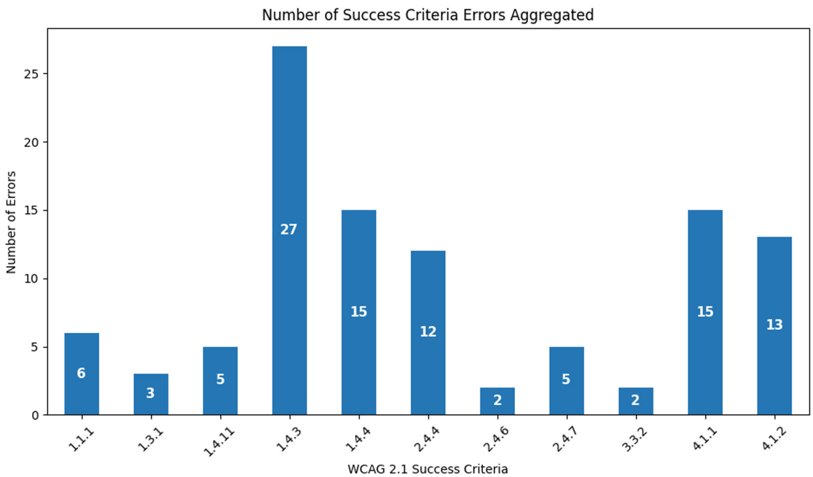

Fig. 2. Example of a graph of aggregated unmet WCAG Success Criteria generated by the software prototype.

3.2 Manual Web Accessibility Evaluation (Second Phase)

To evaluate the effectiveness of this software prototype, a group of 4 web accessibility evaluation practitioners were invited to assess two UK HEI homepages (homepage 'A' and homepage 'B') using WebAIM's WCAG 2 Checklist [18]. The use of 4 evaluators is consistent with similar studies (see, for example, [14]). To control for order effect, the evaluators were split into 2 groups, one group which evaluated homepage 'A' followed by homepage 'B'; and the second group evaluated homepage 'B' followed by homepage 'A'. Whilst evaluators were not asked to formally report on how long they took to conduct the evaluation, they informally reported that the process took at least 45 min.

Results from the expert evaluation were then compared with the aggregated report automatically generated by the software prototype:

- The following unmet WCAG level AA Success Criteria (SC) were missed by the software prototype: 1.3.2 Meaningful Sequence; 1.3.4 Orientation; 1.3.5 Identify Input Purpose; 1.4.11 Non-text Contrast; 1.4.13 Content on Hover or Focus; 2.1.1 Keyboard; 2.1.2 No Keyboard Trap; 2.2.2 Pause, Stop, Hide; 2.4.1 Bypass Blocks; 2.4.3 Focus Order; 2.4.7 Focus Visible; 2.5.2 Pointer Cancellation; 3.2.4 Consistent Identification; 3.3.1 Error Identification; 3.3.3 Error Suggestion.
- The following unmet WCAG level AA Success Criteria (SC) were missed by two or more human evaluators: 1.4.4 Resize Text; 2.4.4 Link Purpose (In Context); 3.3.2 Labels or Instructions; 4.1.1 Parsing; 4.1.2 Name, Role, Value.

An important finding was that accessibility evaluation tools appear to be better suited for finding unmet accessibility guidelines within the WCAG Robust principle and more specifically for SC 4.1.1; similar findings were reported by Frazão and Du-arte [9] and Vigo et al. [14]. Most unmet SC missed by evaluation tools were within the WCAG Perceivable and Operable principles; Vigo et al. [14] suggest that some SC require more than parsing techniques; for example, "ascertaining whether there are keyboard traps ("2.1.2 No Keyboard Trap") requires real interaction or simulation" (p. 8).

4 Conclusion

Web accessibility evaluation is crucial ethically, legally and as a foundation for usa-bility. Web accessibility evaluation requires human input; consistent with [2], it was found that tools cannot replicate the human experience [2], some aspects of web accessibility evaluation cannot currently be automated [14] and particular tools may present false positives or false negatives, and thus require human judgement [6].

Notwithstanding this, web accessibility evaluation tools remain central to enabling accessibility practitioners in determining if web content meets accessibility guidelines. Some of the anticipated benefits of using automated tools include being less time and labour intensive than manual approaches and producing objective results in criteria that is better suited for automated approaches (e.g. HTML parsing).

Research in the field of web accessibility evaluation tools indicates that practitioners are often required to run multiple evaluation tools to overcome the limitations of and reliance on a single tool [6, 6, 6]. Indeed, Abascal et al. [1] report on how "evaluators are often obliged to apply more than one automated tool and then to compare and aggregate the results in order to obtain better evaluation results" (p. 488). Abascal et al. [1] also report on how aggregating and summarising the results from diverse tools can be difficult. No solution to address this need has been found in the literature survey.

In the work presented here, a software prototype is employed for the automatic ensemble of four web accessibility evaluation tools, so that practitioners can best leverage results from existing tools and address the gap identified above. The results from the tools are aggregated and presented in a report for evaluation. The tool enables practitioners to benefit from a coherent report of the findings of different accessibility conformance testing tools without having to run each separately and then combine the results of the tests. Initial findings from the pilot study indicate that the tool has the potential to provide practitioners with reliable data about unmet accessibility guidelines in an efficient manner.

Findings from the literature survey suggest that the automated approach used in this work for results gathering and reporting is novel and merits future work.

5 Future Work

It is planned that web accessibility evaluation practitioners will be involved in refining the reporting of unmet Success Criteria (SC), so that the tool can be more effective at supporting practitioners' evaluation work. In particular, we will be looking at how the results from the evaluation should be presented (for example, how to best report that no errors were found for a given Success Criteria) and whether or not some form of dashboard to support evaluators in prioritising areas that require manual and/or user testing would be a desirable feature.

Additionally, we will be looking at how comparing results from different tools may enable practitioners to identify 'false positives' and 'false negatives' generated by web accessibility evaluation tools [6] more efficiently.

References

1. Abascal, J., Arrue, M., Valencia, X.: Tools for web accessibility evaluation. In: Yesilada, Y., Harper, S. (eds.) Web Accessibility. HIS, pp. 479–503. Springer, London (2019). https://doi.org/10.1007/978-1-4471-7440-0_26
2. Accessibility Evaluation Tools. https://webaim.org/articles/tools/. Accessed 06 March 2022
3. AChecker Web Accessibility Checker. https://achecker.achecks.ca/checker/index.php. Accessed 06 March 2022
4. Aizpurua, A., Harper, S., Vigo, M.: Exploring the relationship between web accessibility and user experience. Int. J. Hum Comput. Stud. **91**, 13–23 (2016)
5. Alim, S.: Web Accessibility of the top research-intensive universities in the UK. SAGE Open, **11**(4). https://doi.org/10.1177/21582440211056614, https://journals.sagepub.com/doi/pdf/. Accessed 06 March 2022
6. Brajnik, G.: Comparing accessibility evaluation tools: a method for tool effectiveness. Univ. Access Inf. Soc. **3**(3–4), 252–263 (2004)
7. Campoverde-Molina, M., Luján-Mora, S., Valverde, L.: Accessibility of university websites worldwide: a systematic literature review. Univ. Access Inf. Soc. 1–36 (2021). https://doi.org/10.1007/s10209-021-00825-z

8. Eraslan, S., Bailey, C.: End-user evaluations. In: Yesilada, Y., Harper, S. (eds.) Web Accessibility. HIS, pp. 185–210. Springer, London (2019). https://doi.org/10.1007/978-1-4471-7440-0_11
9. Frazão, T., Duarte, C.: Comparing accessibility evaluation plug-ins. In: Proceedings of the 17th International Web for All Conference (2020)
10. Ivory, M., Hearst, M.: The state of the art in automating usability evaluation of user interfaces. ACM Comput. Surv. **33**(4), 470–516 (2001)
11. Petrie, H., Kheir, O.: The relationship between accessibility and usability of websites. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 397–406, April 2007
12. TAW. https://www.tawdis.net/. Accessed 06 March 2022
13. Verkijika, S.F., De Wet, L.: Accessibility of South African university websites. Univ. Access Inf. Soc. **19**(1), 201–210 (2018). https://doi.org/10.1007/s10209-018-0632-6
14. Vigo, M., Brown, J., Conway, V.: Benchmarking web accessibility evaluation tools. In: Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility - W4A 2013 (2013)
15. W3C. Introduction to Web Accessibility, Web Accessibility Initiative (WAI). https://www.w3.org/WAI/fundamentals/accessibility-intro. Accessed 06 November 2021
16. W3C. Web Accessibility Evaluation Tools List. https://www.w3.org/WAI/ER/tools. Accessed 08 March 2022
17. WAVE Web Accessibility Evaluation Tool. https://wave.webaim.org/. Accessed 06 March 2022
18. WebAIM's WCAG 2 Checklist. https://webaim.org/standards/wcag/checklist. Accessed 08 March 2022