Can a Web Accessibility Checker Be Enhanced by the Use of AI?

E. A. Draffan\(^{(*)}\), Chaohai Ding, Mike Wald, Harry Everett, Jason Barrett, Abhirami Sasikant, Calin Geangu, and Russell Newman

WAIS, ECS, University of Southampton, Southampton, UK
{ead,cd8e10,mw}@ecs.soton.ac.uk

Abstract. There has been a proliferation of automatic web accessibility checkers over the years designed to make it easier to assess the barriers faced by those with disabilities when using online interfaces and content. The checkers are often based on tests that can be made on the underlying website code to see whether it complies with the W3C Web Content Accessibility Guidelines (WCAG). However, as the type of code needed for the development of sophisticated interactive web services and online applications becomes more complex, so the guidelines have had to be updated with the adoption of new success criteria or additional revisions to older criteria. In some instances, this has led to questions being raised about the reliability of the automatic accessibility checks and whether the use of Artificial Intelligence (AI) could be helpful. This paper explores the need to find new ways of addressing the requirements embodied in the WCAG success criteria, so that those reviewing websites can feel reassured that their advice (regarding some of the ways to reduce barriers to access) is helpful and overcomes issues around false positive or negatives. The methods used include image recognition and natural language processing working alongside a visual appraisal system, built into a web accessibility checker and reviewing process that takes a functional approach.

Keywords: Digital accessibility • Disability • Automatic checkers • Artificial intelligence

1 Introduction

Over the past twelve years, a Web2Access\(^{1}\) system of 15 accessibility checks has been used as a functional review system for the accessibility of websites used for elearning by students, teachers and other academics. Anyone could access the reviews or add an evaluation that went some way to ensuring the online services listed highlighted possible barriers for people with disabilities. The checks were originally based on the W3C Web Content Accessibility Guidelines (WCAG) version 2.0 [1] and took the user on a journey through a website. They started at the login stage catching the issues that might arise with reCAPTCHAs and unlabelled forms. The reviewer then went on to test

---

\(^{1}\) https://web2access.org.uk/.

© Springer Nature Switzerland AG 2020
K. Miesenberger et al. (Eds.): ICCHP 2020, LNCS 12376, pp. 67–73, 2020.
https://doi.org/10.1007/978-3-030-58796-3_9
for a lack of alternative text for images, style sheets that changed the navigation and then to the type of page that might require checks involving the use of keyboard only access, magnification and colour contrast levels. Other important interactive elements included in the review list were videos and audio accessibility, appropriate feedback from forms and access to tables, page integrity and text styles etc. to encompass the concept of readability. A recent experimental study of a Thai translation of Web2access showed that it could be used reliably by novice developers to predict the accessibility of websites where barriers existed for those with disabilities [2].

Updates to WCAG 2.1 with the addition of seventeen new success criteria [3] meant that the Web2Access review system had become outdated and it was time to follow others looking into the potential of AI as a method to support checks [4]. The original Web2Access tests required additional elements and the method for the reviews with a mix of automatic and manual checks needed to be overhauled. It was necessary to abide by the UK Government’s guidance stating that web accessibility compliance must include the specific requirements mentioned in the WCAG 2.1 Success Criteria (SC) that are ‘testable statements’ at levels A and AA. This meant that the update to Web2Access needed to include five additional success criteria at Level A and seven at Level AA. Furthermore, in order to allow the reviewers the chance to evaluate more than one web page at a time the Web Accessibility Conformance Evaluation Methodology (WCAG-EM)2 was included in the update. This enabled the team to produce an automated accessibility statement that could be added to a website as stipulated in the recent Public Sector Bodies (Websites and Mobile Applications) (No.2) Accessibility Regulations 2018 [5].

2 Method

Much of the original database design of Web2Access remained in place with individual tests having additional text added to incorporate the extra information required for the updated success criteria. Some of the tests were merged to allow for the new ones from WCAG 2.1. Particular attention was paid to the additional success criteria that included Orientation (SC1.3.4), Non-text Contrast (SC1.4.11), Text Spacing (SC1.4.12) and Label in Name (SC2.5.3) that is important when form filling as these were the tests that it was felt could be automated.

The web accessibility checker developed to assist with the reviews required a new build using an accessible React, a JavaScript interface3 with additional ARIA attributes. The accordion design involved the participation of five accessibility experts at every stage. Each expert had over ten years’ experience in the field, all used assistive technologies in their day-to-day work and were regularly evaluating the accessibility of digital content. As per user-centred design principles [6], their involvement was ongoing throughout the project, even when testing the outcomes of the checker and commenting on general concerns about the reliability of automated accessibility testing tools [7].

---

2 https://www.w3.org/WAI/test-evaluate/conformance/wcag-em/.
3 https://reactjs.org/.
Behind the interface that presented the results of multiple tests, the team implemented the use of the open source Pa11y\(^4\) accessibility checker with the additional WCAG 2.1 checks. An innovative series of visual appraisal pages were integrated within the drop down results to allow the reviewer to see where issues might arise when the automated checking could have produced false positives or negatives. A false positive was considered an accessibility check that returned a mistake that did not actually exist or did not affect the use of the page and a false negative was one where there might have been an issue, but the automatic checker had not captured it. Algorithms were developed to offer the reviewer the chance to check the results of two particular issues that were arising. Where the success criteria required, for instance no overlaps when using text spacing, a visual representation was designed to outline areas where there were suspected issues. In terms of false negative concerns, the issue was mainly about the alternative text description being accepted, but by adding additional checks there appeared to be a mismatch between the content in the ‘alt tag’ and the actual image. Once again, a visual representation of the image was supplied, alongside the alternative description. The reviewer could then accept or reject the results of the automated tests. All the tests once completed remain available via the Web2Access database if public mode is chosen, or private to the registered reviewer carrying out the checks.

AI models were used when it came to evaluating the accuracy of the alternative text offered for images. Pre-trained neural networks model based on the MobileNet\(^23\) image classifier and the COCO-SSD\(^24\) model for object detection and classification provided a comparison between the actual image used and the alternative text provided. As mentioned, the output could be seen on the appropriate visual appraisal page, providing the reviewer with the opportunity to make a final decision should there be any doubt resulting from the automatic check.

As an addition to the completed work there remains the intention to use the work of Sen (2019) \([8]\) and to explore further the issue related to hypertext links that fail to comply with the WCAG 2.1 Success Criterion 2.4.4 Link Purpose (In Context). This is possible by capturing groups of words, such as three words before and after the hyperlink from the source and then comparing this with a similar amount of words at start of the target page using word-embedding techniques. “A two-layer neural net that processes text by “vectorizing” words”\(^5\) called word2vec \([9]\) was chosen as this works with word association and Word Movers Distance (WMD) \([10]\). Sen hypothesised that if the WMD scores were low this should show that the target text and the link text would be helpful to users.

---

\(^4\) https://pa11y.org/.

\(^5\) https://pathmind.com/wiki/word2vec.
3 Results

The team never intended to evaluate the success or otherwise of the Pa11y automatic checker but to see whether it was feasible to use AI for some of the new WCAG 2.1 success criteria and would a visual appraisal system reduce the questions that might arise from possible false positives or negatives with some checks. A series of website reviews were performed on a sample taken from the top 500 sites, according to alexa.com\textsuperscript{6}. No false positives or negatives were discovered when double checks were made for the checks of the top five sites; where five pages were selected from Google, YouTube, Tmall, Facebook or Qq. All the issues appeared to be relevant when manually checked, even those on the two Chinese sites (Tmall and Qq). Where results were negative, the code producing the fault appeared in the collapsible content under the various sections for each success criteria on the results page, once the checker had gone through all the pages selected (Fig. 1).

![Fig. 1. Google home page issue with WCAG SC 1.4.11 - Non-text contrast](https://example.com/image1)

It was also possible to see the additional results for those success criteria using the visual appraisal system, where issues could be checked for accuracy, such as the accuracy of an alt attribute (tag) providing the alternative text description (Fig. 2).

![Fig. 2. qq.com website where a Chinese logo was captured and checked for its alternative text](https://example.com/image2)

---

\textsuperscript{6} https://www.alexa.com/topsites.
Concerning the contextual hyperlink detection task, a word2vec model was generated based on the contextual hyperlink and the description of target page, compared against other pre-trained word embedding models. The results for the model created on the generated dataset looked promising, but further work was needed to develop a specialised dataset, with the use of Natural Language Processing to improve matches. This result meant that the test could not be included in the final accessibility checker, before the evaluation of the project was completed, due to time constraints. The team agreed that without robust results, this test would also be a concern when outputting content to the accessibility statement (Fig. 3).

![Accessibility statement for https://google.co.uk](https://example.com/image)

**Fig. 3.** Output to the accessibility statement that can be copied to a reviewer’s website.

The five digital accessibility experts tested the application at various times during the design and implementation phases. As the project had to be completed in a few months from design phase to completion, the number of people able to finally evaluate the accessibility checker for usability purposes was low. However, Nielsen stated in 2000; “The best results come from testing [with] no more than 5 users and running as many small tests as you can afford” [11]. Constant feedback via weekly meetings and the use of Slack for messaging, as part of an iterative process, resulted in several changes over the course of the project. All the experts were able to use the system and commented on the helpfulness of the visual appraisal pages as a way of confirming results from the automatic checker. It was noted how useful it was to have automatically captured images from the web pages showing where issues were arising. These can be helpful for the web developer as well as being able to feed issues into the automatically produced accessibility statement.

---

7 https://slack.com/intl/en-gb/.
4 Conclusion

There were some limitations to this research in terms of time constraints, failure to access password protected sites in a secure manner and issues around model windows and contrast levels. Nevertheless, there are WCAG 2.1 Success Criteria that can respond to the use of AI despite the fact that there has been little use of machine learning techniques as a way of supporting web accessibility checks in the last few years. Abou-Zahra et al. commented that the “significant drawback of artificial intelligence for web accessibility at this time is a lack of accuracy and reliability” [3]. Ultimately, it is hoped that further use of machine learning, neural networks and natural language processing can be implemented with increased accuracy and reliability, especially when there are larger data sets available from accessibility checks. However, these need to be open and available in order to build useful corpora. At present models are dependent on external data sets that have the potential to skew results.

Nevertheless, by using a system that not only offers an increased number of automated checks, but also provides a way of visually appraising issues, such as orientation, overlaps, text spacing and image alt tags, a reviewer can be relatively reassured about an accurate result, especially when the tests are carried out on multiple browsers on both mobile and desktop devices. It is felt that this process has provided a means of speeding up the multiple web page checking process for any large organisation. It also enables digital accessibility experts in a team who may not be coders, to highlight issues that are arising with an increased evidence base.

References

1. W3C Web Content Accessibility Guidelines (WCAG). https://www.w3.org/WAI/standards-guidelines/wcag/. Accessed 15 Apr 2020
2. Wald, M., Angkananon, K.: Development and testing of a Thai website accessibility evaluation tool. Int. J. Electron. Commun. Eng. 10(5) (2020)
3. W3C WCAG 2.1 What’s new in WCAG 2.1. https://www.w3.org/WAI/standards-guidelines/wcag/new-in-21/. Accessed 15 Apr 2020
4. Abou-Zahra, S., Brewer, J., Cooper, M.: Artificial Intelligence (AI) for web accessibility: is conformance evaluation a way forward? In: Proceedings of the Internet of Accessible Things, pp. 1–4 (2018)
5. UK Government Legislation Public Sector Bodies (Websites and Mobile Applications) (No.2) Accessibility Regulations (2018). http://www.legislation.gov.uk/uksi/2018/852/contents/made. Accessed 15 Apr 2020
6. Norman, D.A., Draper, S.W.: User Centered System Design; New Perspectives on Human-Computer Interaction. L. Erlbaum Associates Inc., USA (1986)
7. Duran, M.: What we found when we tested tools on the world’s least-accessible webpage (2017). https://accessibility.blog.gov.uk/2017/02/24/what-we-found-when-we-tested-tools-on-the-worlds-least-accessible-webpage/. Accessed 12 June 2020
8. Sen, S.: Artificial Intelligence and Web Accessibility (Unpublished Master’s thesis) (2019). https://shaunaksen.github.io/AI-for-Web-Accessibility/. Accessed 15 Apr 2020
9. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
10. Kusner, M., Sun, Y., Kolkin, N., Weinberger, K.: From word embeddings to document distances. In: International Conference on Machine Learning, pp. 957–966, June 2015
11. Nielsen, J.: Why you only need to test with 5 users (2000). https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/. Accessed 15 Apr 2020