Review of Dissertation

titled

"*New Methods and Algorithms for Testing Web Accessibility*"

By Jinat Ara

1. Topic and timeliness of the dissertation

The dissertation focuses on understanding the scope and limitations of today's widely used web accessibility evaluation methods and aims to improve upon them across various dimensions. On the one hand, the author examines the challenges and limitations of existing approaches, i.e. both automatic and hybrid methods involving humans and machines. On the other hand, based on these investigations, the author proposes a set of improved methodologies that combine the best of multiple approaches, by considering improved requirements modeling, improved knowledge elicitation methods from both everyday users and accessibility experts, mixture-of-experts and other quantitative / machine learning approaches based on both human feedback and automated tools, as well as automated document parsing algorithms informed by both syntactic and semantic considerations. The author validates the proposed approaches on a small set of webpages across multiple domains, in general by comparing the results they provide with evaluations created by experts.

Although the Web itself dates back to the 1990s and some of the core technologies underlying it remain unchanged, significant advancements are routinely made even today that strongly influence not only the user experience, but also the set of capabilities afforded to users. Further, web content accessibility guidelines have also changed over time, and are still evolving in parallel with the changing technological landscape. **Based on this, the topic at the center of the dissertation is timely and worthy of exploration.**

Before proceeding to the evaluation of the dissertation, I would like to note that I previously served as a reviewer of its first draft. Because of this, parts of my evaluation, including observations, remarks and criticisms may refer back to the previous version of the dissertation and to key points I have made in my earlier review.

2. Structure and style of the dissertation

The dissertation consists of 109 numbered pages (not including the appendices and bibliography but including a list of the author's own publications). It is organized into 4 main

chapters, followed by a chapter with conclusions and the aforementioned list of the author's publications, labeled as a separate chapter. The bibliography at the end of the dissertation consists of 107 references, which point to relevant and, in most cases, very recent publications that have appeared in highly regarded scientific conference proceedings and journals. The content of the main chapters (chapters 1-4) can be outlined as follows:

In chapter 1, the author introduces the key motivations for this research, then outlines the research methodology that was used and lays out the research questions and hypotheses that were formulated during the early stages of the research.

In chapter 2, the author introduces the state-of-the-art behind different classes of approaches used in web accessibility evaluations, namely manual approaches, automated approaches and several different variants that belong to these categories or combine them in hybrid ways. Based on a detailed literature overview, the author distills a table with over 20 references and identifies key drawbacks of each existing class of approaches. Overall, the author notes challenges such as the high cost and inefficiency of evaluations based on human feedback, inconsistencies in applying accessibility and usability criteria, difficulties in interpreting and mapping Web Content Accessibility (WCAG) guidelines to actual web features, and a lack of clarity in how evaluation results are presented and understood. The limitations and tradeoffs identified in the chapter form the basis of Thesis group no. 1, which includes 1 thesis with 2 sub-theses.

In chapter 3, the author proposes 3 different and somewhat complementary enhancements to existing hybrid evaluation approaches with the goal of mitigating the drawbacks identified in the previous chapter. The 3 enhancements are termed by the author as i.) the integrated approach; ii.) the variable magnitude approach; and iii.) the machine learning based approach. The novelty of the integrated approach lies in dynamically integrating multiple automated tools with both expert and user feedback in a sequential two-phase process for a more comprehensive accessibility evaluation. The novelty of the variable magnitude approach (in contrast to prior solutions referred to as variable magnitude solutions) is the use of dynamically adjusted weights based on statistical thresholding and expert validations. Finally, the novelty of the machine learning based approach is to use machine learning to infer and integrate additional usability aspects that traditional rule-based methods (like those based on WCAG) cannot directly evaluate. Specifically, the approach leverages ML classifiers to capture indirect indicators – such as manual adjustment options, dynamic webpage states, CAPTCHA handling, and multimedia content usability – to compute a more comprehensive accessibility score. All 3 proposed enhancements are validated in the chapter and form the basis of Thesis group no. 2, which includes 3 theses (Theses 2-4).

Finally, **in chapter 4**, the author shifts focus to automated evaluation tools for web accessibility, proposing enhancements to address several shortcomings of current solutions – related to guideline selection, guideline modeling, rigidity of evaluation criteria and consideration of semantic aspects. In this context, the chapter proposes the Web Content Accessibility Evaluation Environment (WCAEE), which is an integrated automated framework aiming to a.) incorporate user and expert feedback b.) while also using specialized, separate algorithms for textual and non-textual analysis to generate accessibility reports. The second part of the chapter describes the implementation details behind the tool and presents examples run through its user interface. The content of this chapter forms the basis of Thesis group no. 3, which includes 2 theses (Theses 5-6).

The dissertation ends with a concluding chapter, in which a formulation of the 6 theses (organized into 3 thesis groups) is given, and a brief discussion is provided on the author's findings.

In general, the dissertation fulfills the requirements of form and style outlined within the rules and guidelines for PhD dissertations.

3. Methods used

The results of the dissertation were obtained by the author through the combined use of several different methodologies, including:

- Automated and questionnaire-based accessibility evaluation methods (based on both user and expert feedback)
- Dynamic scoring approaches with respect to the variable-magnitude methods
- Classical machine learning approaches, including decision trees and random forests
- Declarative evaluation designs and algorithmic tools for the development of an automated specification and parsing framework
- Complexity and statistical analyses for report generation

In my assessment, these methods were well chosen by the author, as they are highly suited to the challenges addressed in the dissertation. Further, the author often applied these methods by combining them in conceptually novel and useful ways.

Observations and criticisms with respect to the content of the dissertation

In my earlier review, one of my main criticisms with respect to the structure and style of the dissertation was that key concepts that were used were not clearly defined. The author remedied this by providing definitions of the various approaches in Chapter 2 (e.g., hybrid approaches, crowdsourcing approaches, heuristic approaches and various kinds of automated approaches).

However, some inconsistencies still remain in the current form of the dissertation. For example, at the bottom of page 13, it is not clear why scenario 2 (incorporation of user testing and expert testing) and scenario 5 (incorporation of different kinds of automated testing) belong to the hybrid approach, since the definition of the hybrid approach on page 10 talks about the concurrent use of automated and manual solutions, whereas scenario 2 seems to only involve manual approaches and scenario 5 only involves automated approaches.

Similarly, defining crowdsourcing approaches as necessarily involving automated tools is problematic, since the core concept of crowdsourcing typically centers on gathering input from a broad population – often non-experts – without relying on automation. However, the definition used here blends expert evaluation, user testing, and automated tools, which aligns more closely with hybrid or participatory evaluation than with crowdsourcing per se. Moreover, combining automated tools with manual evaluation is not unique to crowdsourcing, and the simultaneous inclusion of both user and expert perspectives may be misleading, as crowdsourcing often involves contributions from non-expert users alone.

In Chapter 3, on page 24, the author states that "*crowdsourcing and heuristic approaches are the two main branches of the hybrid testing or evaluation process. Crowdsourcing approaches generally allow for the integration of both automated testing (incorporating automated tools) and human evaluation (performed by hiring people, including experts and users) to evaluate the accessibility of web. Besides, heuristic approaches generally enable only human inspection in various ways to evaluate the accessibility of the web.*"

This, again, is confusing in light of the earlier definitions: specifically, it is unclear why heuristic approaches would "generally enable **only** human inspection", as this was defined as one of the main branches of hybrid testing, which – according to its definition – involves both manual testing and automated testing.

To be clear, I don't think that these difficulties with respect to the terminology influence the validity of the results achieved. In addition, it is clear from the literature that other researchers have used these terms in somewhat broad contexts, without providing clear-cut

definitions. My impression is that perhaps these categories might be better conceptualized as a spectrum of approaches which could be used in various combinations at a higher level of granularity. For example, some automated approaches might be useful when integrating potentially conflicting assessments coming from a large group of people (in the case of crowdsourcing), but this is a different kind of automation from e.g. robotic process automation or DOM processing techniques focused on automated web accessibility testing.

The author might have added to this discussion in a meaningful way by clearly defining such concepts and their boundaries; however, although I see this as a missed opportunity, it still does not detract from other scientific results achieved by the author.

4. Summary and assessment of the theses

Based on the obtained results, the author formulates 3 thesis groups containing 1, 3 and 2 theses, respectively.

In Thesis no. 1 (thesis group 1), the author highlights several key shortcomings of existing web accessibility evaluation frameworks, and formulates the need for new, more comprehensive frameworks that incorporate concerns related to guideline selection, integration of both expert and user feedback, and canonical visualization and documentation approaches.

In Thesis no. 2 (thesis group 2), the author proposes and validates the so-called integrated approach, which involves the combination of multiple kinds of automated tools with expert and user feedback.

In Thesis no. 3 (thesis group 2), the author proposes and validates the so-called variable magnitude approach, which involves a dynamically weighted summation of scores obtained from various automated evaluation methods.

In Thesis no. 4 (thesis group 2), the author proposes and validates the so-called machine learning-based approach, in which machine learning techniques like Random Forest or Decision Tree models are used to incorporate additional evaluation criteria (of which the author proposes 10) beyond standard Web Content Accessibility Guidelines (WCAG).

In Thesis no. 5 (thesis group 3), the author proposes and validates an automated accessibility evaluation framework that operates on simplified guidelines while also incorporating user feedback and expert suggestions and providing an overall accessibility score as opposed to multiple – potentially conflicting – results.

Finally, **in Thesis no. 6** (thesis group 3), the author proposes and validates the Web Content Accessibility Evaluation Environment (WCAEE), which is an automated evaluation tool that

operates based on common features of the structural and visual elements of webpages (which can be readily derived from the HTML DOM structure), while also considering both semantic and non-semantic aspects.

In my assessment, all 6 theses formulate new scientific findings and are the author's own results.

The formulation of the theses is also much improved compared to the previous draft version of the dissertation, in which I recommended making several changes in terms of wording and style.

5. Assessment of the author's publication performance

Based on the MTMT database, the author has displayed a significant and impressive publication performance.

Specifically, the author has published 15 papers in scientific journals, 13 of which have been published outside of Hungary. In addition, the author has published 15 English-language papers at scientific conferences.

In terms of impact, according to the MTMT database, the author's work has received 68 independent citations, which is a significant achievement for a PhD candidate.

6. Questions

1.) How would you reconcile the seeming contradiction between the use of the term "crowdsourcing-based approach" to refer to hybrid approaches and the common interpretation of the word "crowdsourcing" as simply referring to the involvement of large groups of users? How would you reconcile the seeming contradiction between your description of heuristic methods as ones that "generally enable **only** human inspection" and their categorization as hybrid methods? Are these contradictions simply a reflection of how these terms are used in the literature (e.g., crowdsourcing hasn't been used without some automation in the web accessibility evaluation context) or should the definitions be further refined?

2.) When faced with a brand new webpage, which specific approach or what combination of approaches would you recommend using, since you have proposed several different approaches (integrated approach, variable magnitude approach, machine learning-based approach and WCAEE)? Can these approaches be used in combination, and in what order or according to what workflow would you suggest

they be used? If not, then how would you recommend deciding which approach(es) to use?

7. Conclusions

Based on the above, I have concluded that the dissertation advances the state-of-the-art within the topic of web accessibility assessment methods. Further, based on the publications of the author, it is clear to me that the results are the author's own scientific results. Accordingly, I have accepted all 6 theses contained in the dissertation as new scientific results.

The author has published over 25 peer-reviewed papers in the literature, and her work has received 68 independent citations based on the MTMT database. This is a significant achievement for a PhD candidate.

Aside from minor concerns regarding terminology, I fully support the dissertation being submitted for public defense. Upon a successful defense, I support awarding the PhD degree to the author.

Budapest, 7 August 2025

Ádám B. Csapó

Full professor

Corvinus University of Budapest