Exploring the differences between five accessibility evaluation tools

Marian Pădure
Babes-Bolyai University
Str. Sindicatelor 7, Cluj-Napoca, Romania
marian.padure@ubbcluj.ro

Costin Pribeanu
Academy of Romanian Scientists
Splaiul Independentei 54, Bucharest, Romania
costin.pribeanu@aosr.ro

ABSTRACT

Web accessibility in the public sector is an important objective of the Digital Agenda for Europe and requires a systematic evaluation and monitoring strategy. Given the low accessibility of public websites as well as a large number of websites in the public sector, a tool-based approach is needed. The goal of this paper is to compare five accessibility checking tools as regards the coverage of accessibility-related issues, identification, and reporting of accessibility errors. The comparison is then illustrated with three case studies of municipal websites from Romania. The results show large differences as regards the number of errors which suggest that only one tool is not enough.

Keywords

accessibility checking tools, Web accessibility, WCAG 2.0, municipal websites

ACM Classification

D.2.2: Design tools and techniques. H5.2 User interfaces.

INTRODUCTION

A barrier-free Europe is an important objective of the Digital Agenda for Europe [5] and the main concern for the European Disability Strategy 2010-2020 [6]. More recently, a European directive has been issued that requires the accessibility of public websites by September 2020, the latest [6].

To accomplish this objective, two important steps should be done. First, each member state should reinforce the directive with regulations and accessibility policies at the national level. Second, an accessibility evaluation and monitoring of public websites are needed.

Since 2008, the reference for web accessibility is the Web Content Accessibility Guidelines (WCAG 2.0) document. WCAG 2.0 (WCAG2) defined three levels of conformance (A - lowest, AA, and AAA - highest) [17]. For the public websites in Europe, the AA level of conformance is required.

Given the low accessibility of public websites as well as a large number of websites in the public sector, a tool-based approach is needed. Accessibility evaluation tools have several advantages such as a fast and easy way to check accessibility, cost-effective and affordable for a large number of websites, and reliable to the extent of producing reproducible evaluation results [12].

On another hand, several accessibility evaluation tools exist and selecting the most useful one for a given evaluation target is not easy. The objective of this paper is to explore the differences between five accessibility checking tools about the main features, reporting, and facilities for developers.

Next section presents some related work in analysis and comparison of accessibility checking tools. Then, the main capabilities of the selected tools are presented. The comparison is illustrated with three case studies of Romanian municipal websites. The paper ends with the discussion, conclusion, and intention of future work.

RELATED WORK

Web accessibility guidelines

Web Accessibility Initiative (WAI) is aiming to develop strategies, guidelines, and resources to support web accessibility [14]. The first version of accessibility guidelines (WCAG 1.0) has been published in 1999 [16]. In 2008, the second version of web accessibility guidelines has been published. The accessibility model of WCAG2 has a hierarchical structure based on four accessibility principles: perceivable, operable, understandable and robust [17]. For each accessibility principle, several accessibility guidelines have been defined. For each guideline, several success criteria have been defined (lower level accessibility guidelines).

Various techniques have been defined for each success criteria. The WCAG2 techniques provide guidance for developers and evaluators on how to meet the success criteria. There are three types of user guidance: enough techniques, advisory techniques, and failures.

Accessibility evaluation tools are software programs or online services that are used to check if web content meets accessibility guidelines [17]. Evaluation tools are able to automatically check the content against WCAG2 techniques.

Some potential accessibility issues could be determined automatically by the tool while others need a manual review. The tools differ in many respects: accessibility guidelines used, techniques tested, type of tool (software program / online services), supported technologies (HTML, CSS, WAI-ARIA), errors classification and reporting, guidance to fix errors, and type of license (free/commercial)

Comparison of web accessibility checking tools

Brajnik [3] discussed the effectiveness of accessibility evaluation tools in terms of completeness, correctness, and specificity. Completeness has been defined in terms of conformance with web content accessibility guidelines. Correctness has been defined in terms of the proportion of

true problems (a small number of false positives). The specificity has been related to the number of different problems a tool could detect.

As regards the metrics proposed by Brajnik, we mention the number of false negatives (true problems not detected) and number of tested checkpoints for completeness, number of false positives for correctness, and number of different tests for each checkpoint for specificity.

A more detailed analysis of automatic web accessibility metrics has been done by Vigo & Brajnik [12]. They proposed a quality framework for accessibility metrics having three levels of fulfillment: required, desirable, and optional. In turn, these levels have been discussed within four usage scenarios: quality assurance (web engineering), benchmarking, search engines, and user-adapted interaction. The metrics that have been analyzed in each scenario were validity (related to conformance), reliability (consistency of scores), low sensitivity (robust behavior against small changes in accessibility), adequacy, and complexity.

Vigo et al. [13] analyzed the effectiveness of six frequently used accessibility evaluation tools: AChecker, SortSite, Total Validator, TAW, Deque, and AMP. The effectiveness has been analyzed in terms of coverage, completeness, and correctness about the conformance to WCAG2 guidelines. Since the analyzed tools have specific strengths and weaknesses, they suggested looking for the right combination of tools for each success criteria.

ACCESSIBILITY EVALUATION TOOLS

Previous web accessibility evaluation in Romania in 2010, 2011, and 2014 targeted municipal websites [9]. In all cases, Total Validator has been used. Since the accessibility of public websites is now mandatory, finding the most suitable tool for evaluation and monitoring is important.

Several accessibility evaluation tools exist that are featuring various testing capabilities and reporting facilities. In this work, five tools have been selected: AChecker (AC), Cynthia Says (CS), TAW, Wave, and Total Validator (TV). All are freely available, although some of them are also offering additional facilities on a commercial basis.

AChecker (AC)

Web Accessibility Checker [1] is an online tool for validating the accessibility of web pages for users with different types of disability. The interface is available in English, German and Italian and is offering the possibility of validating a direct online page, an uploaded file or just upload text directly in their editor.

The reports provided are useful for developers, by providing the possibility of checking against various the guidelines, such as BITV 1.0, US Section 508, Stanca Act, WCAG 1.0 and WCAG 2.0 (Level A, AA, AAA).

By registering, users can easily access online all recommendation and international legalization regarding the accessibility online. The validation reports, with recommendations, can be exported offline in PDF, EARL, CSV or HTML format, with full information of HTML or CSS validation, or just potential risks.

Cynthia Says (CS)

The Compliance Sheriff Cynthia Says™ [4] is not a common tool for analyzing at the accessibility of web pages. CS is also intended to be an educational tool for informing and educating the community about the need for accessibility of online contents for all categories of users with disabilities. The portal analysis and report are based on the US Section 508 and WCAG 2.0 guidelines.

The time for analysis and loading the report is longer. The report is displayed together with a version for screen readers, "called friendly version". We would have liked the reader's report to be easy to read without an alternative version.

The report provides a very detailed picture of the errors and warnings, including recommendations needed to help the developers to correct them. This is a useful tool that can be used in correlation with similar tools and contribute to making online content accessible to all users with disabilities.

TAW

TAW (Web Accessibility Test) is an online analytics tool supporting HTML, CSS and JavaScript analysis [10]. Is not just a simple tool, TAW provides a comprehensive service of consulting, certification, training, and development of accessible web content for disability users. TAW also includes a standalone application available for Windows and MacOS.

The reports are easy to read online but can be sent by email free of charge. In the first part of the report, we can read a brief analysis of the fine checks made in the web page content on three dimensions: problems, warnings and those codes that cannot be analyzed automatically. Also, the summary and details report provide an overview based on accessibilities guidelines, divided in perceivable, operable, understandable and robust. The interface is available in three languages: English, Castellano, and Portuguese.

Wave

Wave - Web Accessibility Assessment Tool is a free tool provided by Web accessibility in mind (WebAIM) [15]. This is a very useful validation and verification tool based on information from the community that contributes to international recommendation accessibility guidelines. Wave offers the ability to test a site locally through extensions that can be embedded in Firefox and Chrome, allowing developers a simple way to check the web contents before publishing.

Wave provides a color warning system: red for those errors that need urgently corrected, green for the lines that are correct but still need to be checked, and yellow for the code that needs to be analyzed.

Content evaluation is done in real time, on the left are displayed the summary results, detailed error analysis on full, WCAG 2.0 (A or AA) or USA Section 508, easy access to documentation and information on Outline (The heading structure of the web page).

On the right side, the content is loaded with the warning icons mentioned above. Compared to other instruments used, it allows an evaluation of the contrast and non-styles

contents. A very useful tool but with the offline reports are not available.

Total Validator (TV)

Is a useful tool for checking the accessibility of web pages, can run as desktop application tools under Windows, MacOS, and Linux [11]. TV checks the content against HTML and CSS validation code for W3C, WCAG (1.0, 2.0, and 2.1) and US Section 508. Compared to other validation tools validates pages which are password-protected and pages generated by JavaScript.

It allows HTML5 validation, being the first XHTML5 and WCAG 2.1 validator, and a CSS 3 analysis. Total Validator is a free tool, with a commercial package and comes in four versions of Test, Basic, Professional, and Embedded. Depending on the chosen package, it can also make a linguistic analysis, providing support for five languages. www.totalvalidator.com

CASE STUDIES

Method

The underlying objective of this study is to find the most suitable tools for large scale evaluations. The data has been collected in May 2019 and includes three municipal websites.

For each sample, the home page has been evaluated for conformance against WCAG 2.0. The results have been then analyzed and discussed by the accessibility principle and guideline/success criterion.

The number of WCAG2 accessibility errors has been taken from the report provided by each tool. HTML, CSS, link, and parsing errors have not been considered in this study.

Website of Cluj-Napoca City Hall

A comparison of the evaluation results by accessibility guideline is presented in Table 1.

Table 1. Errors by accessibility guideline

Categories	AC	CS	TAW	Wave	TV
1.1.1 Text alternatives	4	26	17	8	7
1.3.1 Info & rel	11	13	25	0	5
2.4.4 Link purpose	1	8	8	4	2
3.3.2 Labels	9	9	17	16	0
4.1 Robust	0	9	5	0	0
Other	0	0	16	0	0
WCAG2 A	25	7	74	12	15
1.4.3 Contrast	22	329	0	438	0
1.4.4 Rel units	18	529	0	0	0
2.4.6 Unique labels	1	0	0	0	2
WCAG2 AA	41	858	0	438	2
TOTAL	66	924	74	450	17

The number of WCAG2 A errors is varying from 7 to 74. s regards the AA errors, the differences are very large, from none (TAW) to 438 (Wave) and 924(CS).

Website of Timisoara City Hall

A comparison of the evaluation results by accessibility guideline is presented in Table 1.

Table 2 Errors by guideline

Categories	AC	CS	TAW	Wave	TV
1.1.1 Text alternatives	0	17	8	0	23
1.3.1 Info & rel	0	138	130	5	37
2.4.4 Link purpose	0	0	7	7	3
3.3.2 Labels	0	5	5	0	0
4.1 Robust	0	5	7	0	0
Other	0	1	1	1	13
WCAG2 A	0	10	159	8	77
1.4.3 Contrast	0	344	0	169	0
1.4.4 Rel units	0	391	0	0	23
2.4.6 Unique labels	0	0	0	0	0
WCAG2 AA	0	735	0	169	23
TOTAL	0	902	159	177	100

The number of WCAG2 A errors is varying from none (AC) to 159 (TAW). As regards the AA errors, the differences are large, from 0 (AC and TAW) to 735 (CS). What is surprising is the fact that AC didn't find any WCAG2 error which is raising questions as regards the reliability.

Website of Constanta City Hall

A comparison of the evaluation results by accessibility guideline is presented in Table 1.

Table 3. Errors by guideline

Categories	AC	CS	TAW	Wave	TV
1.1.1 Text alternatives	11	40	3	29	12
1.3.1 Info & rel	51	22	50	30	56
2.4.4 Link purpose	1	7	18	6	6
3.3.2 Labels	27	22	43	0	0
4.1 Robust	1	23	44	0	5
Other	2	3	1	1	0
WCAG2 A	93	118	231	67	79
1.4.3 Contrast	1	203	0	65	0
1.4.4 Rel units	12	304	0	0	0
2.4.6 Unique labels	4	0	0	0	24
WCAG2 AA	17	507	0	65	24
TOTAL	110	625	231	132	103

The number of WCAG2A errors is varying from 103 (TV) to 231 (TAW). AChecker, Wave, and Total Validator detected a similar number of errors.

Discussion

A summary of evaluation results is presented in Table 4 that highlights the total number of errors (all three websites) by the accessibility principle.

Table 4 Total number of errors by accessibility principle

Categories	AC	CS	TAW	Wave	TV
1 Perceivable	77	256	256	72	145
2. Operable	2	16	34	12	12
3. Understandable	38	39	70	1	9
4. Robust	1	37	56	0	5
WCAG2 A	118	351	464	87	171
1 Perceivable	53	2100	0	672	23
2. Operable	5	0	0	0	26
WCAG2 AA	58	2100	0	672	49
TOTAL	176	2451	464	759	220

As regards the way of reporting, AChecker, Cynthia Says, and TAW are structuring the reports by accessibility guideline and success criteria. Total Validator and Wave are reporting the errors under level A and AA and document each error with the guideline and success criteria that have been violated.

The results of the three case studies show that Cynthia Says and TAW are reporting the highest number of level A errors. The differences at level AA errors are much higher.

While AChecker and Total Validator are reporting a similar number of errors, Cynthia Says and Wave reported 672, respectively 2200. As regards TAW, it seems that level AA errors are mentioned as potential issues that require a manual review.

A more detailed analysis of the reports provided by each tool shows that the differences come from the different way of counting the accessibility problems.

Cynthia Says provides a total of level A and level AA errors which are actually the number of distinct types of problems, following the WCAG2 techniques. Additionally, it gives access to the source where the number of occurrences is given for each type. Same WCAG2 techniques are mentioned in the report under different success criteria, which is inflating the number of errors. This makes it difficult for the evaluator to determine the number of accessibility errors.

TAW reports the errors in a similar way in that the same results of testing a technique is occurring under several success criteria, thus suggesting that the number of errors is inflated.

Since this is an exploratory study with only three webpages tested, it is not yet possible to fully assess the reliability of each tool and to recommend a tool or a combination of tools to be used in accessibility evaluations.

Overall, the discussion of results confirms the findings of other studies as regards the differences between the results obtained by automated accessibility checker and the suggestion of using more than a single evaluation tool in order to increase the reliability [2, 8, 13].

The results of this exploratory study suggest that Cynthia Says and Wave are useful for developers given the facilities to detect and visualize the accessibility issues. AChecker and Total Validator are more comfortable for evaluators by providing compact reports.

CONCLUSION AND FUTURE WORK

To ensure the conformance level required for the public web by the EU Directive, accessibility regulations and a clear accessibility policy are needed at the national and local level. Given the huge number of public websites, the selection of appropriate tools for accessibility evaluation is an important issue.

This exploratory work contributes to the practice of web accessibility evaluation with a comparison between five frequently used evaluation tools. A preliminary conclusion is that only one tool is not enough. This study suggests that for large scale evaluations, using both AChecker and Total Validator may increase the completeness and reliability of results.

Future work will continue in two directions. First, it will extend the number of checking tools and the number of case studies in order to provide a more useful and reliable comparison. Second, it will deepen and structure the analysis in order to provide more detailed information for evaluators, as a basis for selecting the most suitable tools.

REFERENCES

1. A Checker. Available at http://www.achecker.ca
2. Akgul, Y., & Vatansever, K. (2016). Web accessibility evaluation of government websites for people with disabilities in Turkey. *Journal of Advanced Management Science* Vol, 4(3), 201-210.
3. Brajnik, G. (2004). Comparing accessibility evaluation tools: a method for tool effectiveness. *Universal Access in the Information Society*, 3(3-4), 252-263.
4. Compliance Sheriff Cynthia Says™. Available at http://cynthiasays.com
5. EU (2010). *A Digital Agenda for Europe*. European Commission, Brussels, COM (2010) final/2.
6. EU (2010) European Disability Strategy2010-2020. A Renewed Commitment to a Barrier-Free Europe. COM (2010) 636 final.
7. EU (2016) Directive of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies. Official Journal of the European Union, L327.
8. Karaim, N. A., & Inal, Y. (2019). Usability and accessibility evaluation of Libyan government websites. *Universal Access in the Information Society*, 18(1), 207-216.
9. Pribeanu, C., Gheorghe-Moisii, M., Fogarassy-Neszly, P. (2015) Accessibility of Romanian municipal websites – conformance with WCAG2. In: Dardala M, Rebedea TE (Eds.) *Proceedings of RoCHI 2015*, Bucharest, Romania, 24-25 September, 25-28.
10. TAW. Available at http://www.tawdis.net
11. Total Validator. Available at: http://www.totalvalidator.com/
12. Vigo, M., & Brajnik, G. (2011). Automatic web accessibility metrics: Where we are and where we can go. *Interacting with Computers*, 23(2), 137-155.
13. Vigo, M., Brown, J., Conway, V. (2013) Benchmarking web accessibility evaluation tools: measuring the harm of sole reliance on automated tests. *Proceedings W4A 2013*, Article 1.
14. WAI (1997) *Web Accessibility Initiative*, W3C. Available at: http://www.w3.org/WAI/
15. Wave. Available at: http://wave.webaim.org
16. WCAG1 (1999) *Web Content Accessibility Guidelines 1.0*, W3C, 1999. Available at: http://www.w3.org/TR/WCAG10/
17. WCAG2 (2008) *Web Content Accessibility Guidelines 2.0*, W3C, 2008. Available at: http://www.w3.org/TR/WCAG20/