Comparing Six Free Accessibility Evaluation Tools

Marian PADURE¹, Costin PRIBEANU²
¹ Babes Bolyai University, Cluj-Napoca, Romania
² Academy of Romanian Scientists, Bucharest, Romania
marian.padure@ubb.cluj.ro, costin.pribeanu@gmail.com

The web is continuously growing in content and functionality having an important impact on society. An objective of the Digital Agenda for Europe is to make the web technologies available for all. Recently, a European directive has been issued that requires the accessibility of public websites by September 2020. Given the low accessibility of public websites, systematic evaluation and monitoring strategies are needed. Covering the huge number of public websites with reasonable effort is only possible in a tool-based approach which makes the selection of a suitable tool a critical issue. The goal of this paper is to compare six accessibility evaluation tools as regards the coverage of accessibility-related issues, identification, reporting, and documenting of accessibility errors. The comparison is illustrated with six case studies of Romanian municipal websites. The results show large differences as regards the way of reporting and the number of errors which suggest that only one tool is not enough.

Keywords: Accessibility Checking Tools, Web Accessibility, WCAG 2.0, Municipal Websites
DOI: 10.24818/issn14531305/24.1.2020.02

1 Introduction
The web is continuously growing in scope, content, and functionality thus becoming part of everyday life. As Yesilada & Harper [25] pointed out, making the web accessible for all is important for at least two main reasons: commercial success and understanding the needs of disabled users helps to understand the needs of everyone.

One in six people in Europe has severe to mild disability that affects the possibilities of taking part in society and economy [9]. On another hand, the population aging comes with inherent disabilities thus enlarging the number of people that need an accessible web.

A barrier-free Europe is the main concern for the European Disability Strategy 2010-2020 [9]. Web accessibility is an important objective of the Digital Agenda for Europe [8] aiming to make the web technologies available for all. In order to strengthen this objective, a European directive has been issued that requires the accessibility of public websites by September 2020, the latest [10].

To accomplish this objective, two important steps should be done. First, each member state must elaborate on regulations and accessibility policies at the national level. Second, strategies for a systematic accessibility evaluation and monitoring of public websites are needed in each country.

Since 2008, the Web Content Accessibility Guidelines (WCAG 2.0) document became the reference for web accessibility. WCAG 2.0 (hereinafter WCAG2) defined three levels of conformance (A - lowest, AA, and AAA - highest) [24]. For the public websites in Europe, the AA level of conformance is required. Given the low accessibility of public websites as well as the huge number of websites in the public sector, a large-scale evaluation in 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 many websites, and reliable to the extent of producing reproducible evaluation results [6]. On another hand, since several accessibility evaluation tools exist, selecting the most useful one for a given evaluation target is a critical issue.

This paper extends a previous work [14] that explored the differences between five accessibility checking tools as regards the main features, way of reporting, and facilities for developers. A new tool that is freely available has been added. The differences have been
documented and illustrated with six case studies of Romanian municipal websites. In this paper, the focus is on web accessibility for visually impaired users.

The next section presents related work in analysis and comparison of accessibility checking tools. Then, the main capabilities of the selected tools are presented and comparatively analyzed. The comparison is discussed based on the six case studies.

2 Related Work
2.1 Web Accessibility Guidelines
The World Wide Web Consortium (W3C) launched the Web Accessibility Initiative (WAI) aiming at developing strategies, guidelines, and resources to support web accessibility [21]. WAI developed guidelines for web content (WCAG), authoring tools (ATAG), and user agents (UAAG).

The first version of web content accessibility guidelines (WCAG 1.0) has been published in 1999 [23]. In 2008, the second version of web accessibility guidelines has been published. The accessibility model of WCAG 2 has a hierarchical structure based on four accessibility principles: perceivable, operable, understandable and robust [24].

Several accessibility guidelines have been defined to help in respecting each accessibility principle. For each guideline, several success criteria have been defined as lower-level accessibility guidelines. Various techniques have been defined for each success criterion that provide guidance for developers and evaluators on how to meet the success criteria. According to the type of user guidance, three categories of techniques have been defined: enough techniques, advisory techniques, and failures.

Accessibility evaluation tools are software programs or online services that are used to check the extent to which the web content meets the requirements of the accessibility guidelines [24]. Evaluation tools can automatically check the content against various technical specifications and standards.

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), error classification and reporting, guidance to fix errors, and type of license (free/commercial).

2.2 Use of Accessibility Evaluation Tools
Automated accessibility evaluation tools have several advantages such as suitability for large-scale evaluation and cost-effectiveness (expertise and time). A tool-based evaluation may serve as a first accessibility test to detect accessibility barriers [1]. Although relying only on tools is limiting the results [20], the number of studies taking this approach is continuously increasing, given the pragmatic reasons mentioned above.

Brajnik [5] analyzed the effectiveness of accessibility evaluation tools with respect to fault identification and diagnosis. In this respect, he discussed the tool’s effectiveness in terms of completeness, correctness, and specificity. Completeness refers to the conformance with web content accessibility guidelines (small number of false negatives). Correctness refers to the proportion of true problems (small number of false positives). Specificity refers to the number of different problems a tool could detect.

Vigo et al. [20] compared 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.

Paterno & Schiavone [13] mentioned several issues in the use of automated accessibility tools: expandability (extending the set of guidelines), upgradability (upgrading the existing set), alignment with the latest technology, and limited effectiveness of the reports. The study of Silva et al [16] analyzed the tool support for mobile accessibility evaluation on three platforms: Android, iOS, and Windows.
Phone. They found several differences as regards the accessibility guidelines covered as well as an overall low level of coverage (only 12.5% of recommendations).

Alshamari [4] evaluated the accessibility of three popular e-commerce websites using five tools: AChecker, EvalAccess, Mauve, TAW, and FAE. He found that although AChecker covered most of the guidelines each tool reveals interesting accessibility issues.

### 2.3 Automated Accessibility Metrics

As regards the metrics, Brajnik [5] proposed the number of false negatives (true problems not detected) and the number of tested checkpoints for completeness, the number of false positives for correctness, and the number of different tests per checkpoint for specificity.

A more detailed review and analysis of automatic web accessibility metrics has been done and recently revised by Brajnik & Vigo [6]. They proposed a quality framework of accessibility metrics that specifies five quality attributes: validity, reliability, sensitivity, adequacy, and complexity.

Validity may be defined with respect to accessibility in use and with respect to conformance. Reliability refers to the extent to which results are consistent in different evaluation contexts (different tools). Sensitivity is related to the extent to which changes in the results are mirroring changes in the accessibility. Adequacy is related to the suitability of metrics to the evaluation goal. Complexity refers to the number of variables and the complexity of algorithms used to compute the metrics.

The relevance of these attributes depends on the evaluation scenario. The authors mentioned four scenarios of use: benchmarking, quality assurance (web engineering), search engines, and user-adapted interaction. For each scenario, three levels of fulfillment have been defined: required, desirable, and optional.

As Brajnik & Vigo [6] mentioned, several factors exist that are driving the metrics used: and increasing demand for decision making support (accessibility awareness, policies, and regulations), interest in periodical monitoring of public websites, and increased usage of automated tools in large scale evaluations.

### 3 Accessibility Evaluation Tools

In this study, the following evaluation tools have been considered: AChecker (AC), Cynthia Says (CS), Mauve (M), TAW, Total Validator (TV), and Wave. All these tools are enabling testing against WCAG2.

A comparison at a glance is presented in Table 1 that highlights the main capabilities. All these tools are freely available, although some of them are offering additional facilities on a commercial basis.

| Capabilities | Faile SC | Failed checks | Tested checks | HTML | CSS | ARIA |
|--------------|----------|---------------|---------------|------|-----|------|
| AChecker     | x        | x             | x             | x    | x   | x    |
| Cynthia Says | x        | x             | x             |      |     |      |
| Mauve        |          | x             | x             |      |     |      |
| TAW          | x        | x             |               |      |     |      |
| Total Validator | x   | x             | x             |      |     |      |
| WAVE         |          |               |               |      |     |      |

Some tools make explicit the success criteria that are failed as well as the successful and failed checkpoints. Although it may be assumed that all tools are testing the success criteria somehow, only some tools are reporting the checks failed and only a few of them the checks that have been tested. As regards the HTML, CSS, and ARIA only some tools enable an explicit validation.

### 3.1 AChecker (AC)

Web Accessibility Checker [2, 11] is an online tool available in English, German and Italian. The validation may be performed against various guidelines, such as BITV 1.0, US Section 508, Stanca Act, WCAG 1.0 and WCAG 2.0.
(Level A, AA, AAA). Additionally, is possible to enable HTML validator and CSS validator.
The interface enables the evaluator validating an online page, an uploaded file or just upload
text directly in their editor. The tool can identify three kinds of issues: known problems
(accessibility barriers), likely problems (probable barriers that require human judgment),
and potential problems (possible barriers that require manual check).
The reported could be ordered by accessibility guidelines or line number. In the former case,
the list of problems ordered by success criteria and check identifier is given.
For each issue, the reference to HTML code is given. It is also possible to show the source
and list the accessibility issues where they occur.
Registered users can easily access and manage online the guidelines and save the evaluation
results. The validation reports can be exported offline in PDF, EARL, CSV or HTML format,
with full information of HTML or CSS validation.

3.2 Cynthia Says (CS)
The Compliance Sheriff Cynthia Says™ [7] is an educational portal for educating the community about the accessibility of the online content. A commercial solution is also available. The accessibility analysis could be done against US Section 508 and WCAG2 (A, AA, and AAA) guidelines.
The time for analysis and loading the report is longer. The report is structured by compliance level, success criteria, and WCAG technique.
The report includes a description of each success criterion and each tested technique.
The report provides a detailed description of errors and warnings, including recommendations for developers on how to correct them.
The following results are reported: checkpoints failed, warnings, checkpoints passed, checkpoints not relevant for the page, and possible errors needing a visual check.
For each type of error, the link to the code is provided together with the number of occurrences. An alternative way of reporting (“screen-reader-friendly”) is also available.

3.3 Mauve
Mauve has been developed by the researchers from CNR of Pisa [13] as a free accessibility validation environment. It enables the validation against WCAG2 (both 2.0 and 2.1 versions in the English language) and Stanca Act (English and Italian languages). It also provides the possibility to validate content on various platforms, such as desktop, iPad, tablet, and phone.
Mauve reports the checkpoints successful, checkpoints failed, and warnings. Mauve accessibility percentage is computed as the ratio between the checkpoints passed and total checkpoints tested.
The errors are grouped by the WCAG2 principle. Additionally, the errors may be grouped by tags and HTML vs. CSS. The success criteria are not made explicit in the summary report. Rather, the tool provides a custom set of error types. For each error, the number of occurrences is given.

3.4 TAW
TAW (Web Accessibility Test) is an online analytics tool supporting HTML, CSS and JavaScript analysis [17]. The interface is available in three languages: English, Castellano, and Portuguese. The reports are easy to read and could be also sent by email.
TAW provides services of consulting, certification, training, and development of accessible web content. TAW also includes a standalone application available for Windows and MacOS.
A summary of the report is provided that includes the number of issues (total and by principle), the number of failed success criteria, warnings, and issues needed manual review. The accessibility issues are ordered by guidelines and success criteria. For each success criteria, two links are given in the report: one to brief information about the guideline and the other to the accessibility guideline webpage for more detailed information.

3.5 Total Validator (TV)
Total Validator [18] is a free accessibility evaluation tool that is offered in four versions:
Test, Basic, Professional, and Embedded. A desktop version is provided that can run under Windows, MacOS, and Linux. Depending on the chosen package, it can also make a linguistic analysis, providing support for five languages.

TV checks the content against WCAG1, WCAG2, and US Section 508 guidelines. Can validate pages that are password-protected, and pages generated by JavaScript. It allows HTML and CSS validation, checks for parsing errors and broken links. The reports include errors, warnings, and possible errors. Total Validator provides a page report, issue report, and a detailed report page. The summary page could be expanded to see all errors and warnings grouped onto parsing, link, HTML, WCAG2A, and WCAG2AA. For each error, the number of occurrences is given together with a link to the information on the TV validation reference: success criteria, explanation, and technique.

3.6 Wave

Wave - Web Accessibility Assessment Tool is a free tool provided by Web accessibility in mind (WebAIM) [22]. Wave enables testing a site locally through Firefox and Chrome extensions. The accessibility validation is done against WCAG2 and US section 508. Wave provides a color-coding system: red for errors that need urgently corrected, green for the lines that are correct but still need to be checked, and yellow for potential issues that need manual review. Content evaluation is very fast, and the results are given on a two-pane view. On the left side is a brief online report including a summary (errors, alerts, features, structural elements, HTML and ARIA, and contrast errors. More details are given also in a compact form, by using colored icons and links to more detailed information.

On the right side, the content is loaded with error and warning icons. Compared to other instruments used, it allows an evaluation of the contrast and non-styles contents. Although this two-pane view is very useful, an offline report is not given.

4 Evaluation

4.1 Method

The current situation in Romania shows poor website accessibility. A recent study on the accessibility of municipal websites on a large sample showed low accessibility, only one website passing the requirements of WCAG2 [15]. Therefore, in order to avoid wasting resources, a pragmatic strategy for an accessibility evaluation at the national level should start with the evaluation of the homepage of all websites using an accessibility evaluation tool [15]. Provided that clear accessibility regulations at national level enter into force, next step is the evaluation of all webpages. Only after reaching an acceptable accessibility level it is worse to use a systematic evaluation methodology, such as WCAG-EM / (Website Conformance Accessibility Evaluation Methodology) [19].

In order to analyze in more detail and to illustrate the differences between the accessibility evaluation tools, six websites have been checked. The data has been collected in August-September 2019 and includes six municipal websites. For each sample, the home page has been evaluated for conformance against WCAG2 AA. The results have been then analyzed and comparatively discussed by conformance level, accessibility principle, success criterion, checkpoint, and errors (occurrences). The number of WCAG2 accessibility errors has been taken from the report provided by each tool. In a previous study [14], we noticed that some tools are including in the report the same errors several times thus inflating the number of errors. In this study, we reviewed the reports and counted the number of possible duplicates of level A errors. HTML, CSS, link, and parsing errors have not been considered since not all tools enable this kind of validation.

4.2 Case studies

Website of Cluj-Napoca City Hall

A comparison of the evaluation results by conformance level and accessibility principle
is presented in Table 2. The number of WCAG2 level A errors is varying from 13 to 58. As regards the AA errors, the differences are very large, from none (TAW) to 416 (Wave) and 765 (Mauve). AChecker and Mauve detected three success criteria that have been violated.

Mauve reported only warnings related to the lack of alternative text for non-text content (1.1.1) while other tools reported 4 to 19 errors for this success criteria. TAW was the only tool that reported only warnings or issues needing a manual review as regards the AA errors.

|               | AC | CS | M  | TAW | TV  | Wave |
|---------------|----|----|-----|-----|-----|------|
| 1. Perceivable| 15 | 48 | 56  | 27  | 12  | 13   |
| 2. Operable   | 1  | 1  | 0   | 9   | 1   | 0    |
| 3. Understandable| 3  | 1  | 2   | 4   | 1   | 0    |
| 4. Robust     | 0  | 0  | 0   | 4   | 0   | 0    |
| Total level A errors | 19 | 50 | 58  | 43  | 14  | 13   |
| Total level AA errors | 38 | 21 | 765 | 0   | 2   | 429  |
| Checks failed | 12 | 9  | 9   | 7   | 6   | 4    |
| SC failed     | 7  | 7  | 7   | 8   | 6   | 5    |

Website of Timisoara City Hall
A comparison of the evaluation results is presented in Table 3. The number of WCAG2 level A errors is varying from 0 to 98. Most errors were related to the first accessibility principle. As regards the AA errors, the differences are large, from none (AC and TAW) to 361 (Mauve).

|               | AC | CS | M  | TAW | TV  | Wave |
|---------------|----|----|-----|-----|-----|------|
| 1. Perceivable| 0  | 30 | 46  | 90  | 52  | 6    |
| 2. Operable   | 0  | 1  | 0   | 2   | 2   | 0    |
| 3. Understandable| 0  | 2  | 2   | 2   | 1   | 1    |
| 4. Robust     | 0  | 0  | 0   | 4   | 0   | 0    |
| Total level A errors | 0  | 33 | 48  | 98  | 55  | 7    |
| Total level AA errors | 0  | 10 | 361 | 0   | 23  | 154  |
| Checks failed | 0  | 12 | 9   | 14  | 11  | 5    |
| SC failed     | 0  | 7  | 7   | 8   | 8   | 6    |

Surprisingly, AChecker didn’t detect any accessibility error which may suggest that it doesn’t work well on some websites. Wave reported only 7 errors related to three success criteria.

Website of Constanta City Hall
The comparison of the evaluation results is presented in Table 4. The number of WCAG2 A errors is varying from 45 to 87. Most errors are related to the lack of text alternatives. Except for Mauve, that reported warnings, all tools reported 14 to 43 errors related to this success criterion.
Table 4. Errors by accessibility principle

|                  | AC  | CS  | M   | TAW | TV  | Wave |
|------------------|-----|-----|-----|-----|-----|------|
| 1. Perceivable   | 38  | 84  | 68  | 52  | 59  | 17   |
| 2. Operable      | 4   | 1   | 0   | 16  | 2   | 6    |
| 3. Understandable| 2   | 2   | 2   | 2   | 0   | 32   |
| 4. Robust        | 1   | 0   | 5   | 7   | 5   | 0    |
| Total level A errors | 45 | 87  | 75  | 77  | 66  | 55   |
| Total level AA errors | 17 | 64  | 489 | 0   | 24  | 65   |
| Checks failed    | 16  | 14  | 11  | 12  | 11  | 7    |
| SC failed        | 9   | 8   | 8   | 8   | 7   | 6    |

As regards the AA errors, the differences are large, from 0 to 489. As in the previous two case studies, TAW reported only warnings for the WCAG2 AA violations.

*Website of Craiova City Hall*

A comparison of the evaluation results by conformance level and accessibility principle is presented in Table 5. The number of WCAG2 A errors is varying from 23 to 46, most of them being related to the first accessibility principle.

As regards the AA errors, the differences are varying from none to 64. TAW and Total Validator didn’t detect AA errors.

Table 5. Errors by accessibility principle

|                  | AC  | CS  | M   | TAW | TV  | Wave |
|------------------|-----|-----|-----|-----|-----|------|
| 1. Perceivable   | 24  | 34  | 45  | 14  | 16  | 21   |
| 2. Operable      | 0   | 0   | 0   | 10  | 7   | 0    |
| 3. Understandable| 1   | 1   | 1   | 2   | 0   | 2    |
| 4. Robust        | 0   | 1   | 0   | 1   | 1   | 0    |
| Total level A errors | 25 | 36  | 46  | 27  | 23  | 23   |
| Total level AA errors | 5  | 47  | 92  | 0   | 0   | 42   |
| Checks failed    | 7   | 8   | 6   | 10  | 6   | 4    |
| SC failed        | 5   | 8   | 5   | 6   | 5   | 6    |

*Website of Bacau City Hall*

A comparison of the evaluation results is presented in Table 6. The number of WCAG2 A errors is varying from 12 to 47. As regards the AA errors, the differences are large.

Table 6. Errors by accessibility principle

|                  | AC  | CS  | M   | TAW | TV  | Wave |
|------------------|-----|-----|-----|-----|-----|------|
| 1. Perceivable   | 11  | 15  | 46  | 5   | 10  | 0    |
| 2. Operable      | 0   | 1   | 0   | 11  | 5   | 11   |
| 3. Understandable| 0   | 1   | 1   | 0   | 0   | 2    |
| 4. Robust        | 1   | 0   | 0   | 7   | 1   | 0    |
| Total level A errors | 12 | 17  | 47  | 23  | 15  | 13   |
| Total level AA errors | 8  | 11  | 538 | 0   | 0   | 16   |
| Checks failed    | 7   | 5   | 8   | 6   | 7   | 6    |
| SC failed        | 5   | 5   | 6   | 6   | 5   | 5    |

*Website of Iasi City Hall*

A comparison of the evaluation results by conformance level and accessibility guideline is presented in Table 7. The number of WCAG2 A errors is varying from 53 to 146.
Table 7. Errors by accessibility principle

|                  | AC  | CS  | M   | TAW | TV  | Wave |
|------------------|-----|-----|-----|-----|-----|------|
| 1. Perceivable   | 44  | 103 | 145 | 105 | 86  | 89   |
| 2. Operable      | 7   | 0   | 0   | 30  | 10  | 3    |
| 3. Understandable| 1   | 0   | 1   | 1   | 0   | 1    |
| 4. Robust        | 1   | 0   | 0   | 0   | 10  | 0    |
| Total level A    | 53  | 103 | 146 | 136 | 106 | 93   |
| Total level AA   | 55  | 3   | 132 | 0   | 2   | 2    |
| Checks failed    | 13  | 7   | 8   | 7   | 10  | 5    |
| SC failed        | 9   | 6   | 5   | 5   | 7   | 4    |

Most errors are related to the lack of a text alternative for non-text content and the lack of a text describing the purpose of a link. As regards the AA errors, the differences are varying from none to 132.

4.3 Discussion

Summary of results

A summary of evaluation results is presented in Table 8 that highlights the total number of errors (all six websites) by the accessibility principle, the number of checks failed, and the number of success criteria failed. Most level A errors are related to the first accessibility principle. The differences between the numbers of errors reported by each tool are large, varying from 204 (Wave) to 420 (Mauve). We didn’t include ACChecker into this range since the total of 154 errors was for only five websites. Most of the level AA errors are related to the guideline 1.4 (distinguishable). There are few errors related to guideline 2.4 (navigable) since only ACChecker and TAW reported such errors.

AChecker, CynthiaSays, Mauve, TAW, and Total Validator reported a similar number of failed checks. Wave does not report the failed check, these being inferred from the distinct types of errors reported.

Table 8. Summary of results

|                  | AC  | CS  | M   | TAW | TV  | Wave |
|------------------|-----|-----|-----|-----|-----|------|
| 1. Perceivable   | 132 | 314 | 406 | 293 | 249 | 146  |
| 2. Operable      | 12  | 4   | 0   | 78  | 14  | 20   |
| 3. Understandable| 7   | 7   | 9   | 11  | 2   | 38   |
| 4. Robust        | 3   | 1   | 5   | 23  | 15  | 0    |
| Total level A    | 154 | 326 | 420 | 404 | 280 | 204  |
| Total level AA   | 123 | 156 | 2377| 0   | 51  | 708  |
| Checks failed    | 55  | 55  | 51  | 56  | 48  | 31   |
| SC failed        | 35  | 41  | 38  | 41  | 38  | 34   |

As regards the failed success criteria, CynthiaSays and TAW reported the largest number of errors.

Reporting

As regards the way of reporting, ACChecker, Cynthia Says, and TAW are structuring the reports by accessibility guidelines 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. Mauve reports the checks tested and failed and is structuring the report by accessibility principle.

The detailed analysis of the reports suggests that there are several errors that refer to the same issue. This was easier to observe in the reports of ACChecker, TAW, and Cynthia Says, which are structured by accessibility.
guidelines and success criteria since the same check (in the same place) is mentioned in several success criteria. Although these repetitions have been eliminated in this study, it is difficult to manually review the reports in a large-scale evaluation of hundreds of websites.

The usefulness of evaluation tools
The results of this exploratory study suggest that Cynthia Says and Wave are more useful for developers given the facilities to detect and visualize the accessibility issues. Mauve is the only tool that evaluates the conformance against WCAG 2.1. However, the summary report is structured on the accessibility principle and checks and does not mention the success criteria. AChecker, TAW, and Total Validator seem to be the most suitable for large-scale evaluations by providing compact still detailed reports.

As regards the metrics, the number of failed success criteria and the number of failed checks seem to be most reliable. The number of errors could be an additional metric provided that is counted by failed checks. Overall, this study confirms the findings of other studies as regards the pretty large differences between the results obtained with different accessibility evaluation tools and the suggestion of using more than a single tool in order to increase the confidence in results [6, 12, 14, 20].

Limitations
There are several limitations of this study. The first limitation is related to the reliance on automatic test [5, 20]. Although the results take into account to some extent the possible duplications of errors this is based on a rough estimation after a manual examination of the reports. Another limitation is related to the fact that some tools are also reporting warnings and possible errors that require a manual review. In this study, only errors reported as known issues were considered.

5 Conclusion
All European countries should accomplish the objectives of the EU Directive as regards the accessibility of the public web by September 2020. Given the huge number of public websites, the selection of appropriate tools for large-scale accessibility evaluation is an important issue. Existing accessibility evaluation tools are featuring a large range of capabilities. However, the differences between the way of reporting and documenting the accessibility errors are too large and seriously undermine the confidence in the evaluation results. It would be expected when checking the conformance with the same accessibility guidelines (e.g. WCAG2) similar reports should be produced. In this respect, a summary report should mention at least the number of tested, successful, and failed checks for each success criteria, the number of failed success criteria, and the number of occurrences for each failed check. The detailed report should mention the checks failed (techniques) and the number of occurrences under the same check.

References
[1] Abascal, J., Arrue, M., Valencia, S. “Tools for web accessibility evaluation”. Ch. 26 in Web Accessibility – A Foundation for Research, 2019, 479-504.
[2] AChecker. Available at http://www.achecker.ca
[3] Akgul, Y., & Vatansever, K. “Web accessibility evaluation of government websites for people with disabilities in Turkey”. Journal of Advanced Management Science, 4(3), 2016, 201-210.
[4] Alshamari, M. “Accessibility Evaluation of Arabic E-Commerce Web Sites Using Automated Tools”. Journal of Software Engineering and Applications, 9, 2016, 439-451.
[5] Brajnik, G. “Comparing accessibility evaluation tools: a method for tool effectiveness”. Universal Access in the Information Society, 3(3-4), 2004, 252-263.
[6] Brajnik, G., Vigo, M. “Automatic web accessibility metrics. Where we are and
where we went”. Ch. 27 in Web Accessibility—A Foundation for Research, 2019, 505-522.

[7] Compliance Sheriff Cynthia Says™. Available at http://cynthiasays.com

[8] EU (2010). A Digital Agenda for Europe. European Commission, Brussels, COM (2010) final/2.

[9] EU (2010) European Disability Strategy 2010-2020. A Renewed Commitment to a Barrier-Free Europe. COM (2010) 636 final.

[10] 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.

[11] Gay, G., Qi Li, C. Open, interactive, customizable, web accessibility checking. In: Proceedings of the 7th international cross-disciplinary conference on web accessibility. ACM, New York, p 2, 2010.

[12] Karaim, N. A., & Inal, Y. (2019). “Usability and accessibility evaluation of Libyan government websites”. Universal Access in the Information Society, 18(1), 207-216.

[13] Paternò, F., & Schiavone, A. G. (2015). “The role of tool support in public policies and accessibility”. interactions, 22(3), 60-63.

[14] Padure M, Pribeanu, C. “Exploring differences between five accessibility evaluation tools”. In: Moldoveanu A, Dix A (Eds.) Proceedings of RoCHI 2019, Bucharest, Romania, 17-18 October 2019, 87-90.

[15] Pribeanu, C. “Large-scale accessibility evaluation of Romanian websites”. International Journal of User-System Interaction, 12(2), 83-98.

[16] Silva, C., Eler, M., Fraser, G. "A survey on the tool support for the automatic evaluation of mobile accessibility." In Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, 286-293. ACM, 2018.

[17] TAW. Available at http://www.tawdis.net

[18] Total Validator. Available at: http://www.totalvalidator.com/

[19] Velleman, E., Abou-Zahra, S. Website Accessibility Conformance Evaluation Methodology (WCAG-EM) 1.0, W3C Working Group Note 10 July 2014. Tech. Rep.; World Wide Web Consortium (W3C); 2014. URL: http://www.w3.org/TR/WCAG-EM/.

[20] Vigo, M., Brown, J., Conway, V. “Benchmarking web accessibility evaluation tools: measuring the harm of sole reliance on automated tests”. Proceedings W4A 2013, Article 1, 2013.

[21] WAI (1997) Web Accessibility Initiative, W3C. Available at: http://www.w3.org/WAI/

[22] Wave. Available at: http://wave.webaim.org

[23] WCAG1 (1999) Web Content Accessibility Guidelines 1.0, W3C, 1999. Available at: http://www.w3.org/TR/WCAG10/

[24] WCAG2 (2008) Web Content Accessibility Guidelines 2.0, W3C, 2008. Available at: http://www.w3.org/TR/WCAG20/

[25] Yesilada, Y., Harper, S. Web Accessibility—A Foundation for Research, 2nd Edition, Springer, 2019.

Marian PĂDURE received the PhD degree in Psychology from the Babeș-Bolyai University from Cluj-Napoca in 2011. Currently he is research assistant at Special Education Department in Babes-Bolyai University, coordinator at The Center for Counseling in Assistive Technologies for Students with Visual Impairment and Responsible at the Office for students with disabilities. His research interests include assistive technologies and accessibility, CAI, e-learning, learning styles, disability inclusion. He is author/co-author of 3 books and over 100 journal and conference papers/presentations.
Costin PRIBEANU received the PhD degree in Economic Informatics from the Bucharest University of Economic Studies in 1997. Currently he is a co-Editor-in-Chief of the International Journal of User-System Interaction. His research interests include: usability and accessibility evaluation, usage and acceptance of social networking websites, e-learning, usability heuristics, and usability guidelines. He is author / co-author of 4 books, 6 edited books, 8 book chapters, and over 100 journal and conference papers.