




A large-scale web accessibility analysis considering technology adoption

Beatriz Martins¹ · Carlos Duarte¹ 

Accepted: 20 June 2023
© The Author(s) 2023

Abstract

This paper reports the results of the automated accessibility evaluation of nearly three million web pages. The analysis of the evaluations allowed us to characterize the status of web accessibility. On average, we identified 30 errors per web page, and only a very small number of pages had no accessibility barriers identified. The more frequent problems found were inadequate text contrast and lack of accessible names. Additionally, we identified the technologies present in the websites evaluated, which allowed us to relate web technologies with the accessibility level, as measured by A3, an accessibility metric. Our findings show that most categories of web technologies impact the accessibility of web pages, but that even for those categories that show a negative impact, it is possible to select technologies that improve or do not impair the accessibility of the web content.

Keywords Web accessibility · Accessibility evaluation · Web technology · Accessibility barriers

1 Introduction

“Web accessibility means that websites, tools and technologies are designed and developed so that people with disabilities can use them” [1]. By making web content accessible, we are not only respecting a human right defined in the United Nations Convention on the Rights of Persons with Disabilities,¹ we are also ensuring that our content becomes accessible to approximately 15% of the world population that experiences some form of disability [2] and the many of us that experience temporary disabilities [3, 4].

To guarantee web access, one of the processes to be carried out is the evaluation of the accessibility of the web content developed [5]. While evaluating a web page or a web site allows the web site administrator to detect accessibility issues and act in conformance, only large-scale accessibility

evaluations portray the status of the accessibility of (a subset of) the web at a given point in time [6, 7]. Large-scale accessibility evaluations allow understanding the level of conformance [8], understanding how accessible the web is for certain user groups [9], comparing different domains of activity [10, 11], comparing geographical areas [12, 13] or determining the evolution of a sector in a given period of time [14].

Besides assisting in understanding the impact of web accessibility, large-scale evaluations also afford the analysis of factors that might increase or decrease the accessibility of the content made available. One of those factors is the technology used for web content development and deployment [15].

In this paper, we investigate the current status of accessibility by evaluating the accessibility of 2,884,498 pages from

Beatriz Martins and Carlos Duarte have contributed equally to this work.

✉ Carlos Duarte
caduarte@fc.ul.pt

Beatriz Martins
bmartins@lasige.di.fc.ul.pt

¹ LASIGE, Faculdade de Ciências da Universidade de Lisboa, Campo Grande 016, 1749-016 Lisboa, Portugal

¹ <https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html>.

166,311 websites. This large sample allowed us to portray the current status of the accessibility of web content, analyzing the errors that are more frequently found. Additionally, we also identified the technologies present in the websites. This allowed us to compare pages that use technologies, from different categories, with those that do not. From these comparisons, we could understand that most technologies are related to different levels of accessibility, as measured by an accessibility metric, but also, that it is always possible to identify technologies related to pages with better accessibility, even inside those categories that usually relate to pages with lower accessibility.

In the remainder of the article, we start by reviewing past large-scale accessibility studies, before describing the methodology followed. We then describe two sets of results: the first portraying the current status of web accessibility; the second describing the relations we found between web accessibility and web technologies. In the following section, we discuss the results, before concluding.

2 Related work

2.1 Large-scale accessibility evaluations

A large-scale accessibility evaluation assesses the accessibility of hundreds, thousands or tens of thousands of web pages [16]. These evaluations are useful to draw meaningful conclusions about the state of accessibility [8] as well as to locate potential problems in order to shape and improve the accessibility of the Web [17].

Evaluating the accessibility of a web page usually translates to checking its conformance with the Web Content Accessibility Guidelines (WCAG),² either directly, or indirectly through different standards that reference the WCAG, such as the European Standard EN 301 549³ or the US Section 508.⁴ This is also noticeable in the multiple large-scale evaluation studies that have been reported in the literature, from older studies checking conformance to WCAG 1.0 [6, 12, 14] to more recent studies considering WCAG 2.0 [18–21], while large-scale evaluations using the WCAG 2.1 are still missing.

One noticeable aspect of accessibility evaluation apparent when multiple evaluation tools are used, is related to the different findings reported by different tools [22–24]. As a consequence of this situation, the W3C sponsored the creation of a community of evaluation tool developers and manual methodology experts to promote an harmonized

interpretation of the WCAG and ensure that different tools and methodologies report consistent results. The Accessibility Conformance Testing (ACT) community⁵ has so far created 91 ACT-Rules, representing common interpretations of how certain aspects of the WCAG should be assessed.

By means of large-scale evaluations, it becomes possible to conclude general aspects about a certain context. For instance, if there is the need to understand the level of accessibility between two time periods, a large-scale evaluation is executed in order to have more representative data results. One study [14] carried out an analysis in 2014, where it was possible to observe the changes and examine the evolution regarding web accessibility in China from 2009 and 2013. For this purpose, the authors studied the accessibility of websites in 2009 and in 2013. First, they choose the 100 most popular websites from 2009. Then, they classified each website into categories according to their content and removed the categories with no interest for this study. The authors only evaluated the home pages of the most popular websites, due to constraints related to time and resources. None of the websites met the basic accessibility requirements. However, the results show that in 2013 people were more conscious about web accessibility.

Besides analyzing the evolution of the accessibility or even the accessibility status of various countries [12], large-scale evaluations enable the investigation of factors that can somehow impact web accessibility. For instance, a 2016 study [15] evaluated 1669 web pages and identified the web technologies used in their development in order to understand if web technologies could influence web accessibility. After conducting the accessibility evaluation and technology identification, the authors computed three web accessibility metrics (conservative, optimistic and strict) and concluded that web technologies have a significant impact in web accessibility. Another study [25] performed a similar analysis regarding web accessibility and web technologies, where results could identify a set of technologies that may lead to more accessibility errors. The authors of this study conducted accessibility assessments over three years, starting in 2019. They concluded that the number of accessibility errors and WCAG conformance failures decreased in 2021.

Other aspects like the correlation between the number of HTML nodes (i.e., website complexity) and accessibility levels can also be inspected through a large-scale evaluation. In a 2010 study [6], it was possible to verify that as the number of HTML elements increases, the accessibility quality rate decreases, which leads to the hypothesis that a more complex website tends to have a lower accessibility quality.

Generally, large-scale accessibility studies aim to characterize different aspects regarding the accessibility of web content. These large-scale assessments can help locating

² <https://www.w3.org/WAI/standards-guidelines/wcag/>.

³ <https://eur-lex.europa.eu/eli/dir/2016/2102/oj>.

⁴ <https://www.access-board.gov/law/ra.html#section-508-federal-electronic-and-information-technology>.

⁵ <https://act-rules.github.io/pages/about>.

and evaluating potential barriers, as well as encouraging developers to improve the accessibility of the Internet [17]. In this study, we draw from a large sample to characterize the current status of web accessibility but also to try to understand how the technologies used in the development and deployment of websites correlate with their accessibility levels. To guide the study, we formulated the following research questions:

1. What are the more frequently violated WCAG success criteria?
2. Do specific categories of technology have a positive or negative impact on Web accessibility?
3. Within a technology category do all technologies have a similar impact on Web accessibility?

2.2 Web developers' perspective on accessibility

Research on the impact of web technologies on accessibility can contribute to foster the uptake of accessibility practices by web developers by increasing the knowledge pool available to them. Several studies have identified that accessibility practices are still not accounted for in many instances. Inal et al. [26] surveyed a group of Turkish developers that consider themselves trained or educated in web accessibility. However, they were mostly found to be unfamiliar with web accessibility standards and assistive technologies. Antonelli et al. [27] conducted a similar effort in Brazil. They surveyed over 400 developers and found that two-thirds do not consider accessibility in their development projects, with only half of these planning on doing it in the future. Gupta et al. [28] conducted a smaller study interviewing web developers in Mozambique. They also found that developers, in general, do not consider accessibility in their product development.

The causes for the lack of adoption of accessibility practices have been identified in several studies. Leitner et al. [29] identify lack of evaluation tools, time and resources. These are consequence of untrained staff, but also of organizational factors such as promotion of accessibility resulting from diminished media attention. Farrelly [30] found another set of factors. His study identified social and individual values, inadequate guidelines and support and monetary demands as barriers impeding the diffusion of web accessibility.

Inal et al. [31] did a similar study, though focusing on a sample of UX professionals instead of web developers. Nevertheless, the findings and the underlying causes found are similar, raising the universality of the identified issues. According to this study, UX professionals spend limited work time on accessibility issues and have limited knowledge about accessibility guidelines and standards. Their main challenges in creating accessible systems are related to time constraints, lack of training and cost.

Table 1 Number of web pages by top-level domain

Top-level domain	Number of web pages
.asia	322,208
.au	174,371
.com	166,388
.org	154,154
.pt	139,807
.gov	130,689
.info	125,233
.uk	122,543
.es	120,085
.it	120,085
.fr	113,513
.de	109,135
.us	106,432
.news	105,196
.eu	102,995
.net	102,671
.br	100,353
.edu	93,956

Interestingly, a study of accessibility practitioners by Azenkot et al. [32] uncovered how accessibility champions mitigate these issues in their organizations. Their main focus is on education and development of tools and resources to allow designers and developers throughout the organization to implement accessibility, confirming the lack of training and resources available to developers. Our study can, as aforementioned, contribute by increasing the resources available to developers, specifically, by guiding technology selection.

3 Methodology

3.1 Materials

To collect a sample fit for a large-scale accessibility evaluation, we obtained individual URLs from the CommonCrawl⁶ set. We used the crawl data from November and December 2020. We removed possibly duplicated URLs, as well as unwanted URLs, such as the ones pointing to robots.txt files or image resources. We did not limit in any way the number of pages per website. We conducted the accessibility evaluation and technology identification in the period from March 2021 to September 2021, starting from the most recently crawled pages. During this period, we were able to evaluate the accessibility of 2,884,498 web pages belonging to 166,311 websites (an average of 17 pages per website). The

⁶ <https://commoncrawl.org/>.

distribution of pages per top level domain is presented in Table 1.

Given the scale of the task to be performed, automated evaluation tools are the most cost-effective option [33]. We used QualWeb⁷ [34], an automated web accessibility evaluation engine that performs a set of tests on a web page that check conformance with ACT-Rules⁸ and WCAG 2.1 Techniques.⁹ The aim was to use an engine that was free and available as a package that could be integrated into our large-scale evaluation architecture. Additionally, the engine should evaluate conformance with ACT rules. ACT-rules test a web page against a set of community approved checks, while WCAG techniques test a web page against the tool developer's interpretation of specific WCAG techniques. To ensure that only checks that correspond to consensual interpretation of the WCAG and increase the validity of the results, we only used the outcomes of the ACT-Rules tests in this study. QualWeb was one of the few options that met these criteria at the time of the study (the other ones being Deque's aXe and Siteimprove's alpha) and the one that was easier to integrate in our testing architecture. We used version 0.6.1 of QualWeb, which tested a total of 72 ACT-Rules that check different aspects of conformance with 30 WCAG 2.1 success criteria (38% of the total success criteria).

To identify the technologies of the 166,311 websites assessed, we used two technology identification tools: Wappalyzer¹⁰ and SimilarTech.¹¹ Two different tools were used in order to increase the coverage of the identified technologies that were used in the websites. Wappalyzer was used as QualWeb includes a Wappalyzer module (which represented another advantage of QualWeb for this study). In order to choose another tool to use alongside Wappalyzer, we identified the six tools with more page visits on the SimilarWeb service and performed a coverage test. This test checked the coverage of several technology identification tools by comparing the technologies the tool could find with the actual technologies from a set of 25 web pages. The 25 pages were selected from the Hunter.io platform that lists pages that use a given technology. We picked pages from technologies listed by the WebAIM One million study [25]. SimilarTech was the tool with the best coverage in this test, capable of identifying 60.88% of the technologies used in the 25 web pages of the test set.

During the selection process, we were also able to compare the outcomes of Wappalyzer and SimilarTech. We found that SimilarTech identified, on average, three times more technologies than Wappalyzer. However, many of these

are not relevant for our purposes since they are not related to the front-end component, such as, for example, the database system supporting the website.

Since we were collecting data from three different sources (QualWeb, Wappalyzer and SimilarTech), we built a system to orchestrate the data collection. An Express server issued requests to Docker containers of three types: QualWeb with Wappalyzer, QualWeb solo, and SimilarTech. Responses were processed and stored in a PostgreSQL database for analysis. The QualWeb with Wappalyzer container was responsible for handling the first evaluation of a URL from a specific domain. Given that Wappalyzer is available as a QualWeb module, the first URL of a domain was processed by QualWeb with the Wappalyzer option, to ensure the result of the same web request was processed for both accessibility evaluation and technology. Subsequent requests to the same domain did not need to further evaluate the technology, since the answer is domain dependent, and therefore, they were routed to the QualWeb only containers. We also needed to ensure that the SimilarTech technology identification was processed at a time as close as possible to the Wappalyzer and QualWeb processing to prevent unwanted effects from changes to the website. The domain of each call to QualWeb with Wappalyzer was added to a pool of domains for SimilarTech checking. We had multiple SimilarTech containers running, which ensured the pool was quickly processed and the interval between a domain entering the pool and being processed was small. This means that, unlike Wappalyzer, SimilarTech was not integrated with QualWeb.

The outcomes of the requests to each container were processed in the Express server. QualWeb returns a JSON answer. The outcomes of the accessibility evaluation and the technology identification are provided in different properties of the JSON document. They are split before individual processing and being sent to different tables in the PostgreSQL database. SimilarTech also replies with a JSON document. We parsed the outputs of Wappalyzer and SimilarTech and discarded information not relevant for our purposes. We kept the information provided by the "technologies" property, which consists of an array with the identification of the technologies used in the URL analyzed, a categorization of the technology and a version of the technology. The categorizations were dependent on the service used (Wappalyzer or SimilarTech), and we needed to merge them as explained in the following sub-section.

3.2 Measurement

QualWeb produces an evaluation report that details, for each web page, the ACT rules that pass, fail, are inapplicable or that QualWeb can't tell (those instances where an automated tool is able to process just part of a test, but not complete it without human assistance). Each ACT rule checks

⁷ <http://qualweb.di.fc.ul.pt/evaluator/>.

⁸ <https://act-rules.github.io/rules/>.

⁹ <https://www.w3.org/TR/WCAG21/>.

¹⁰ <https://www.wappalyzer.com/>.

¹¹ <https://www.similartech.com/>.

requirements that allow detecting failures of compliance with a WCAG success criteria. Therefore, by analyzing the QualWeb reports it is possible to understand how success criteria are being failed and produce a first description of the status of web accessibility.

In addition to counting the number of tests failing, passing and identifying the success criteria being violated, we also computed an accessibility metric to facilitate the analysis of the relation between technology and accessibility. To compute the accessibility level of a web page we used an accessibility metric, more specifically the A3 aggregation function [8]. This metric is an extension of the UWEM 0.5 metric [8, 35] but improves upon its predecessor by considering the complexity of the resources and the needs of different ability groups. A3 is characterized by a limited range of scores (from 0 to 1). The higher the score, the worse the accessibility level of the evaluated web resource. Equation 1 presents the formula for computing the A3 metric, where B_{pb} is the number of actual points of failure of a checkpoint b in page p , b is the barrier (checkpoint violation), N_{pb} is the number of potential points of failure of a checkpoint b in page p , and F_b identifies the severity of a certain barrier b [8].

$$A3 = 1 - \prod_b (1 - F_b)^{\frac{B_{pb}}{N_{pb}} + \frac{B_{pb}}{B_p}} \quad (1)$$

We used the A3 metric since a study comparing eleven accessibility metrics [36] found it to be one of the metrics with the expected behavior (being capable to correctly judge the level of accessibility of the pages in the study). Additionally, from the three metrics with the expected behavior, A3 was the most discriminating one.

The two technology identification tools produce reports identifying the technologies present in a given website and categorize them with a set of predefined criteria. Some categories are named the same in both tools, while others are named differently. We reviewed the identified categories and merged the ones that had different names representing the same category. For our analysis, we measured both the number of instances of each technology, and the number of instances of the different categories present in a web page. Given the volume of different technologies found (near 3500 technologies), we did not review the assignment of technologies to categories and opted to use the outcome of the technology identifications tools.

3.3 Data analysis

In order to characterize the current accessibility state, we computed descriptive statistics considering the number and the type of errors and respective success criteria. In this

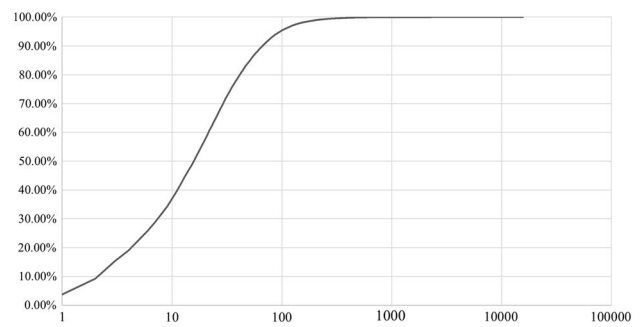


Fig. 1 Percentage of web pages and respective maximum of errors (logarithmic scale)

analysis, we also considered the top-level domain (TLD) of the web pages evaluated.

To inspect the relation between web technologies and accessibility, we carried out two analyses. The first studies the relation between web technologies and accessibility at the technology category level. Given that we found a total of 166 categories, we conducted a selection process to define the main focus of this study. We selected a total of 32 categories that belong to general areas that are used and associated with web development, such as programming languages, libraries, frameworks and software. For each category, we compared the accessibility levels, as measured by the A3 metric, of web pages that use technologies from that category with the accessibility level of web pages that do not. Since the A3 scores of our samples are not normally distributed, we applied the Mann–Whitney U rank test for the comparison. As we considered 32 different categories, we applied 32 tests. Given the large number of hypotheses being tested, we corrected for type I errors by establishing the significance level at 0.0016 (0.05/32 categories).

The second study analyzes the relation between specific technologies and the accessibility level of pages that use them. In this study, we considered only technologies belonging to categories that were identified in more than 1 million pages and had a statistically significant difference found in the prior study. From this set of categories, we selected all the technologies that were identified in no less than 2% of the web pages of their category. We applied this criterion, since some categories have a large amount of technologies. Because the Advertising category had 43 technologies that were identified in more than 2% of the web pages, for this category we studied only categories that were identified in at least 8% of the pages belonging to the category. Applying this criteria, we ended up with 6 categories. To understand if there are differences between the technologies in a specific category we used the Kruskal–Wallis test. Post hoc tests, in particular Dunn’s tests, were applied to identify significant differences between technologies of the same category. The

Fig. 2 Proportion of accessibility errors by page by each analyzed top-level domain

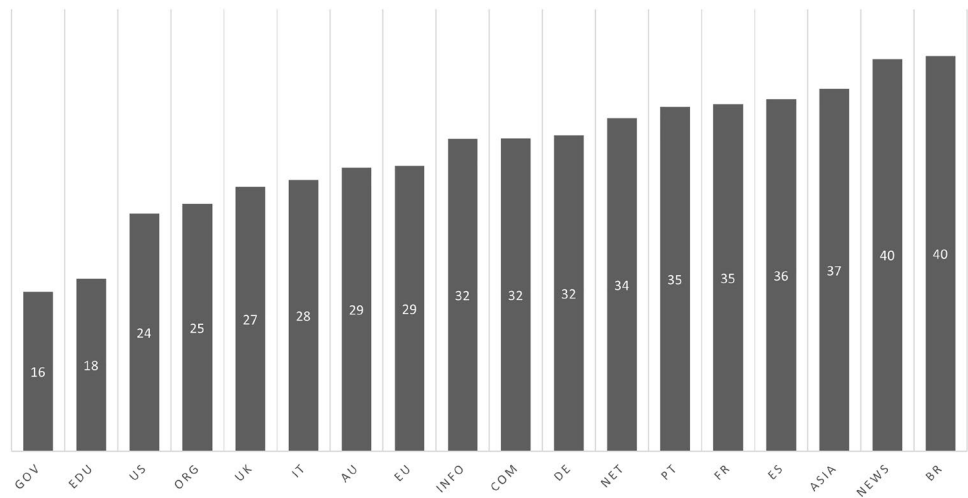


Table 2 Violated ACT rules and their respective percentage of pages

ACT-rule	Description	Success criteria	% of pages
09o5cg	Text has enhanced contrast	1.4.6 Contrast (Enhanced)	79
afw4f7	Text has minimum contrast	1.4.3 Contrast (Minimum)	66
c487ae	Link has non-empty accessible name	4.1.2 Name, role, value 2.4.4 Link purpose (In Context) 2.4.9 Link purpose (Link Only)	52
3ea0c8	id attribute value is unique	4.1.1 Parsing	31
23a2a8	Image has non-empty accessible name	1.1.1 Non-text content	30
e086e5	Form field has non-empty accessible name	4.1.2 Name, role, value	22
b4f0c3	meta viewport allows for zoom	1.4.4 Resize text 1.4.10 Reflow	22
cae760	iframe element has non-empty accessible name	4.1.2 Name, role, value	19
b5c3f8	HTML page has lang attribute	3.1.1 Language of page	18

p values for Dunn’s test were adjusted by applying the Bonferroni correction.¹²

4 Results

4.1 Web accessibility descriptive analysis

The evaluation found a total of 86,644,426 errors, averaging 30 errors per page and 521 errors per website. The highest number of errors on a single webpage was 15,645, and the lowest was 0. The highest number of errors on a website was 878,776 and the lowest 0. Only 15,963 pages have 0 errors, which corresponds to less than 1% of the total number of pages.

Table 3 Violated success criteria and their respective percentage of pages

Success criteria	Description	% of pages
1.4.6	Contrast (Enhanced)	79
4.1.2	Name, role, value	68
1.4.3	Contrast (Minimum)	66
2.4.4	Link purpose (In Context)	52
2.4.9	Link purpose (Link Only)	52
1.1.1	Non-text content	33
4.1.1	Parsing	31
1.3.1	Info and relationships	24
1.4.10	Reflow	22
1.4.4	Resize text	22
3.1.1	Language of page	19

¹² <https://onlinelibrary.wiley.com/doi/10.1111/opo.12131>.

The total number of detected accessibility issues is represented in Fig. 1 in the x-axis using a logarithmic scale. The y-axis represents the percentage of pages that have at most that number of errors. According to this illustration, approximately 37% of the web pages have between 0 and 10 errors. This means that more than the majority of the web pages (63%) have more than 10 errors.

Regarding errors by top-level domain (TLD), Fig. 2 lists 18 different TLD, ordered by their average number of errors per page. The TLD presented in Fig. 2 are the ones that have at least 90,000 pages evaluated by this study. The best results are achieved by .gov and .edu pages with 16 and 18 errors per page, respectively. The worst results are found in .news and .br pages with 40 errors per page on average.

Table 2 presents the ACT rules that failed in more than 10% of the pages evaluated, whereas Table 3 represents the violated success criteria. In the following sections, we analyze the most common issues found.

4.1.1 Text contrast

By analyzing both tables, we reach the conclusion that the majority of the pages have accessibility issues regarding the contrast of the text. If the contrast is not sufficient, users with visual impairments may have difficulties reading and distinguishing information. Interestingly, in other studies [25, 37] this aspect is also one of the most common issues affecting the legibility of the content.

4.1.2 Name, role, value

This criterion refers to the need for user interface components to have names, roles and values that can be read and set by user agents and assistive technologies. If this is not followed, users of assistive technologies will not be able to perceive and operate the interface. User interface components include standard HTML controls like links or form controls, but also custom controls that developers might create. The majority of the pages (68%) violated this criterion at least once, implying that users of assistive technologies frequently face difficulties in understanding and interacting with most web pages.

Links

The third aspect that was more frequently violated (in 52% of the web pages) concerns links and their accessible name. This problem typically arises when images are used as the single content of the link, and no accessible name is provided for the image and, therefore, for the link. We found that this situation happens at least once in approximately half of the pages evaluated. Compared to [25], our analysis

reported more accessibility problems related to links without accessible names.

4.1.3 Non-text content

All non-text content (e.g., images, videos) must be presented in an alternative format that is readable by assistive technology; otherwise, users that cannot perceive the medium the content is presented in, will not be able to perceive the content. In this study, about 33% of the pages missed the alternative description for non-text content present in the page. Images, being the most common of the media content presented on web pages, are responsible for these violations in approximately 30% of the pages evaluated. As reported in [25], a similar, even if lower, percentage of pages (26%) with missing alternative text was detected.

4.1.4 Parsing

This criterion guarantees that user agents can parse the web content, as it is properly defined and described to produce a logical data structure. To meet this requirement, elements must (1) have complete start and end tags, (2) be nested according to their specifications, (3) not contain duplicate attributes and (4) contain unique IDs. Not meeting this criterion can cause problems for user agents and assistive technologies to correctly process the content, resulting in the inability to present the content correctly if at all. We found that in approximately 31% of the pages the same id is repeated at least once.

4.1.5 Info and relationships

Whenever the relationship between pieces of information is properly specified, user agents are able to correctly provide the structure of the information irrespectively of the presentation format. This is particularly useful for screen reader users that cannot perceive the structure of the information that is provided by visual cues (e.g., visual proximity between two elements). We identified violations of this criterion in about 24% of the pages. This corresponds mainly to improper use of ARIA roles and properties and to missing relationships in table elements.

4.1.6 Reflow and resize text

The intent of these criterion is to ensure that users that need to zoom in to be able to perceive and operate web content can do so, without loss of information or functionality. When a design is responsive and correctly adapts to the zoom level, users with low vision can easily access and obtain the same information as other users. According to [37], 24% of desktop homepages do not allow the user to zoom and scale. In

Table 4 Mann–Whitney tests to analyze the impact of the web categories in the web accessibility

Category	ρ value	Average A3 in non-category pages	Average A3 in category pages	Number of Pages with category
Accessibility	<0.001	0.666	0.526	8,505
Advertising	<0.001	0.640	0.700	1,240,981
CMS	<0.001	0.711	0.627	1,776,338
Comment system	<0.001	0.664	0.731	120,676
Editors	<0.001	0.664	0.773	31,723
JavaScript frameworks	<0.001	0.666	0.663	2,324,247
JavaScript graphics	<0.001	0.666	0.645	80,260
JavaScript libraries	<0.001	0.686	0.661	2,626,325
LiveChat	<0.001	0.664	0.720	146,116
LMS	<0.001	0.666	0.366	4,130
Maps	<0.001	0.664	0.701	163,485
Message boards	<0.001	0.665	0.764	22,734
Mobile frameworks	0.005	0.666	0.662	34,159
PaaS	<0.001	0.670	0.622	1,449,356
Page builders	<0.001	0.668	0.627	162,592
Photo galleries	0.052	0.666	0.668	26,176
Programming languages	<0.001	0.688	0.655	1,928,355
Rich text editors	<0.001	0.668	0.473	38,861
Security	<0.001	0.662	0.681	540,850
Social logins	<0.001	0.655	0.742	338,538
Static site generator	<0.001	0.666	0.559	3,076
UI frameworks	<0.001	0.672	0.655	1,072,937
Video players	<0.001	0.661	0.695	376,306
Wikis	<0.001	0.666	0.371	6,662
Web frameworks	<0.001	0.666	0.660	261,150
Audio video media	<0.001	0.663	0.672	910,612
Captcha	<0.001	0.663	0.682	436,015
Forum software	<0.001	0.665	0.873	2,323
Multilingual	<0.001	0.669	0.632	255,044
Online forms	<0.001	0.666	0.662	122,336
Online video platform	<0.001	0.664	0.672	746,105
Website builder	0.839	0.666	0.684	2,254

our study, we found that 22% of the pages disabled the zoom option by using the user-scalable property with “no” value or by specifying the maximum-scale property with a value smaller than 2.

4.1.7 Language of page

To ensure the web content is correctly transmitted, the language of the page must be identified. To achieve this, the lang attribute must be defined. This allows, for instance, screen readers to ensure the correct pronunciation rules. About 18% of our set of pages failed at specifying the lang attribute, which means that 82% of the pages positively verified this requirement. Similarly, in [37], around 81% of desktop websites present a valid lang attribute. The [25] study

also reported a significant percentage of pages that specified the document language (72%).

4.2 Relation between accessibility and categories of technology

We could identify 3482 different technologies from 166 categories in the 2,884,498 web pages. From the 3482 technologies, Wappalyzer detected 1197 technologies, whereas SimilarTech detected 2733 technologies. Each category contains, on average, 21 technologies. After the accessibility evaluations, the A3 metric was computed over the generated reports for all pages. The average A3 score for all pages was 0.6657.

To assess the relation between the categories of technology and accessibility of web pages, we compared the A3

scores of web pages that use any technology of a given category and web pages that use no technology of that category. The results of the Mann–Whitney tests are presented in Table 4. They indicate that all categories impact the accessibility of the evaluated web pages, except for Mobile Frameworks, Photo Galleries and Website Builder categories.

By inspecting the values in Table 4, we can identify the categories of technology that have pages with better or worse A3 scores. Categories that are related to improved accessibility, as measured by the A3 score, are: Accessibility, CMS, JavaScript Frameworks, JavaScript Graphics, JavaScript Libraries, LMS, PaaS, Page Builders, Programming Languages, Rich Text Editors, Static Site Generator, UI Frameworks, Wikis, Web Frameworks, Multilingual and Online Forms.

Categories that are related to decreased accessibility, as measured by the A3 score, are: Advertising, Comment System, Editors, LiveChat, Maps, Message Boards, Security, Social Logins, Video Players, Audio Video Media, Captcha, Forum Software and Online Video Platform.

While the use of technologies of a specific category is related to improved or decreased accessibility, as measured by the A3 score, this does not mean that any technology from that category has the same positive or negative relation. In the following section, we report the analysis of specific technologies belonging to some of these categories.

4.3 Relation between accessibility and specific technologies

For categories of technology that are present in more than one million pages, we examined the differences between the representative technologies of the category (i.e., technologies that are present in more than 2% of the pages of the category). Table 5 presents the results of this analysis. For all categories considered, the Kruskal–Wallis tests show there is a significant difference between the means of A3 scores of the technologies belonging to the category.

In order to analyze what technologies have a statistically significant difference within the same category, we performed pairwise comparisons through Dunn's tests [38]. Figures 3, 4, 5, 6, 7 and 8 present boxplots of the A3 metric scores for all web technologies of the above categories. For each category, we only report the differences between technologies that are statistically different.

The Advertising category encompasses services that serve, text, images, video or interactive media advertisements. In the Advertising category (Fig. 3), the A3 score for DoubleClick ($\mu = 0.652$) is significantly lower (better accessibility) than the A3 scores of AppNexus ($\mu = 0.718$), Google AdWords Advertiser ($\mu = 0.752$), Google AdSense ($\mu = 0.717$) and Twitter Ads ($\mu = 0.669$). The A3 score

of Google AdWords Advertiser ($\mu = 0.752$) is significantly higher (lower accessibility) than the A3 scores of AppNexus ($\mu = 0.718$), DoubleClick ($\mu = 0.652$), Google AdSense ($\mu = 0.717$) and Twitter Ads ($\mu = 0.669$).

The Content Management Systems (CMS) category includes traditional content management systems-software platforms that support the creation and modification of content-as well as website builders-tools allowing the creation of websites without manual coding. Regarding the Content Management Systems category (Fig. 4), the A3 metric score for Joomla ($\mu = 0.566$) is significantly lower than the A3 scores of the remaining CMS. The A3 metric score for Drupal ($\mu = 0.586$) is significantly lower than the A3 scores of Jimdo ($\mu = 0.855$), TYPO3 ($\mu = 0.633$) and WordPress ($\mu = 0.624$). Jimdo's A3 score ($\mu = 0.855$) is significantly higher when compared with Drupal ($\mu = 0.586$), Joomla ($\mu = 0.566$), TYPO3 ($\mu = 0.633$), Wix ($\mu = 0.855$) and WordPress ($\mu = 0.624$).

The JavaScript Framework category includes software frameworks that are designed to support the development of web applications. Additionally, this category considers template systems and web modules. For the JavaScript Frameworks category (Fig. 5), Dunn's tests found that all the pairs of technologies are significantly different, with the exception of the AMP and Stimulus pair. In this category, we can highlight the A3 metric score for MooTools ($\mu = 0.583$) being significantly lower than the A3 metric scores for all the other technologies. On the other hand, the A3 metric score for Mustache JS ($\mu = 0.809$) is significantly higher than all the other.

The JavaScript Libraries category comprises software libraries that, when included in web pages or web applications, facilitate the development of dynamic interfaces. With regard to the JavaScript Libraries category (Fig. 6), Dunn's tests found that the A3 metric score for Isotope ($\mu = 0.328$) is significantly lower than the A3 metric scores for Hammer.js ($\mu = 0.645$), jQuery UI ($\mu = 0.628$), jQuery ($\mu = 0.472$), LightBox ($\mu = 0.689$), Lodash ($\mu = 0.743$), Moment.js ($\mu = 0.738$), Polyfill ($\mu = 0.769$), prettyPhoto ($\mu = 0.708$) and animate.css ($\mu = 0.669$). jQuery Migrate is not statistically different from any of the remaining web technologies, except for Isotope and Slick. Polyfill ($\mu = 0.769$) has an A3 metric score significantly higher than the A3 metric scores of Hammer.js ($\mu = 0.645$), Isotope ($\mu = 0.328$), jQuery UI ($\mu = 0.628$), jQuery ($\mu = 0.472$), LightBox ($\mu = 0.689$), Lodash ($\mu = 0.743$), Modernizr ($\mu = 0.646$), prettyPhoto ($\mu = 0.708$) and animate.css ($\mu = 0.669$).

The Programming Languages encompasses general programming languages as well as languages geared to web development. Notably, it also includes Node.js that is not a programming language but a runtime environment for JavaScript. In the Programming Languages category (Fig. 7), the A3 metric score for NodeJS ($\mu = 0.547$) is significantly lower than the A3 metric scores for the remaining programming languages.

Table 5 Results of the Kruskal–Wallis tests to analyze the impact of the web technologies in their categories

Categories	Technologies (Pages with technology)	<i>p</i> -value	Average A3 score	
Advertising	AppNexus (102,685)	<0.001	0.718	
	DoubleClick (830,058)		0.652	
	Facebook Advertiser (310,209)		0.755	
	Google AdSense (443,042)		0.717	
	Google AdWords Advertiser (173,995)		0.752	
	Twitter Ads (154,270)		0.669	
Content management systems	Drupal (126,866)	<0.001	0.586	
	Elementor (61,427)		0.789	
	Jimdo (28,341)		0.855	
	Joomla (80,764)		0.566	
	TYPO3 (37,902)		0.633	
	Wix (77,181)		0.610	
	WordPress (1,302,963)		0.624	
JavaScript frameworks	AMP (26,473)	<0.001	0.616	
	Angular JS (21,709)		0.730	
	Backbone.js (38,627)		0.703	
	GSAP (109,116)		0.648	
	Handlebars (36,006)		0.719	
	MooTools (48,986)		0.583	
	Mustache JS (51,118)		0.809	
	Prototype (31,768)		0.725	
	React (167,677)		0.613	
	Require JS (36,818)		0.679	
	Stimulus (16,394)		0.635	
	Vue JS (55,722)		0.636	
JavaScript libraries	Hammer.js (71,599)	<0.001	0.645	
	Isotope (208,397)		0.328	
	jQuery Migrate (1,031,833)		0.766	
	jQuery UI (635,742)		0.628	
	jQuery (2,265,370)		0.472	
	LightBox (162,919)		0.689	
	Lodash (424,009)		0.743	
	Modernizr (555,621)		0.646	
	Moment.js (103,047)		0.738	
	Polyfill (88,426)		0.769	
	prettyPhoto (124,110)		0.708	
	Slick (154,677)		0.378	
	animate.css (168,733)		0.669	
	Programming languages		Java (67,395)	<0.001
Lua (28,849)		0.763		
NodeJS (25,117)		0.547		
PHP (1,780,370)		0.650		
Python (43,765)		0.816		
Ruby (26,616)		0.630		
UI frameworks	Bootstrap (953,689)	<0.001	0.654	
	ZURB Foundation (58,886)		0.641	

On the other hand, the A3 metric score for Python ($\mu = 0.816$) is significantly higher than the others.

The UI frameworks category includes front-end frameworks and CSS focused libraries. In the UI Frameworks category (Fig. 8), all technologies are statistically significantly

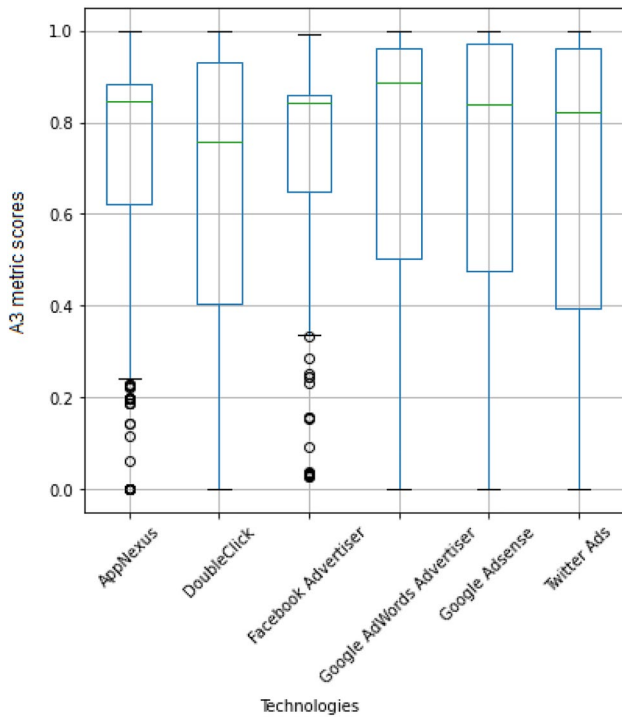


Fig. 3 Boxplot of the A3 metric scores for each technology of the Advertising category

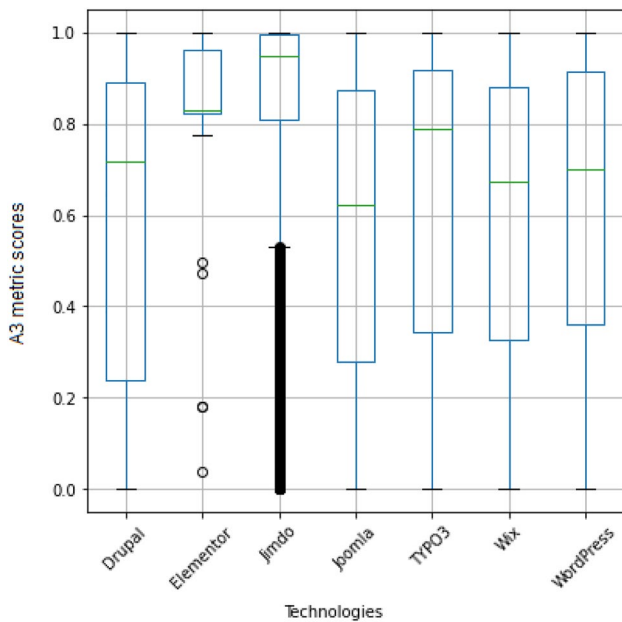


Fig. 4 Boxplot of the A3 metric scores for each technology of the Content Management Systems category

different from each other. The A3 metric score for ZURB Foundation ($\mu = 0.641$) is significantly lower than the A3 metric scores for the other technology of this category.

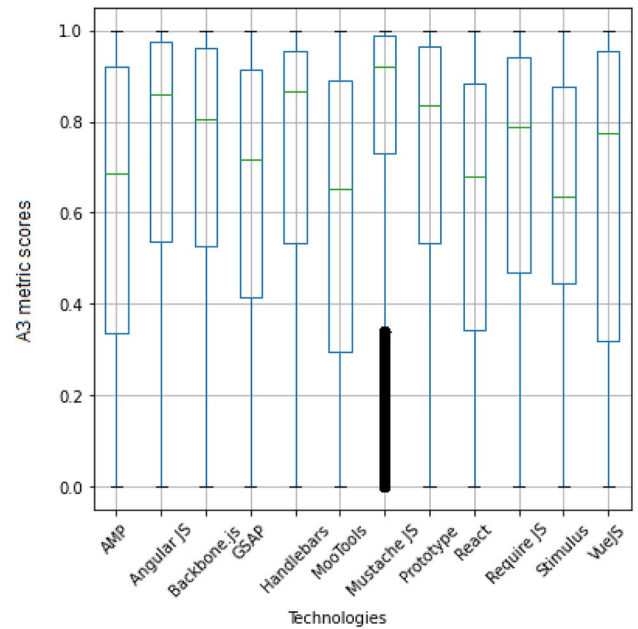


Fig. 5 Boxplot of the A3 metric scores for each technology of the JavaScript Frameworks category

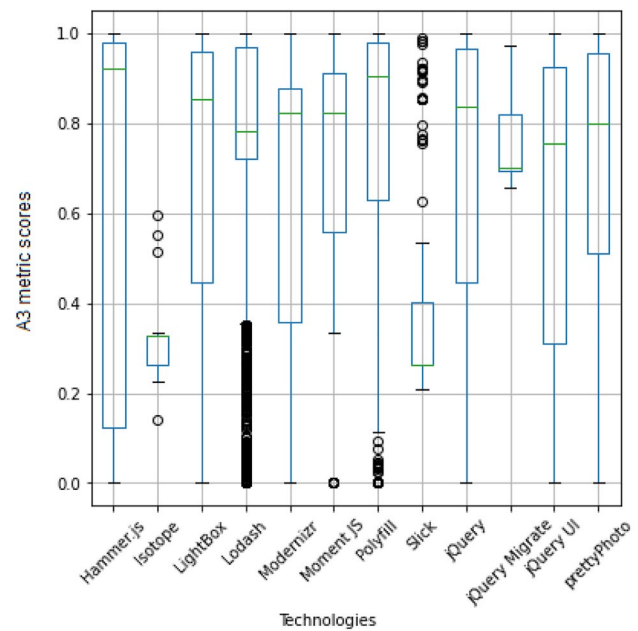


Fig. 6 Boxplot of the A3 metric scores for each technology of the JavaScript Libraries category

5 Discussion

Our analysis of the state of the accessibility of the Web revealed what can be classified as a mediocre outlook for

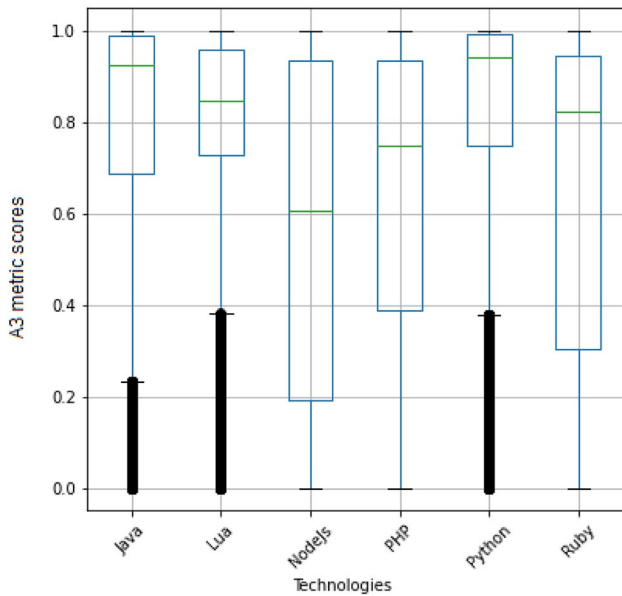


Fig. 7 Boxplot of the A3 metric scores for each technology of the Programming Languages category

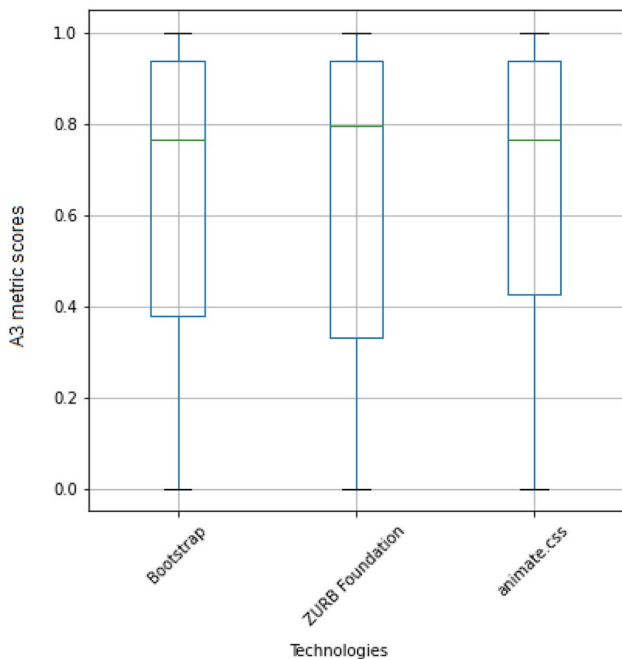


Fig. 8 Boxplot of the A3 metric scores for each technology of the UI Frameworks category

the year 2021. Assessed web sites had an average of 30 errors per web page. Only 15,963 pages out of 2,884,498 had no errors detected by our automated tool, a mere 0.5% of our sample. In almost two-thirds (63%) of the web pages, we found at least 10 accessibility errors.

Still, this is a smaller number when compared with the 51.4 errors per page reported in WebAIM’s analysis of one million pages for 2021 [25]. The major factor to explain this difference is the different tools used in each study. In our study, we used a tool that is able to test ACT-rules, and we used only the results of those checks. WAVE¹³ (the tool used for the WebAIM’s study) has not (to our knowledge) implemented ACT rules and includes more tests that have not been validated by the accessibility testing community. Therefore, it is reasonable to expect that WAVE will report more errors than QualWeb in this configuration. However, both tools report a similar proportion of errors of the same type, and a similar proportion of pages violating specific success criteria.

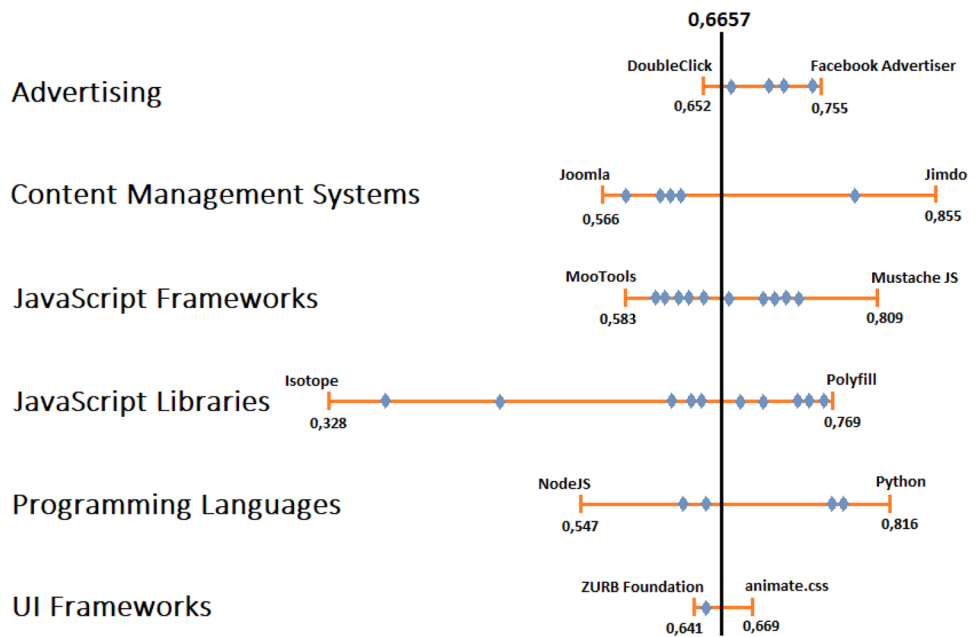
Problems related to insufficient text contrast are the most frequent, followed by problems regarding the absence of accessible names. These two aspects complicate the interaction between users and the web page, insofar as users will not be able to perceive the page’s content. Contrast problems impact users with low vision that do not use contrast-enhancing technology. This description can fit any of us given the right conditions (for example, browsing a web page on a mobile device outdoors with bright sunlight). Blind users, who are not impacted by the contrast issues, are impacted by the lack of accessible names, which prevent them from perceiving and understanding web content when using a screen reader to browse.

Our analysis also shows that the level of accessibility of a web page, as measured by an accessibility metric, and the web technology it uses are related for most of the technology categories examined. From the total of 29 categories for which we found that pages using technologies from that category have significantly different A3 scores than pages that do not use those technologies, 16 categories lead to significantly lower A3 scores (better accessibility), while 13 categories lead to significantly higher A3 scores (lower accessibility). LMS and Wikis were the categories with the lower A3 scores. Both LMS and Wikis represent technologies that structure and constrain the ways that web content can be presented. Of course, this can lead to better or worse accessibility, depending on how accessible the “templates” used are. What our analysis shows is that most LMS and Wiki based sites seem to be using “templates” that have taken accessibility into account, given the good A3 scores their pages achieve.

Another interesting observation that can be made is that development frameworks (JavaScript, UI, Web) and systems (CMS, LMS, Static Site Generators, Page Builders) are related to improved A3 scores, implying that the use of technologies from these categories prevents some of the accessibility issues that automated tools can detect. On the

¹³ <https://wave.webaim.org/>.

Fig. 9 Distribution of the A3 metric scores for each technology in those categories that were identified in more than 1 million pages



other hand, we can observe that categories of technology representing components that are “plugged-in” to the website (Advertising, Captcha, Maps, LiveChat, Social Logins) are related to decreased A3 scores, implying that these components are a cause of accessibility issues. Unfortunately, the same can be said about categories of technology related to media content (Video Players, Audio Video Media, Online Video Platform).

We followed our analysis by inspecting specific technologies within the most used categories. This allowed us to pinpoint the technologies that are present in the web pages with the best and worst A3 scores, for each of those categories. Figure 9 illustrates the distribution of technologies, for each category, around the average A3 score of all the evaluated pages ($\mu = 0.6657$). In the figure, we identify the technologies with the best and worst scores and represent the other technologies with the relative position in the range from the best to the worst technology.

In Fig. 9, we can see that three of the categories present a medium range of A3 values (from 0.226 to 0.289), while one category (JavaScript Libraries) has a large range of A3 values (0.441) and two categories have small ranges (0.103 and 0.028). The largest and smallest ranges are categories that have the higher and lower number of technologies, but it should be highlighted that that is not the only reason, since there are categories in the medium range with the same number of technologies. What this seems to indicate is that the choice of JavaScript Library can have a higher impact on the accessibility of the web page than the UI Framework or the Advertising technology to use.

It is also interesting to observe that none of these six categories have only technologies with an A3 score that is

above or below the average of the A3 score of all the pages evaluated in our study. What this shows is that, for these categories, it is always possible to select a technology that can improve (or worsen) the accessibility of web content, even if that category has shown to worsen (or improve) the accessibility of web pages in general.

By comparing our results with the WebAIM Million study [25], we noticed similar ordering of the technologies. For instance, our UI Frameworks category has the ZURB Foundation ($\mu = 0.641$) technology as the technology with lowest A3 score, followed by Bootstrap ($\mu = 0.654$). The range of these metrics is the smallest one in our analysis. The same happens in the WebAIM study, whose results have the technologies ordered in the same way, as well as a similar average for the number of errors reported for each technology.

6 Limitations of the study

In this work, we used QualWeb, Wappalyzer and SimilarTech, either to evaluate the accessibility of web pages or to identify web technologies. The fact that automatic tools were the only used evaluation procedure in this study limits the amount of detected accessibility issues, among other drawbacks [39]. Given that by using only tests from ACT rules we are adopting an approach that reduces the number of false positives (i.e., flagging something as an error when it is not) at the cost of eventually increased false negatives (i.e., failing to flag errors), we are aware that we are not capturing all accessibility problems,

therefore offering a more positive perspective about the overall web accessibility than the reality.

We used the categorization of technologies from the tools themselves. We are aware that some technologies might be categorized differently and that would mean a specific technology would be compared with other technologies when in another category. However, due to the large size our sample, we do not expect that would impact the results of the comparison between categories.

7 Conclusions

Following the accessibility assessment of 2,884,498 web pages, we believe the state of web accessibility continues to lag behind what needs to be achieved to ensure universal access to web services and content. We found an average of 30 errors per page and that only a very small number of web pages (0.5% of our sample) did not have errors. Taking into account that the accessibility assessment was done with an automated tool, which cannot detect all accessibility barriers, this grim outlook is still an optimistic view of the real status, that, at best, is as bad as portrayed, and, probably, even worse.

We complemented the accessibility analysis, with a study of the relationship between web accessibility, as measured by an accessibility metric, and web technologies. Almost all of the categories of technology we analyzed were found to lead to differences in the value of the metric when comparing pages that used technologies from the category to those pages without those technologies. Our inspection found that using development frameworks and systems seems to lead to pages with improved accessibility, while technologies that represent components that are just plugged-in to the web page seem to lower the accessibility. Still, by further analyzing the differences between technologies within a category, we learned that, irrespective of category, technologies can be selected that can lead to improved or worsened accessibility.

According to this study's findings, developers should take into consideration two important aspects: (1) the prioritization and awareness of the importance of access by all groups of users and (2) the impact that the technologies have on accessibility, upon the development of web content. Hence, design and development teams should focus on applying good practices when designing and developing web content and choosing the technologies that lead to better accessibility.

Acknowledgements This work was supported by FCT through the LASIGE Research Unit, ref. UIDB/00408/2020 and ref. UIDP/00408/2020

Funding Open access funding provided by FCTIFCCN (b-on). The research leading to these results received funding from FCT through the LASIGE Research Unit under Grant Agreements ref. UIDB/00408/2020 and ref. UIDP/00408/2020.

Data availability The datasets generated during and/or analyzed during the current study are available from <https://doi.org/10.5281/zenodo.7494722>.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. W3C: Introduction to Web Accessibility. (2005). Accessed in 18 of January of 2022. <https://www.w3.org/WAI/fundamentals/accessibility-intro/>
2. World Health Organization and World Bank: World report on disability 2011. World Health Organization (2011)
3. Henry, S.L.: Understanding web accessibility. In: Web Accessibility: Web Standards and Regulatory Compliance, pp. 1–51. Apress, Berkeley (2006). https://doi.org/10.1007/978-1-4302-0188-5_1
4. Macík, M., Sporka, A.J., Slavík, P.: An initial study of effects of temporary disabilities. SIGACCESS Access. Comput. **103**, 3–13 (2012). <https://doi.org/10.1145/2335867.2335868>
5. Abou-Zahra, S.: Web accessibility evaluation. In: Harper, S., Yesilada, Y. (eds.) Web Accessibility: A Foundation for Research, pp. 79–106. Springer, London (2008). https://doi.org/10.1007/978-1-84800-050-6_7
6. Lopes, R., Gomes, D., Carriço, L.: Web not for all: A large scale study of web accessibility. In: Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A). W4A '10. Association for Computing Machinery, New York, NY, USA (2010). <https://doi.org/10.1145/1805986.180601>
7. Paternò, F., Pulina, F., Santoro, C., Gappa, H., Mohamad, Y.: Requirements for large scale web accessibility evaluation. In: Miesenberger, K., Manduchi, R., Covarrubias Rodriguez, M., Peñáz, P. (eds.) Computers Helping People with Special Needs, pp. 275–283. Springer, Cham (2020)
8. Bühler, C., Heck, H., Perlick, O., Nietzio, A., Ulltveit-Moe, N.: Interpreting results from large scale automatic evaluation of web accessibility. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) Computers Helping People with Special Needs, pp. 184–191. Springer, Berlin, Heidelberg (2006). https://doi.org/10.1007/11788713_28

9. Sirithumgul, P., Suchato, A., Punyabukkana, P.: Quantitative evaluation for web accessibility with respect to disabled groups. In: Proceedings of the 2009 International Cross-Disciplinary Conference on Web Accessibility (W4A). W4A '09, pp. 136–141. Association for Computing Machinery, New York, NY, USA (2009). <https://doi.org/10.1145/1535654.1535687>
10. Kimmons, R.: Open to all? Nationwide evaluation of high-priority web accessibility considerations among higher education websites. *J. Comput. High. Educ.* (2017). <https://doi.org/10.1007/s12528-017-9151-3>
11. Acosta-Vargas, G., Acosta-Vargas, P., Jadán-Guerrero, J., Salvador-Ullauri, L., Gonzalez, M.: Improvement of accessibility in medical and healthcare websites. In: Nunes, I.L. (ed.) *Advances in Human Factors and System Interactions*, pp. 266–273. Springer, Cham (2021)
12. Snaprud, M., Sawicka, A.: Large scale web accessibility evaluation - a European perspective. In: Stephanidis, C. (ed.) *Universal Access in Human-Computer Interaction. Applications and Services*, pp. 150–159. Berlin Heidelberg, Springer (2007). https://doi.org/10.1007/978-3-540-73283-9_18
13. Costa, D., Fernandes, N., Neves, S., Duarte, C., Hijón-Neira, R., Carriço, L.: Web accessibility in Africa: A study of three African domains. In: Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M. (eds.) *Human-Computer Interaction - INTERACT 2013*, pp. 331–338. Springer, Berlin, Heidelberg (2013)
14. Rau, P.-L.P., Zhou, L., Sun, N., Zhong, R.: Evaluation of web accessibility in China: changes from 2009 to 2013. *Univers. Access Inf. Soc.* (2014). <https://doi.org/10.1007/s10209-014-0385-9>
15. Duarte, C., Matos, I., Vicente, J.a., Salvado, A., Duarte, C.M., Carriço, L.: Development technologies impact in web accessibility. In: Proceedings of the 13th Web for All Conference. W4A '16. Association for Computing Machinery, New York, NY, USA (2016). <https://doi.org/10.1145/2899475.2899498>
16. Lazar, J., Goldstein, D., Taylor, A.: *Ensuring Digital Accessibility Through Process and Policy*. Morgan Kaufmann, Burlington (2015)
17. Snaprud, M.H., Ulltveit-Moe, N., Pillai, A.B., Olsen, M.G.: A proposed architecture for large scale web accessibility assessment. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) *Computers Helping People with Special Needs*, pp. 234–241. Springer, Berlin, Heidelberg (2006). https://doi.org/10.1007/11788713_35
18. Mirri, S., Muratori, L.A., Salomoni, P.: Monitoring accessibility: Large scale evaluations at a geo political level. In: The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility. ASSETS '11, pp. 163–170. Association for Computing Machinery, New York, NY, USA (2011). <https://doi.org/10.1145/2049536.2049566>
19. Fernandes, N., Carriço, L.: A macroscopic web accessibility evaluation at different processing phases. In: Proceedings of the International Cross-Disciplinary Conference on Web Accessibility. W4A '12. Association for Computing Machinery, New York, NY, USA (2012). <https://doi.org/10.1145/2207016.2207025>
20. Martins, J., Gonçalves, R., Branco, F.: A full scope web accessibility evaluation procedure proposal based on iberian ehealth accessibility compliance. *Comput. Hum. Behav.* **73**, 676–684 (2017). <https://doi.org/10.1016/j.chb.2016.12.010>
21. Kous, K., Kuhar, S., Pavlinek, M., Heričko, M., Pušnik, M.: Web accessibility investigation of Slovenian municipalities' websites before and after the adoption of European standard EN 301 549. *Univers. Access Inf. Soc.* **20**(3), 595–615 (2021). <https://doi.org/10.1007/s10209-020-00732-9>
22. Frazão, T., Duarte, C.: Comparing accessibility evaluation plugins. In: Proceedings of the 17th International Web for All Conference. W4A '20. Association for Computing Machinery, New York, NY, USA (2020). <https://doi.org/10.1145/3371300.3383346>
23. Padure, M., Pribeanu, C.: Comparing six free accessibility evaluation tools. *Informatica Economica* **24**(1), 15–25 (2020)
24. Brajnik, G.: Comparing accessibility evaluation tools: a method for tool effectiveness. *Univers. Access Inf. Soc.* **3**(3), 252–263 (2004). <https://doi.org/10.1007/s10209-004-0105-y>
25. WebAIM: The WebAIM Million. (2021). Accessed in 18 of January of 2022. <https://webaim.org/projects/million/>
26. Inal, Y., Rızvanoğlu, K., Yesilada, Y.: Web accessibility in Turkey: awareness, understanding and practices of user experience professionals. *Univers. Access Inf. Soc.* **18**(2), 387–398 (2017). <https://doi.org/10.1007/s10209-017-0603-3>
27. Antonelli, H.L., Rodrigues, S.S., Watanabe, W.M., de Mattos Fortes, R.P.: A survey on accessibility awareness of brazilian web developers. In: Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion. DSAI 2018, pp. 71–79. Association for Computing Machinery, New York, NY, USA (2018). <https://doi.org/10.1145/3218585.3218598>
28. Gupta, S., Gjørseter, T., Giannoumis, G.A.: Web accessibility and web developer attitudes towards accessibility in Mozambique. In: *Universal Access in Human-Computer Interaction. Design Methods and User Experience*, pp. 213–231. Springer, Berlin (2021). https://doi.org/10.1007/978-3-030-78092-0_14
29. Leitner, M.-L., Strauss, C., Stummer, C.: Web accessibility implementation in private sector organizations: motivations and business impact. *Univers. Access Inf. Soc.* **15**(2), 249–260 (2014). <https://doi.org/10.1007/s10209-014-0380-1>
30. Farrelly, G.: Practitioner barriers to diffusion and implementation of web accessibility. *Technol. Disabil.* **23**(4), 223–232 (2011). <https://doi.org/10.3233/tad-2011-0329>
31. Inal, Y., Guribye, F., Rajanen, D., Rajanen, M., Rost, M.: Perspectives and practices of digital accessibility: A survey of user experience professionals in nordic countries. In: Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society. NordiCHI '20. Association for Computing Machinery, New York, NY, USA (2020). <https://doi.org/10.1145/3419249.3420119>
32. Azenkot, S., Hanley, M.J., Baker, C.M.: How accessibility practitioners promote the creation of accessible products in large companies. *Proc. ACM Hum.-Comput. Interact.* **5**(CSCW1), 1–27 (2021). <https://doi.org/10.1145/3449222>
33. Abascal, J., Arrue, M., Valencia, X.: Tools for web accessibility evaluation. In: Yesilada, Y., Harper, S. (eds.) *Web Accessibility: A Foundation for Research*, pp. 479–503. Springer, London (2019). https://doi.org/10.1007/978-1-4471-7440-0_26
34. Fernandes, N., Costa, D., Neves, S., Duarte, C., Carriço, L.: Evaluating the accessibility of rich internet applications. In: Proceedings of the International Cross-Disciplinary Conference on Web Accessibility. W4A '12. Association for Computing Machinery, New York, NY, USA (2012). <https://doi.org/10.1145/2207016.2207019>
35. Brajnik, G., Vigo, M.: Automatic web accessibility metrics. In: Yesilada, Y., Harper, S. (eds.) *Web Accessibility: A Foundation for Research*, pp. 505–521. Springer, London (2019). https://doi.org/10.1007/978-1-4471-7440-0_27

36. Martins, B., Duarte, C.: Large-scale study of web accessibility metrics. *Univ. Access Inf. Soc.* (2022). <https://doi.org/10.1007/s10209-022-00956-x>
37. Tait, A., Davis, S., Niyi-Awosusi, O., Wilhelm, G., Paige, K.: Accessibility. (2021). Accessed in 18 of January of 2022. <https://almanac.httparchive.org/en/2021/accessibility>
38. Dinno, A.: Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. *Stata J.* **15**(1), 292–300 (2015). <https://doi.org/10.1177/1536867X1501500117>
39. Vigo, M., Brown, J., Conway, V.: Benchmarking web accessibility evaluation tools: Measuring the harm of sole reliance on automated tests. In: Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility. Association for Computing Machinery, New York, NY, USA (2013). <https://doi.org/10.1145/2461121.2461124>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.