



An inclusive framework for automated web content accessibility evaluation

Jinat Ara¹ · Cecilia Sik-Lanyi¹ · Arpad Kelemen² · Tibor Guzsvinecz³

Accepted: 14 October 2024 / Published online: 25 October 2024
© The Author(s) 2024

Abstract

Since 1996, web accessibility evaluation has been an important aspect of web development to increase social inclusion for people with special needs. Several web accessibility evaluation and testing tools have been developed to automatically evaluate websites in order to identify barriers for people with disabilities. The developed tools are significant since their aim is to effectively represent accessibility issues. However, a few issues with website accessibility cannot be identified through the existing accessibility testing tools due to several limitations, including (i) inappropriate guideline selection, (ii) ambiguities in guideline understanding, (iii) avoiding user and expert suggestions as evaluation criteria, (iv) limited consideration of semantic perspectives, and (iv) unwillingness to incorporate the updated engineering methods. Therefore, reported results may be unclear and inappropriate for some users. Such limitations are critical factors that reduce the effectiveness of the developed tools. These issues cause unwillingness to use a particular tool or possible adoption of other tools. In this situation, we must identify which aspects are important to incorporate in the development in order to make the developed solution more effective as it allows users to make their website accessible to people with disabilities. In this paper, first, we present a literature review of different existing solutions for web accessibility testing to identify their challenges and limitations. Following the reported findings, we propose an automated web accessibility evaluation framework addressing several accessibility aspects to improve the evaluation results by mitigating the limitations of existing solutions. The proposed framework is validated by comparing it with existing automated solutions considering their functional properties. The proposed accessibility framework might be beneficial for web developers, accessibility engineers, and other practitioners to incorporate into their development and research.

Keywords Automated testing · Acceptability · Accessibility testing · Effectiveness · Fairness · Hybrid testing · Reliability · Textual and non-textual complexity · Web accessibility

✉ Jinat Ara
jinat.ar@mik.uni-pannon.hu

Cecilia Sik-Lanyi
lanyi.cecilia@mik.uni-pannon.hu

Arpad Kelemen
kelemen@umaryland.edu

Tibor Guzsvinecz
guzsvinecz.tibor@zek.uni-pannon.hu

¹ Department of Electrical Engineering and Information Systems, University of Pannonia, Egyetem u. 10, Veszprém 8200, Hungary

² Department of Organizational Systems and Adult Health, University of Maryland Baltimore, 655 W. Lombard st #455B, Baltimore, MD 21201, USA

³ Department of Information Technology and Its Applications, University of Pannonia, Zalaegerszeg, Hungary

Abbreviations

ACT	Accessibility Conformance Testing
AI	Artificial Intelligence
API	Application Programming Interface
CSS	Cascading Style Sheets
CVD	Color Vision Deficiency
DOM	Document Object Model
EDBA	Evaluator Decision Based Assignment
HTML	Hyper Text Markup Language
NCAM	National Center for Accessible Media
NLP	Natural Language Processing
UI	User Interface
VCS	Visual Complexity Score
W3C	World Wide Web Consortium
WCAG	Web Content Accessibility Guidelines

1 Introduction

Nowadays, the web has become the most important platform to get information for people with disabilities (e.g., blindness, cognitive, vision impairment, hearing difficulties, etc.). With the web, people have an opportunity to access a wide array of information (e.g., news, healthcare information, educational resources, banking information, etc.) and do several activities (e.g., online transactions, shopping, doctor appointments, e-health services, etc.) that would be difficult without proper accessibility.

Generally, accessibility is the ease of use of any services, tools, and environments in terms of user capability [1]. In the context of web, accessibility is the ability to ensure consistent web navigation, prototype identification, information extraction, and execution of website functionalities without experiencing any difficulties [2]. For example, the navigation path of websites should be equal for people with disabilities and without disabilities. Recently, several statistics have confirmed that a significant number of users with some sort of disabilities are actively participating online and this number is steadily increasing [3]. About 81% of users with various disabilities still experience difficulty, and sometimes it's quite impossible for them to effectively perceive the content [4]. A few recent studies showed that the majority of websites even fail to maintain the basic accessibility requirements or minimum standards of accessibility [5, 6]. As a result, people with disabilities experience several difficulties with web access. For example, web content information might be difficult to read and understand the meaning, placement of the user interface elements might be difficult to identify or remember, and some interactive designs (dropdown menu, sub tasks, landing page, etc.) might make the content partially or completely inaccessible. Besides, as the accessibility problems are distinct according to the type of disability, thus the problems or difficulties might vary from person to person and in different situations. Among several scenarios of difficulties, more particularly people with vision problems have difficulties understanding content that is written in very small font and specific theme (italic, and bolded). People with color blindness have difficulties recognizing specific colors, and people with cognitive difficulties have issues understanding the meaning of some complex or advanced words, notations, abbreviations, and alerts. Also, people with motion difficulties have issues with scrolling and pointing dropdown menu. In that case, people with disabilities are forced to spend more time on the website to find their required information than people without disabilities. In this context, in our preliminary investigation, we validated a set of 15 webpages from different parts of the world considering different criteria to understand the accessibility of the selected web pages [7]. Our investigation confirmed

that most of the tested websites had accessibility issues that should cause further concern.

Addressing these particular issues, in the last few decades, legislations has been ratified; for instance, Sect. 508 of the Rehabilitation Act was initiated by the United States in 2006 concerning the human rights of people with disabilities [8]. In 2010, the European Union accepted these guidelines and declared that to ensure accessibility of online platforms including the web, imposing these guidelines is mandatory to improve social inclusion. Nowadays, the World Wide Web Consortium (W3C) has taken several steps to improve the accessibility of web platforms by initiating several protocols, recommendations, and guidelines about the accessibility of software and application tools, and web content to make them accessible to all. The W3C published a set of standards called the ‘Web Content Accessibility Guidelines (WCAG)’ which is considered the most effective guideline for web designers and developers to improve access opportunities of web content or web platform [9, 10].

Currently, the most effective technique for improving such an inaccessible scenario is the detection of accessibility issues using web accessibility evaluation tools [11]. Such tools help to evaluate and identify the issues associated with accessibility in terms of accessibility guidelines and provide additional information about how to address the detected issues for future improvement. To assist web practitioners (e.g., web designers, web developers, etc.) and end users, several automated and semi-automated tools have been designed and implemented for website accessibility evaluation. For example, Schiavone and Paterno (2015) [12] proposed an accessibility evaluation tool implementing the updated version of web content accessibility guidelines to address and improve the shortcomings of current evaluation tools. A few authors proposed automatic evaluation tools for dynamic webpage evaluation [13, 14]. Some tools have been developed for personalized web accessibility evaluation for specific disabilities, such as vision impairment [15, 16]. Details about several accessibility evaluation tools can be found in Sect. 2. Though these accessibility evaluation tools are effective in investigating the websites, unfortunately, most of them have several issues that lead to misleading the evaluation process and in the process, reduce the acceptability and reliability of the evaluated result. For example, it is difficult to determine which guidelines are implemented and which cannot be implemented. It is also unclear which guidelines necessitate additional testing, such as user or expert testing, for further validation. Besides, it is difficult to determine the website’s accessibility percentage for each disability, and the assessment terminologies are not clearly defined which makes it difficult to understand how the overall score was calculated.

According to [17], considering accessibility guidelines alone is not enough; additional accessibility requirements should be concurrently considered. Kaur and Vijay claimed that only 50% of the accessibility problems with web content are addressed by the web content accessibility guidelines [18]. As a result, additional specific issues cannot be resolved by simply adhering to web content accessibility guidelines. Therefore, to improve website effectiveness, efficiency, and satisfaction, accessibility guideline incorporation should be together with user and expert suggestions as an additional evaluation criterion. Indeed, accessibility guidelines are written in natural language format that is very general and cumbersome to be applied by designers, developers, and web practitioners which puts pressure on and delays the design and development process. Some recent research accepted this limitation and addressed some additional consequences of this limitation. For example, as WCAG remains subjective, it could be interpreted and implemented in several ways depending on the designers' and developers' individual preferences. As website designers and developers are not accessibility experts and have a limited understanding of accessibility, some guidelines and requirements may be applied differently depending on the scenario and the context. Besides, a certain level of knowledge is required to understand the natural language formatted guidelines and user perspectives. Due to these assorted issues, web development and evaluation processes could be deceptive.

Some studies [19, 20] claimed that most accessibility evaluation tools have limited consideration of semantic aspects and developments are not following advanced engineering techniques. Recently, researchers conducted an extensive investigation on the improvement of the effectiveness of website evaluation process, specifically the automated web evaluation process to assist web practitioners in understanding the accessibility status of their developed web-based applications [21]. Therefore, in this paper, we investigate several proposed solutions for accessibility evaluation to identify frequently appeared issues that limited the effectiveness of their developments. According to our observation and reported findings, we propose a framework considering several aspects that are crucial to include in the development process to limit the existing issues of current automated web accessibility evaluation tools. The proposed accessibility framework demonstrates five aspects considering several criteria to improve ambiguities, such as accessibility guidelines, user and expert suggestion, guideline simplification, automated testing, and issue identification and visualization. As the proposed approach focuses on a wide array of aspects including guidelines, additional criteria, evaluation result computation, visualization, etc., it could be useful in facilitating the evaluation process and

representing the computed results as reliable, acceptable, and fair. Also, compared to the existing system, the aspects addressed in the proposed system are not considered in the existing system which makes the proposed system distinctive. Besides, the proposed framework will be helpful for web practitioners and web researchers to understand the web evaluation process.

Moreover, the first objective of this work is to observe the existing available web accessibility evaluation techniques or solutions to identify their limitations to answer our RQ. Based on the findings of the addressed research question, the second objective is to propose an accessibility testing framework that focuses on several aspects for enhancing the automatic accessibility evaluation process against the observed drawbacks. The contribution of our research aligns with:

- Evaluating the effectiveness of the recently developed accessibility testing and evaluation systems/tools;
- Determining the challenges and limitations of the current accessibility testing processes/methods;
- Presenting an extensive accessibility testing framework considering a wide array of aspects to mitigate the investigated issues and improve the effectiveness of the accessibility testing results.

This paper is structured as follows: Sect. 2 provides the literature review of past studies. Section 3 provides the research methodology and discusses multiple accessibility evaluation techniques or solutions illustrating how they evaluate the accessibility issues, their strength, challenges, and drawbacks. Also, Sect. 3 explains our proposed framework with its main benefits and how it can address the existing research challenges and drawbacks. Section 4 provides the validation results of the proposed framework. Following that Sect. 5 concludes the discussion in detail. Finally, Sect. 6 outlines the conclusions of this study.

2 Literature review

This study aims to evaluate the existing literature related to web content accessibility evaluation process in order to identify their challenges and limitations by conducting a literature review. This phase is divided into three steps: (i) planning the literature review, (ii) conducting the literature review, and (iii) reporting the findings.

2.1 Planning the literature review

The main sub-activities related to the planning of the literature review are (i) specification of the research question, (ii)

formulation of the search string, and (iii) database selection. All these sub-activities are described below.

2.1.1 Research questions

The formulation of the research question is the initial stage of a literature selection. As a result, we developed the research question based on our focused research area:

RQ1 What are the challenges and drawbacks of the existing web accessibility testing process?

2.1.2 Search strings

In order to answer the research question, it is crucial to identify and evaluate the available web accessibility evaluation framework that contributes to improving the accessibility of webpages. To achieve this aim, preliminary research activities such as extracting literature from the scientific database are the first important tasks [22]. Therefore, we defined a list of keywords based on our research question regarding web accessibility framework or web accessibility improvement process in order to choose the relevant search strings. We manually searched several scientific databases using the developed list of keywords, and we then fine-tuned it based on how well the results aligned with the goal of the study.

The following is the final list of keywords that were chosen and represented by a Boolean operation: (“web accessibility evaluation framework” OR “techniques” OR “methods” OR “tools”) AND (“Automated web accessibility evaluation criteria” OR “statistics”) AND (“Web accessibility validation by user” OR “expert”) AND (“Web accessibility evaluation for people with disabilities” OR “impairments” AND “Accessibility issues with web”) AND (“Accessibility validation”).

2.1.3 Database selection

Database selection is essential for identifying the most recent and pertinent publications. There are many scientific databases available, so choosing the right one is essential. Here, we have chosen six well-known databases that offer high-quality scientific publications and provide multidisciplinary research work including information science which categorised our research objectives. These databases extracted the most relevant literature based on the user’s interests using sophisticated search algorithms. Six databases used in this literature selection process are IEEE Xplore, Google Scholar, ACM digital library, Springer, Scopus, and Web of Science directories.

2.2 Conducting the literature review

The aim of this phase is to describe review activities by explaining (i) literature extraction, (ii) inclusion and exclusion implication, and (iii) data extraction and quality assessment. These sub-activities are described in detail in the following subsections.

2.2.1 Literature extraction

To extract past literature, we examined the search terms in six different databases. Scholarly committees approve these databases for scientific publication. Almost all of the literature is freely accessible. These databases extract relevant literature based on search strings using their sophisticated search algorithms and semantic technologies. In total, 140 records were found from six databases in the period from 2016 to November 2023 (IEEE Xplore: 15, Google Scholar: 53, ACM digital library: 6, Springer: 7, Scopus: 25, and Web of Science: 34.). The search result is shown in Fig. 1 by taking into account the number of papers that were chosen from each database by the search query. Nonetheless, compared to other databases, Google Scholar, Web of Science, and Scopus offer a greater number of literatures. Among 140 papers, we have chosen the most pertinent papers needed for this evaluation using inclusion and exclusion criteria (explained in the following section).

2.2.2 Inclusion and exclusion criteria

The most appropriate studies for this research were selected from the obtained literature after the evaluation. The literature that didn’t fit the review’s inclusion requirements was eliminated. The following criteria were used for inclusion: publications that are written in English, published between 2016 and 2023 (November) in peer-reviewed conferences or journals (not books), and that discuss the advancement or growth of web accessibility evaluation framework.

The purpose of the exclusion process was to remove publications from this review. The exclusion criteria were the following: duplicate papers, papers written in languages other than English, papers that are not directly related or irrelevant, papers that are not publicly available, and non-research papers like posters, letters, theses, and editorials. After applying inclusion and exclusion criteria to 140 papers, the following observation was made: forty-four (44) papers were duplicates and were not freely accessible or downloadable, ten (10) articles were literature reviews or non-English papers, forty-seven (47) articles were not appropriate for our research objective. For example, they did not focus on the web accessibility framework or web accessibility improvement process. Only few articles considered the

Fig. 1 Number of selected studies per database

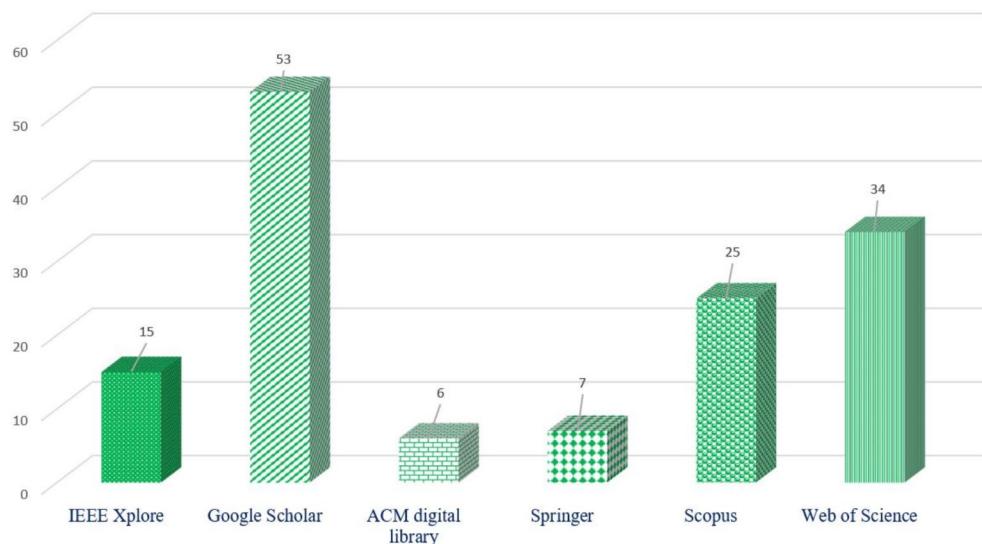
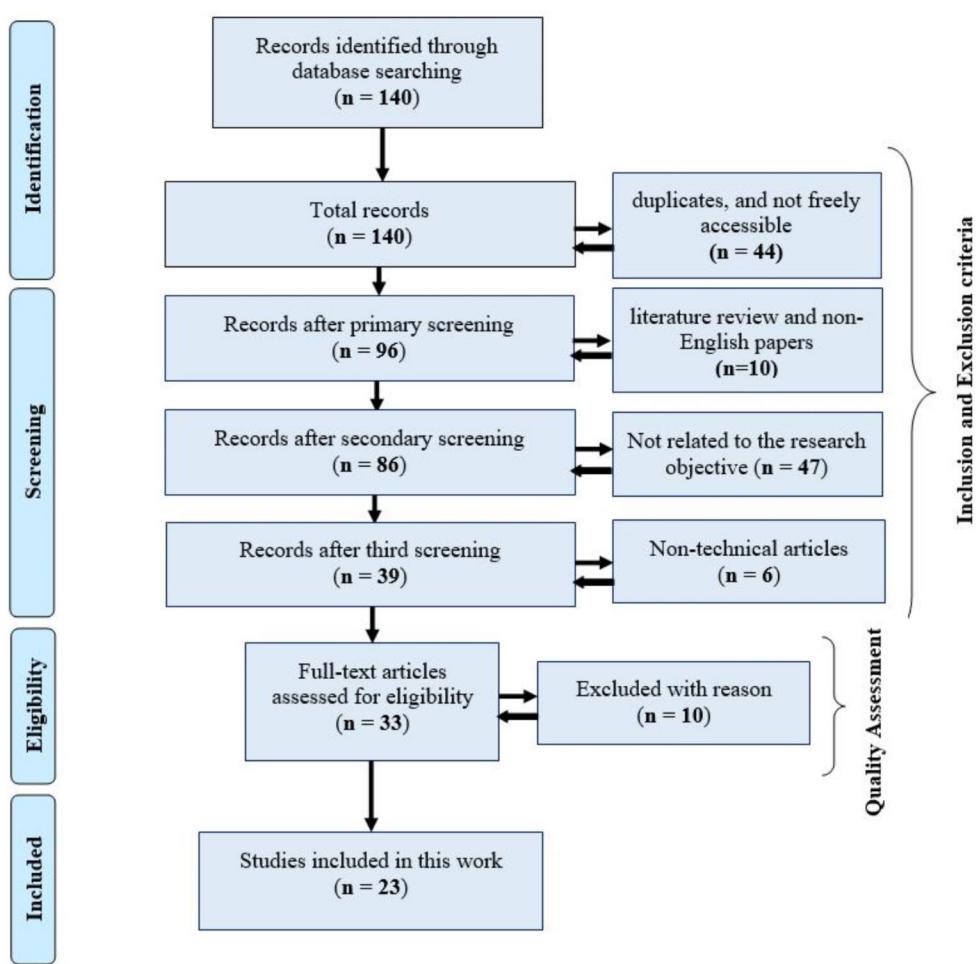


Fig. 2 PRISMA flow diagram for literature selection



guidelines and their effectiveness, and few were focusing on the improvement suggestions and six (6) articles were non technical articles. In total, we excluded 107 papers from the preliminary screening. Finally, 33 articles met all inclusion criteria and were eligible for this study. The entire literature

selection process has been performed through the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) technique [23]. Figure 2 shows the flowchart of the article selection process based on PRISMA. After finalizing the related literature, we analyzed the selected articles

considering our stated research question and reported the findings in Sect. 3 to answer our research question.

2.2.3 Data extraction and quality analysis

We did this analysis based on the eligible 33 papers. Quality assessment and data extraction are crucial for identifying the most relevant paper with the research aim. This method was used in a number of previous literature reviews for the main assessment of the chosen research. Consequently, in order to find high-quality related studies, finish the paper reading process, and respond to our research questions, we adhered to several assessment criteria. The assessment criteria for the evaluation of certain studies are described in Table 1.

We set the score to either 0 or 1 for each question. A paper receives a score of 1 for each positive response. The score is zero if it is not pertinent to the evaluation questions. The extra points for the Q1 indexed journal are +0.50. Likewise, the additional points are +0.40, +0.30, and +0.20 for the Q2, Q3, and Q4 indexed journals, respectively. Equations 1 and 2 were utilized to determine the final score and the normalization score which was used to estimate the standard of each selected study.

$$\text{Score} = \sum (QA_1 + QA_2 + QA_3 + \text{additional points}) \quad (1)$$

$$\text{Normalization} = \frac{\text{Score} - \min(\text{Score})}{\max(\text{Score}) - \min(\text{Score})} \quad (2)$$

Following the quality analysis, we only take into account the studies that obtained $\alpha >= 0.4$ normalized ratings on at least three assessment questions. However, based on the results of the quality assessment criteria, ten (10) of the thirty-three (33) selected studies were eliminated from this review (as indicated in PRISMA diagram). The results of the 23 qualifying papers' quality assessments are shown in Table 2 for this review.

Table 1 Quality assessment questionnaire of selected studies

Questionnaire for analysis	Options	outputs
QA.1: Is the paper a journal article or indexed in SJR or JCR or conference paper?	(+1) Yes/ (+0) No	If yes, rank them
QA.2: Is web accessibility evaluation method or process proposed in the paper?	(+1) Yes/ (+0) No	If yes, what method is presented?
QA.3: Does the paper have significant findings for improvement of accessibility evaluation process?	(+1) Yes/ (+0) No	If yes, what are the findings?

3 Methodology

3.1 Reporting the findings

In this section, our prime aim is to describe our findings from the selected literature in the context of our stated research question.

RQ1: What are the challenges and drawbacks of the existing web accessibility testing process?

To answer the addressed research question, we analyzed 23 selected studies. From these studies, it can be concluded that, accessibility testing is the process of validating accessibility status of online content (e.g., websites/webpages) considering the requirements of people with disabilities (e.g., vision impairment, cognitive impairment, motion difficulty, etc.) [33]. After evaluating 23 selected studies, we provided information related to their research area, evaluation type, and categorical information in Table 3. This table indicates that all of the selected studies are from the web accessibility domain. In terms of evaluation type, all the reviewed studies are grouped into two groups of studies. The first group of studies deals with how to test web accessibility features automatically without human involvement, which is also called Automated Testing. The second group of studies deals with how to incorporate automated testing and human evaluation (performed by hiring people including experts and users), also known as Hybrid Evaluation. Furthermore, in terms of category, all the studies focusing on automated testing proposed three categories of studies (Declarative model, Ontological model, and Algorithmic evaluation), and studies focusing on hybrid evaluation mostly proposed two categories of studies (Crowdsourcing system and Heuristic Approaches).

3.1.1 Automated testing

In the context of web accessibility, automated testing refers to the validation of accessibility features of website content through computer programs against accessibility guidelines. In other words, automatic testing is a process of automatically executing a set of tasks to validate a set of patterns of websites. The importance of automated accessibility testing has increased in recent times as it reduces testing time, minimizes the associated cost, and makes the testing process faster than other testing processes [24]. Besides, it allows testing a wide array of websites without having any difficulties. Focusing on these advantages, several research works have been conducted to develop automated accessibility testing tools or methods in the context of website evaluation that are classified into three categories: declarative model,

Table 2 Quality assessment result of the selected studies

No	References	Type of paper		Quality assessment parameter			Normalization (min_score = 1.0; max_score = 3.50)		
		Journal	Conference	QA.1	QA.2	QA.3	points	score	Quality
01	Hilera and Timbi-Sisalima [1]	✗	✓	1	1	1	0.0	3.0	0.8
02	Giraud et al. [2]	✓	✗	1	1	1	0.50	3.50	1.0
03	Mazalu and Cechich [10]	✓	✗	1	1	1	0.20	3.20	0.8
04	Acosta-Vargas et al. [3]	✓	✗	1	1	1	0.50	3.50	1.0
05	Li et al. [4]	✗	✓	1	1	1	0.0	3.0	0.8
06	Rashida et al. [6]	✓	✗	1	1	1	0.30	3.30	0.9
07	Alsaedi [11]	✓	✗	1	1	0	0.40	2.40	0.5
08	Martins et al. [13]	✓	✗	1	1	1	0.50	3.50	1.0
09	Bonacin et al. [15]	✓	✗	1	1	1	0.40	3.40	0.9
10	Michailidou et al. [16]	✓	✗	1	1	1	0.50	3.50	1.0
11	Robal et al. [17]	✗	✓	1	1	0	0.0	2.0	0.4
12	Duarte et al. [19]	✗	✓	1	1	0	0.0	2.0	0.4
13	Giovanna et al. [21]	✓	✗	1	1	1	0.40	3.40	0.9
14	Jens Pelzetter [22]	✗	✓	1	1	1	0.0	3.0	0.8
15	Mohamad et al. [24]	✗	✓	1	1	0	0.0	2.0	0.4
16	Boyalakuntla et al. [25]	✗	✓	1	1	1	0.0	3.0	0.8
17	Shrestha [26]	✗	✓	1	0	1	0.0	2.0	0.4
18	Ingavélez-Guerra et al. [27]	✗	✓	1	1	1	0.0	3.0	0.8
19	Song et al. [28]	✓	✗	1	1	1	0.40	3.40	0.9
20	Li et al. [29]	✗	✓	1	1	0	0.0	2.0	0.4
21	Alahmadi [30]	✗	✓	1	1	0	0.0	2.0	0.4
22	Song et al. [31]	✗	✓	1	1	0	0.0	2.0	0.4
23	Hambley [32]	✗	✓	1	1	0	0.0	2.0	0.4

Table 3 Summary of the evaluated studies

References	Research area	Type	Category
Hilera and Timbi-Sisalima [1]	Web accessibility	Automated Testing	Ontological model
Rashida et al. [6]	Web quality	Automated Testing	Algorithmic evaluation
Alsaedi [11]	Web accessibility	Automated Testing	Algorithmic evaluation
Bonacin et al. [15]	Web accessibility	Automated Testing	Ontological model
Michailidou et al. [16]	Visual accessibility	Automated Testing	Declarative model
Robal et al. [17]	Web usability	Automated Testing	Ontological model
Duarte et al. [19]	Web accessibility	Automated Testing	Algorithmic evaluation
Jens Pelzetter [22]	Web accessibility	Automated Testing	Declarative model
Boyalakuntla et al. [25]	Web accessibility	Automated Testing	Declarative model
Shrestha [26]	Image Accessibility	Automated Testing	Declarative model
Ingavélez-Guerra et al. [27]	E-learning platform Accessibility	Automated Testing	Ontological model
Giraud et al. [2]	Web accessibility and usability	Hybrid evaluation	Heuristic Approach
Mazalu and Cechich [10]	Web accessibility	Hybrid evaluation	Heuristic Approach
Acosta-Vargas et al. [3]	Web accessibility	Hybrid evaluation	Crowdsourcing System
Li et al. [4]	Web accessibility	Hybrid evaluation	Heuristic Approach
Martins et al. [13]	Web accessibility and usability	Hybrid evaluation	Crowdsourcing System
Giovanna et al. [21]	Web accessibility	Hybrid evaluation	Crowdsourcing system
Mohamad et al. [24]	Web accessibility	Hybrid evaluation	Crowdsourcing system
Song et al. [28]	Web accessibility	Hybrid evaluation	Crowdsourcing system
Li et al. [29]	Web accessibility	Hybrid evaluation	Crowdsourcing System
Alahmadi [30]	Web accessibility	Hybrid evaluation	Crowdsourcing system
Song et al. [31]	Web accessibility	Hybrid evaluation	Crowdsourcing system
Hambley [32]	Web accessibility	Hybrid evaluation	Crowdsourcing system

ontological model, and algorithmic evaluation (as shown in Table 3).

3.1.1.1 Declarative model (DM) This section describes the framework that is considered a declarative model for web content accessibility testing. Among 23 papers, four (4) presented a declarative model (representing 17% of the total literature) that emphasized an automated accessibility testing process. These studies were grouped into seven major drawbacks, as presented in Table 4.

Despite having several advantages of automated accessibility testing, a few researchers claimed that existing automated testing processes have several limitations. For example, most of the existing automated accessibility testing tools do not support browser plugins and need the installation of multiple packages, which makes the evaluation process challenging and may discourage users from utilizing them. Boyalakuntla et al. [25] developed an automated accessibility evaluation tool to assist web accessibility testing through the command line and browser plugin facilities to address these challenges. The tool supports WCAG 2.1 and WCAG 2.2 focusing on aria, color-contrast, Hyper Text Markup Language (HTML) checking, and interaction-related issues. It displays a list of errors as well as suggestions for how to repair them with a snippet of code. Although the proposed approach is effective, few issues are limiting its effectiveness as they assess websites in terms of 16 success criteria of WCAG 2.1, and 2.2, even though additional success criteria must be implemented or verified to represent the full picture of the accessibility situation; they do not compute an overall accessibility score, that might not accurately reflect the accessibility situation of the tested website.

Pelzetter [22] addressed how the vagueness of the accessibility requirements causes a number of anomalies in the results of automated accessibility testing. They proposed a declarative model to evaluate the accessibility status of the websites by incorporating small test sets considering the Accessibility Conformance Testing (ACT) rule set and ontology modelling. Even though the proposed system is capable, certain potential issues have been observed that could restrict the effectiveness of the evaluation result. For instance, ontology modelling introduces ambiguity during

the testing process, which reduces the effectiveness of the evaluated result. Also, implementing ACT rules is quite difficult as it requires resources and experience, which may not be convenient for practitioners.

Others asserted that understanding visual complexity is an emerging requirement for web accessibility evaluation though many of the automated tools don't consider it due to its associated weaknesses or problems or difficulties. For example, as picture descriptions are either written manually or generated automatically, they may not be appropriate or suitable to understand for people with disabilities. Especially, it might be difficult for those with vision impairment to interpret the content of an image owing to an improper image description. Considering this, Michailidou et al. [16] proposed an automated tool to assess the visual complexity of the content and generate the Visual Complexity Score (VCS) based on common aspects of an HTML Document Object Model (DOM) to predict and visualize the complexity of the web page in the form of a pixelated heat map. Addressing the same issue, Raju Shrestha [26] proposed a neural network framework for an automatic evaluation of image descriptions according to the National Center for Accessible Media (NCAM) principles. In order to increase the effectiveness of the proposed framework, they also incorporated expert knowledge (people who understand image accessibility) and the universal design process. Even though these two proposed systems can accurately predict accessibility issues, especially for those who have visual impairments, a few issues reduce the effectiveness of the suggested solution, such as evaluating images for visual complexity, the proposed systems consider a small number of assessment features, incorporate a small number of guidelines and checkpoints, and do not compute the overall accessibility score to indicate the accessibility status of the evaluated component.

3.1.1.2 Ontological model (OM) The framework for web content accessibility testing which is regarded as an ontological model is explained in this section. Out of 23 publications, four (4) proposed ontological models emphasizing automated accessibility testing procedures (representing

Table 4 The four studies related to declarative model (DM), grouped by seven major challenges

References	Identified drawbacks
Boyalakuntla et al. [25]	DM ₁ . Evaluate against a limited number of guidelines
Shrestha [26]	DM ₂ . Absence of accessibility score computation
Jens Pelzetter [22]	DM ₃ . Ontology modelling introduces ambiguity
Michailidou et al. [16]	DM ₄ . Implement guidelines that are not a standard guideline
Shrestha [26]	DM ₅ . Lack of consideration of user and expert evaluation
	DM ₆ . Limited number of assessment feature
	DM ₇ . Limited checkpoints considered

Table 5 The four studies related to ontological model (OM), grouped by four major challenges

References	Identified drawbacks
Bonacin et al. [15]	OM ₁ . Expert opinion/assessment is required
Robal et al. [17]	
Ingavélez-Guerra et al. [27]	
Robal et al. [17]	OM ₂ . Update the ontology is challenging and requires a level of expertise OM ₃ . Adding new guidelines is a complex and laborious task
Hilera and Timbi-Sisalima [1]	OM ₄ . Observed inconsistency in the knowledge base and database

17% of the total literature). These studies were grouped into four major drawbacks, as presented in Table 5.

According to several studies, the process of color-intensive website design may alter how accessible a website is for people who have color vision issues, such as color perception or differentiation. In order to increase the accessibility and interaction of Color Vision Deficiency (CVD) persons with the web, Bonacin et al. [15] designed an ontology-based framework for adaptive interface development. Using this method, it is possible to identify the ideal recoloring interface for CVD users based on their personal preferences. Additionally, Robal et al. [17] addressed that the user interface of websites should be developed with end-user requirements in mind so that users can navigate the structure easily and smoothly and understand the information being shown to them. To ensure these aspects, they developed an ontology-based automated evaluation of the website user interface (UI) to determine the accessibility and usability of the UI.

Furthermore, to assess website accessibility, Hilera and Timbi-Sisalima [1] designed a universal architecture that focuses on web services and semantic web technologies. They incorporated multiple evaluation tools and generated results according to the semantic similarity of multiple reports obtained by each tool. Similarly, Ingavélez-Guerra et al. [27] provided a strategy based on ontology and knowledge modelling that supports accessibility analysis and evaluation of learning objects, highlighting the relevance of knowledge representation about learning objects with a focus on WCAG.

These developments are effective, although there are a few difficulties that have made these improvements less successful such as frequent updating and adding new guidelines are arduous processes in ontology-based solutions that practitioners may not want to employ. Without a professional review of the validated outcome, evaluation results may be deceptive, and the user may not find them fully acceptable.

3.1.1.3 Algorithmic evaluation (AE)

The framework for algorithmic evaluation of web content accessibility testing

Table 6 The three studies related to algorithmic evaluation (AE), grouped by four major challenges

References	Identified drawbacks
Rashida et al. [6]	AE ₁ . Limited number of assessment feature
Duarte et al. [19]	AE ₂ . No expert evaluation
Abdullah Alsaeedi [11]	AE ₃ . A limited number of tools have incorporated AE ₄ . Missing possible web accessibility violations and the corresponding error messages generation function

is described in this section. Out of 23 studies, three (3) presented algorithmic evaluation with a focus on the automated accessibility testing procedure (representing 13% of the total literature). These studies were grouped into four major drawbacks, as presented in Table 6.

Recently, a few researchers and academics have expressed concerns about website quality, which needs extra attention to satisfy end users' demands. They claimed that website quality also gives an indication of how accessible a website is. With this in mind, Rashida et al. [6] presented an automatic evaluation process by putting forward three algorithms for the evaluation of content information, loading time, and overall performance attributes that are usually disregarded in many approaches. Results from experiments demonstrate the usefulness of the suggested automated tool. In addition to tackling the same problem, Alsaeedi [11] presented an algorithmic evaluation framework that would incorporate different automated accessibility testing tools to evaluate the accessibility of websites. The framework enables the selection of sets of evaluation tools prior to the test and allows comparison of the accessibility status of the old and new versions of a given website.

Furthermore, concerning the semanticity of web content, Duarte et al. [19] reported that current automated web accessibility testing tools are unable to assess rules and techniques semantically and evaluate the web content incorrectly. To mitigate this problem, they proposed an automated tool that determines the similarity between content and its textual description from the perspective of web content accessibility guidelines. The semantic similarity measurement performs through the SCREW algorithm to measure similarity between a set of textual descriptions of web content. They represent the accessibility of web content in terms of their computed similarity score.

Algorithmic evaluation is effective in automated accessibility evaluation, although the methodologies mentioned above have a number of limitations that restrict their usefulness and applicability. For example, assessment features are very few and other features must be taken into account in the evaluation process. To validate the algorithmic solution or evaluation result, a wide array of user and expert

interventions is required that will help to validate the evaluated results and improve the acceptability to the end user.

3.1.2 Hybrid evaluation

The W3C has been offering a list of web accessibility testing tools in recent years. Unfortunately, these tools are not frequently updated to use the most recent version of the accessibility guidelines and are not able to keep up with the latest technology [34]. For example, mobile browsing is growing in popularity every day, but in some cases, due to the technological design and development process, they are not usable or viewable on different devices and screen resolutions. Generally, accessibility evaluation of websites is a challenging task that demands the incorporation of several aspects (web features, requirements of people with disabilities, expert opinion, etc.) to reduce the shortcomings of automated accessibility testing results. Therefore, these kinds of problems might be omitted from the automated accessibility evaluation process.

To avoid such scenarios, researchers suggested that hybrid evaluation could be effective in mitigating such issues to improve the accuracy of the evaluated result and retain the fairness of the evaluation process as it is not only a code-oriented evaluation, but also allows human investigation. In other words, hybrid evaluation process is the way of evaluating websites in terms of automated, user, and expert evaluation. Moreover, it incorporates the user and expert requirements and suggestions to improve the effectiveness of the evaluation process. Several researchers have conducted hybrid evaluations of web accessibility which is divided into two categories: Crowdsourcing systems and Heuristic approaches (as shown in Fig. 2).

3.1.2.1 Crowdsourcing system (CS) The framework that is thought of as a crowdsourcing method for testing the accessibility of web content is described in this section. Nine (9) of the 23 studies (or 39% of the total literature) presented crowdsourcing systems with an emphasis on accessibility testing procedures. As shown in Table 7, these studies can be categorized into nine main shortcomings.

Numerous studies have been conducted focusing on several issues to improve the effectiveness of the crowdsourcing system. For example, evaluating some specific issues (e.g., access, navigation, etc.) with disabilities, requires manual assessment (e.g., user and expert assessment). Non-expert evaluation could result in inaccurate validation of these concerns. Songet al. [28] developed a crowdsourced-based web accessibility evaluation system merging user and expert evaluators with an inferencing technique to produce trustworthy and accurate web accessibility evaluation results to improve user and expert assessment processes. Web designers may find accessibility issues and solutions with the help of the provided accessibility reports. Besides, Mohamad et al. [24] proposed a hybrid approach to assess websites' accessibility incorporating accessibility knowledge, automated tools assessment, and expert feedback to validate multiple webpages/websites. They employed a set of rules through a rule engine to do the inferencing process and a decision support system to compute the accessibility evaluation report that makes the process effective. Furthermore, Li et al. [29] addressed that automated testing is effective, though few checkpoints/success criteria of web content accessibility guidelines demand manual judgment. However, in manual judgment/testing, the main challenges are associated with the burdensome and excessive workload for the evaluator. Addressing these challenges, the authors of this paper proposed an advanced crowdsourcing-based

Table 7 The nine studies related to crowdsourcing systems (CS), grouped by nine major challenges

References	Identified drawbacks
Song et al. [28]	CS ₁ . A limited number of checkpoints
Song et al. [28]	CS ₂ . Not time efficient/time-consuming
Mohamad et al. [24]	
Li et al. [29]	
Tahani Alahmadi [30]	
Giovanna et al. [15]	
Acosta-Vargas et al. [3]	
Mohamad et al. [24]	CS ₃ . Absence of empirical validation
Li et al. [29]	CS ₄ . Cost-sensitive
Acosta-Vargas et al. [3]	
Tahani Alahmadi [30]	CS ₅ . Difficult to perceive
Hambley [32]	
Giovanna et al. [21]	CS ₆ . Limited number of evaluation tools have incorporated
Martins et al. [13]	
Song et al. [31]	CS ₇ . Evaluation results could be inappropriate due to the lack of accessibility knowledge
Hambley [32]	CS ₈ . A limited number of evaluation metrics might bias the performance
	CS ₉ . Expert and user opinion/assessment is required

web accessibility evaluation process that is effective for facilitating the manual testing process (e.g., user, and expert testing). The proposed technique configured and simplified the evaluation system focusing on the learning system, task assignment system, and task review process.

Furthermore, addressing the accessibility and usability issues, Alahmadi [30] proposed a crowdsourcing system for web accessibility evaluation considering the subjective and objective measurements. They incorporated several accessibility and usability criteria, automated systems, and human incorporation to reduce the amount of effort and time, for interactivity on the webpage during the evaluation process. In order to provide flexible and open support for a variety of accessibility difficulties, Broccia et al. [21] presented a crowdsourcing system to evaluate website accessibility including the results of automated accessibility testing and usability testing. They incorporated usability testing to validate the result in a qualitative and quantitative manner. A heuristic method based on the user barrier computation process was also proposed by Acosta-Vargas et al. [3] to increase user satisfaction, productivity, security, and effectiveness. The proposed approach was developed by incorporating UX Checker, evaluators (who are experts in web accessibility), and users with low vision. Another study by Martins et al. [13] noted that there are several accessibility and usability problems associated with existing guidelines that are challenging to pinpoint using automated methods alone. Thus, they suggested a hybrid process for evaluating web accessibility that combines automatic evaluation (using the ACCESSWEB platform) and manual tasks (e.g., user and expert evaluation).

In order to improve the task assignment process, the crowdsourcing system could be divided into tedious micro-tasks that can be solved in parallel by workers. Most of the crowdsourcing systems validate websites using an expert who has a level of knowledge in accessibility. They addressed that the validity and reliability of the evaluation result could be significantly improved by incorporating non-experts as they will observe the websites from an end-user viewpoint. Therefore, Song et al. [31] introduced a new crowdsourcing system implementing the Golden Set Strategy and Time-Based Golden Set Strategy. They incorporated automated testing for a few checkpoints and other checkpoints were evaluated manually. The manual evaluation was performed

Table 8 The three studies related to the heuristic approach (HA), grouped by four major challenges

References	Identified Challenges
Li et al. [4]	HA ₁ . Not cost-efficient HA ₂ . Limited number of evaluation checkpoints
Giraud et al. [2]	HA ₃ . Limited number of assessment feature
Mazalu and Cechich [10]	HA ₄ . Expert opinion/assessment required

through non-experts' observation where all the associated task was distributed through Golden Set Strategy and task completion time was allocated according to the Time-Based Golden Set Strategy. The evaluation result depicts that the evaluation time reduces to half of the expert evaluation time and improves the evaluation accuracy.

In addition, Hambley [32] addressed that automated tools have limited coverage of accessibility issues that reduce the acceptance of the accessibility evaluation result. Therefore, they proposed an accessibility testing system incorporating pre-existing evaluation tools in terms of sampling, clustering, and developer testing. The proposed approach is effective in reducing the inaccuracy of the evaluation results.

Although these approaches are effective, some issues reduce their effectiveness. For example, since only a small subset of WCAG checkpoints are possible to evaluate using these methodologies, the results may not accurately reflect the accessibility perspective; it is challenging to implement because it takes a lot of effort, time, financial support, and empirical validation; a wide array of evaluation tools are available, though a few tools are possible to incorporate to these systems that might alter the evaluation result depending on the other evaluation tools taken into account; considering a limited number of evaluation metrics may bias the evaluated result; lack of accessible knowledge may skew the evaluation process and results because the majority of the evaluation process involves human evaluation; in some cases, they avoid expert evaluation due to several complexities, that may reduce the effectiveness of the evaluation results.

3.1.2.2 Heuristic approach (HA) The methodology for heuristic evaluation of web content accessibility testing is described in this section. Among 23 papers, three (3) studies included heuristic evaluations that focused on accessibility testing processes (representing 13% of the total literature). These studies were grouped into four major drawbacks, as presented in Table 8.

Heuristic approaches generally enable human inspection in various ways to evaluate web accessibility. According to past research, several studies have concentrated on heuristic approaches to evaluate the effectiveness and accessibility issues of the web. For example, Li et al. [4] addressed that a certain number of checkpoints during the evaluation of web accessibility require human inspection, such as volunteer participation or expert opinion. Due to the lower level of expertise and inappropriate knowledge, the evaluation task might seem complicated and bring poor evaluation results. To address this issue, they proposed a heuristic approach considering a task assignment strategy called Evaluator-Decision-Based Assignment (EDBA) to

Fig. 3 Number of reviewed studies on each identified limitations or challenges

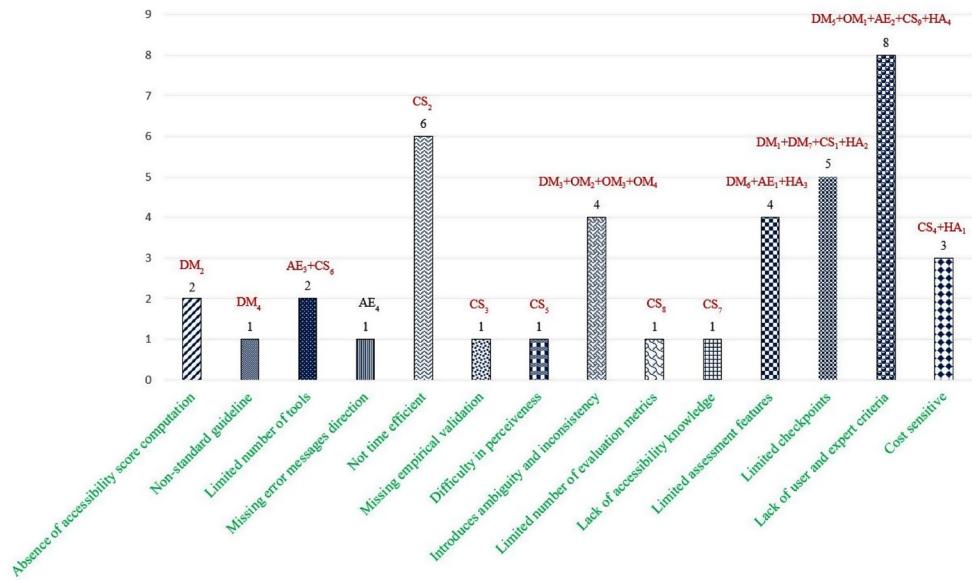
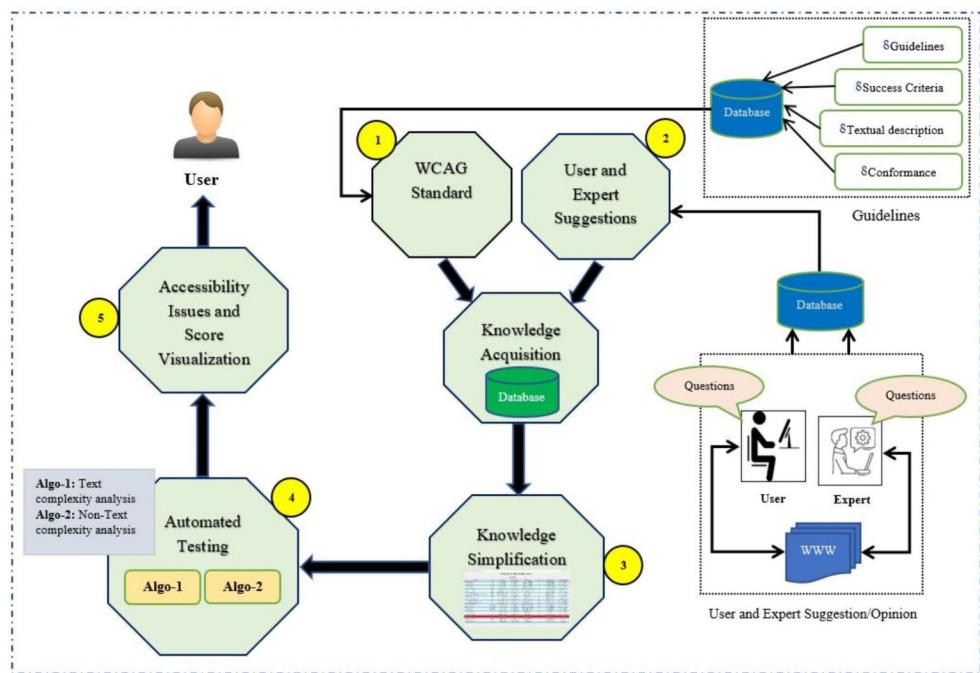


Fig. 4 The proposed automated accessibility testing framework



enhance the selection of participants and experts by using evaluators' prior evaluation records and knowledge of their areas of competence.

Additionally, Giraud et al. [2] argued that adherence to accessibility standards does not guarantee that a website is fully accessible for users with impairments, particularly users with blindness. They pointed out that usability standards are crucial for making a website fully accessible to users with all kinds of disabilities. One critical criterion to enhance the usability of websites is redundant information filtering. They proposed a heuristic approach focusing on redundant and irrelevant information filtering including

participants with blindness. They found that eliminating redundant information and information filtering enhances website accessibility, user satisfaction, and navigation performance.

Furthermore, Mazalu and Cechich [10] highlighted that it is important to consider both developer and end-user requirements to encourage accessibility support for individuals with impairments. To evaluate the intelligent feature specifically for users with visual impairment, they proposed a web accessibility assessment methodology that incorporates a multiagent system. The proposed system was validated through several assessment tools and results.

Although these approaches are effective, several factors, including a limited number of assessment features, guideline implementation, cost-associated difficulties, and expert assessment process reduced the outcome of these proposed systems.

Referring to (Tables 1, 2, 3, 4 and 5), it can be concluded that the existing solutions have several disadvantages that make the developed techniques less effective. For example, in automated evaluation, more specifically in the declarative model, it was noted that consideration of a limited number of guidelines and checkpoints, fewer assessment features, and lack of consideration of user and expert evaluation for accessibility assessment and validation are factors making these processes less effective. Also, lack of attention has been observed in most of the tools as to what samples or objects of the website they have evaluated and what criteria they have implied to assess those website objects. Besides, for the ontological model, the major limitations were found associated with a lack of consideration of expert opinion/assessment, ontology updating issues for adding new guidelines which are complex, laborious, and challenging tasks for the developer, and inconsistency in knowledge base and database that might alter the evaluated results. For the algorithmic evaluation process, a limited number of assessment features, lack of user and expert evaluation, and limited statistics for accessibility error computation are the primary factors to reduce their effectiveness. In contrast, in the context of hybrid evaluation, specifically for the crowdsourcing system, our findings reported that the crowdsourcing system has several issues with user and expert assessment, consideration of accessibility and usability criteria, task distribution, assessment time minimization, and cost reduction that might hamper the evaluation process and limit the advancement of the developed crowdsourcing system. Besides, a few drawbacks to heuristic approaches have been identified, including a limited number of evaluation checkpoints and assessment features, cost-sensitiveness, and lack of an expert assessment process. The majority of the proposed solutions only considered error identification or highlighted the violated guidelines. The effectiveness of the developed methods was constrained by not taking into account the computation of the overall accessibility score. Most works focused on all types of disabilities. In some works, only persons with vision impairments were taken into account.

Figure 3 presents the number of papers on each identified limitation of the reviewed papers. This figure depicts that the most frequent issues or challenges or limitations addressed in the majority of the studies are: lack of consideration of user and expert criteria, most of the solutions are not time efficient, and limited checkpoints and assessment feature consideration. The observation result of our research question concludes that a number of issues appeared in

the existing approaches that make the evaluation results ineffective, which addresses a further concern of the web researcher. In summary, addressing our research question (RQ), the following drawbacks and challenges have been observed frequently in multiple developed solutions that reduced their effectiveness:

- Difficulties understanding the guidelines (WCAG) that are written in natural language format;
- Consideration of only a limited number of WCAG success criteria;
- Insufficient attention is given to user behavior, user requirements, and expert suggestions;
- Challenges in mapping process among success criteria and websites features;
- Semantic features of websites and related engineering techniques are not given enough attention;
- The process of accessibility evaluation and score visualization is ambiguous, thus it is difficult to determine which criteria have been looked into and which have been skipped;
- Terminologies used in assessments are ambiguous and do not accurately reflect their intended meaning.

From the above findings, it could be depicted that the majority of the web accessibility testing processes (both automated and hybrid evaluation process) have several issues that hinder their effectiveness. Following several limitations of the existing web accessibility testing system, there is an emerging need to develop an updated accessibility testing tool to mitigate the existing shortcomings of the current tools. Therefore, in the following part of the paper, our objective is to propose an accessibility testing framework considering the determined aspects that can facilitate to improvement of the limitations of available automated accessibility testing systems and improve the effectiveness of the evaluation results. The proposed accessibility testing framework for automated evaluation is demonstrated in the following section.

3.2 Proposed automated accessibility testing Framework

According to the outcome of the conducted literature review that listed several existing shortcomings (in Sect. 3.1), our observation is that the main aspects leading to incorrect perception, encoding, and development of the accessibility evaluation tool are:

- Understanding difficulties of natural language formatted web content accessibility guidelines;

- Limited consideration of user requirements and expert suggestions;
- Lack of semantic concern.

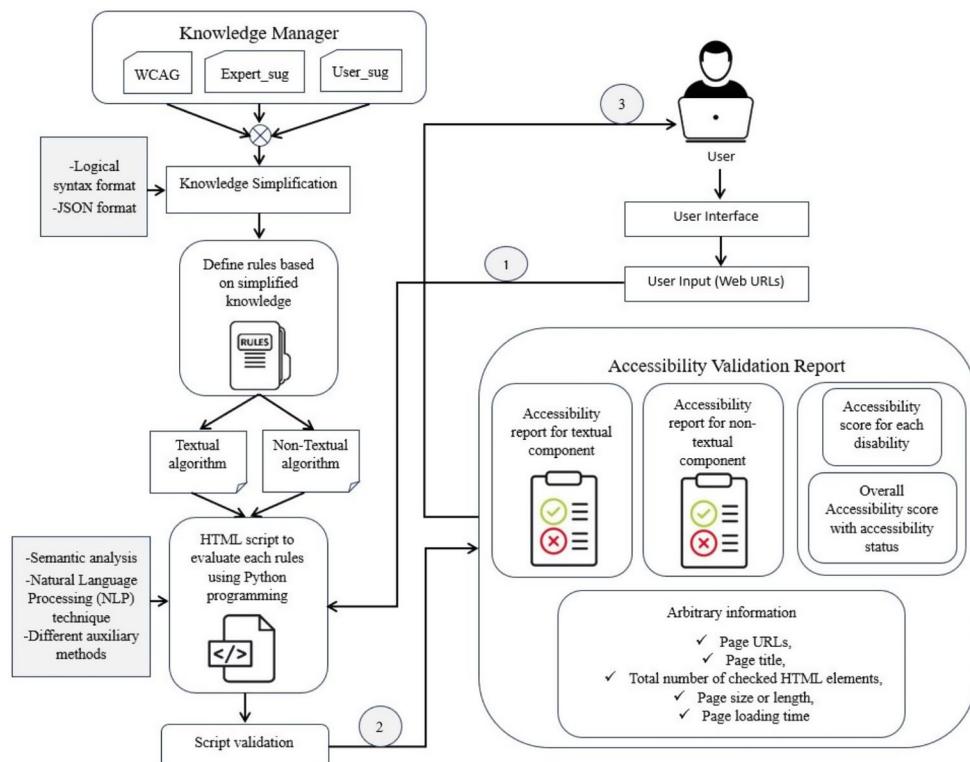
These make the evaluation results less credible and less acceptable. Most of the accessibility testing tools only check a specific number of WCAG success criteria which is around 50% of the total guidelines. As a result, it restricted the evaluation process and the overall evaluation result might be inaccurate. As many web accessibility guidelines cannot be assessed automatically, they do not specify whether guidelines require user/expert testing or not. This could also be a cause of incorrect calculation of matrices and evaluation report formulation. Without incorporating the user's requirements/opinions and expert suggestions during the development of accessibility testing tools, the evaluation process may overlook some crucial aspects and inadvertently inflate the final accessibility score. Besides, a lack of consideration of semantic aspects may reduce the effectiveness of the evaluated results. Therefore, to minimize such issues, an accessibility aspects framework for automated web accessibility testing considering the following aspects could help the development procedure and improve the evaluation process with accurate results:

- Simplifying the updated web content accessibility guidelines to represent the guideline knowledge in the easiest and most effective manner;

- Incorporating all success criteria in the evaluation process to make the evaluation results more effective, while improving the fairness of the evaluated result;
- Incorporating user requirements/opinions with expert suggestions during the evaluation process as an additional evaluation criterion;
- Incorporating separate complexity analysis algorithms for textual feature, and non-textual feature analysis focusing on semantic aspects to improve the effectiveness of the evaluated result;
- Categorizing the evaluated guidelines in terms of user evaluation and expert evaluation when the guideline is not applicable for automatic evaluation;
- Displaying the evaluation result with the overall accessibility score along with specific accessibility scores for each disability type.

Figure 4 shows the proposed automated web accessibility evaluation framework. The proposed framework considers several aspects related to appropriate accessibility guidelines selection, user requirements, and expert suggestions (as additional evaluation criteria) consideration, and guideline knowledge simplification prior to the algorithmic coded process that facilitates the appropriate accessibility evaluation scores computation. Figure 5 depicts a use case diagram that explains the specifics of the accessibility assessment procedure carried out by the proposed

Fig. 5 Use case diagram of the proposed automated accessibility testing framework



framework. All the aspects shown in Fig. 4 are discussed in the following sections.

3.2.1 Aspects of web content accessibility guideline

Several governments and organizations from different countries have presented various accessibility guidelines in recent years. Referring to past studies [35, 36], the Web Content Accessibility Guidelines (WCAG) is the most sophisticated and most widely accepted set of guidelines. WCAG has several versions and it is the most used guideline by most of the existing testing tools. Its newer version (WCAG 2.2) has more features/success criteria than its previous versions WCAG (1.0, 2.0, 2.1). It contains 13 guidelines with 87 success criteria that are distributed into three conformance levels: A, AA, and AAA where 33 success criteria are assigned to level A, 24 success criteria are assigned to level AA and 30 success criteria are assigned to level AAA. The conformance levels ensure the priority of the success criteria in terms of their importance, where A refers to what might be included based on the development/evaluation criteria, AA refers to what should be included and AAA refers to what must be included. However, to make the evaluation process effective and correct, it is crucial to incorporate all success criteria under three conformance levels. Therefore, in our proposed automated accessibility testing framework, we considered every success criterion of the updated web content accessibility guideline (WCAG 2.2) that could facilitate to improve the overall evaluation result. An overview of WCAG 2.2 is shown in Fig. 6. Detailed information about WCAG could be found in [37].

3.2.2 Aspects of user opinions and expert suggestions

From the evaluation of the existing work (Sect. 3.1), it could be depicted that consideration of user opinions and expert suggestions is the most common limitation in their work. As the development of automated web accessibility testing

tools is limited to incorporating few/specific guidelines, the evaluation process might not consider every aspect of difficulties associated with disabilities. Sometimes, incorporation of every success criterion of guidelines is not possible through automated means. Following this limitation, consideration of user opinion and expert suggestion might be a valuable resource for identifying other supplementary requirements as an additional evaluation criterion along with WCAG during the development of an automated accessibility testing tool.

Generally, user opinion refers to the expressed opinion of users based on difficulties they have encountered during the experimentation or web evaluation process [38]. Depending on the sort of experiment and the user opinion collection, user opinions may be collected in different forms. The most common and effective way is questionnaire-based evaluation (questions asked to the user) to express their opinion [39]. In the context of accessibility evaluation of websites, a few researchers mentioned additional aspects that are encountered frequently during website accessing and introduce accessibility issues such as the necessity of webpage availability, manual text size adjustment availability, manual color adjustment availability, necessity of user information, and availability of textual and image CAPTCHA. Unexpectedly, web content accessibility guidelines, including WCAG, do not address these aspects. Therefore, we prepared our questionnaire concerning these aspects that help us understand the user's perspective and obtain their particular requirements regarding every single aspect. Also, understanding user requirements might be helpful to improve the overall accessibility evaluation process.

In the context of expert suggestion, it refers to the recommendation to improve the prototypes of some particular aspects to avoid some unconsidered situations [40]. From the web content accessibility perspective, experts are the people who have a thorough understanding of accessibility standards as well as technical knowledge of website design and development process. Depending on their role,

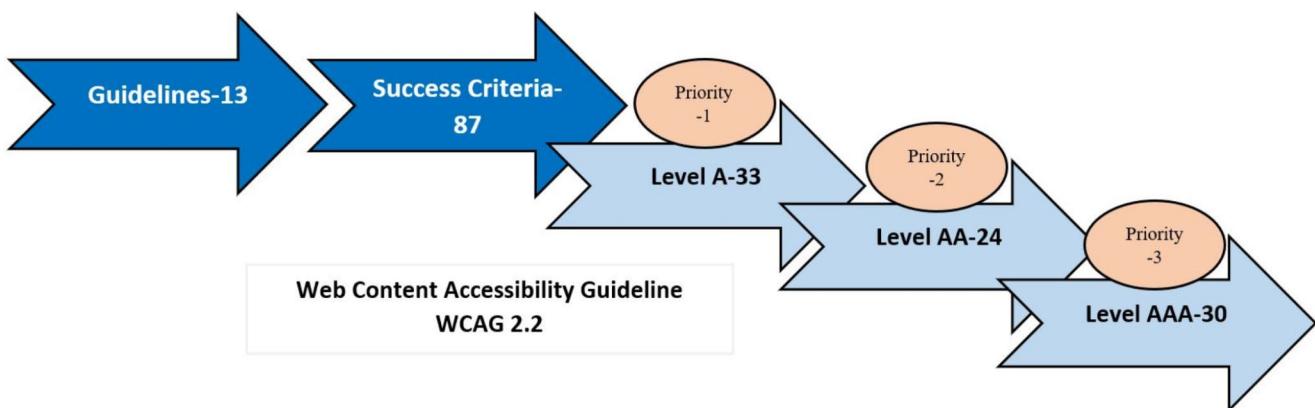


Fig. 6 Overview of WCAG 2.2 (the latest version of WCAG)

accessibility experts may be web developers, web designers, accessibility test specialists, UX/UI experts, and researchers [41]. A few additional aspects such as word/sentence length, specific font family, font size, and other factors require fixed determinants to make the website accessible. To determine these aspects, we considered expert suggestions or perspectives as an additional factor, as they have more expertise, experience, and knowledge in making critical judgments. We interviewed five experts from the Electrical Engineering and Information Systems Department of the University of Pannonia, Hungary to provide insightful guidance based on their knowledge and experience. Out of the five experts, three were professors with over 20 years of experience in digital platform accessibility perspectives, and the other two were PhD students in their final year focusing on digital inclusion and human-computer interaction with over 5 years of experience in accessibility aspects. From expert suggestion, we considered 13 additional factors that are beyond the mentioned criteria in WCAG. These additional requirements include proper loading time, page length, appropriate number of internal/external links, number of images and video content, accessible color pair, proper word length, sentence length, paragraph length, text content length, font size, font family, text pattern complexity (e.g., italic, bold), and content types. By considering expert suggestions along with WCAG, including any new elements they recommend, it might be possible to enhance the website's evaluation results.

After obtaining all the additional criteria from users and expert suggestions/opinions, we incorporated these criteria into our evaluation process or in our proposed framework along with WCAG as additional rules or guidelines to improve the entire accessibility evaluation process.

3.2.3 Aspects of guideline simplification

Guideline simplification is a part of text simplification as guidelines are written in natural language format. It is the process of simplifying the existing guidelines to make the guidelines more comprehensible for the users or associated authorities [42]. As web content accessibility guidelines

are written in natural language format and there is no logical representation of these guidelines, it is relatively difficult to understand and implement these guidelines during web accessibility evaluation tool development [43]. To understand these complex guidelines, adequate accessibility knowledge and high-level technical competence are required. In that regard, we considered the concept of web content accessibility guidelines simplification that can help to represent the guideline simply and effectively. In the guideline simplification process, we categorized all the success criteria of web content accessibility guidelines into eight criteria: guidelines, objects, attributes, components type, requirements, conformance level, beneficiary type, and evaluation type/phase as shown in Fig. 7. This classification process might represent the simplified guideline more systematically which will help to encode every guideline and perceive the website feature appropriately through the developed accessibility evaluation algorithm. Also, guideline modelling helps to encode each guideline semantically.

3.2.4 Aspects of automated testing

Automated accessibility testing methodologies proposed by several accessibility research groups such as Accessibility Conformance Testing (ACT), Source code mining, Application Programming Interface (API) based testing, and Ontology-based testing are prominent. However, these approaches have several limitations when it comes to conducting accessibility testing. For example, for ACT rules' implementation, maintaining a unique ID is necessary for mapping web attributes to one or more WCAG success criteria, which is challenging [44]. In source code mining, the analyzed code might not be well-structured to extract the relevant regularities due to poor precision and recall and sometimes a large number of user involvement is required [45]. Besides, API testing is effective but it does not allow interaction with real user activity and proceeds only with a raw request [46]. In ontology-based evaluation, inconsistency in the knowledge base and database might reduce the effectiveness of the result [47].

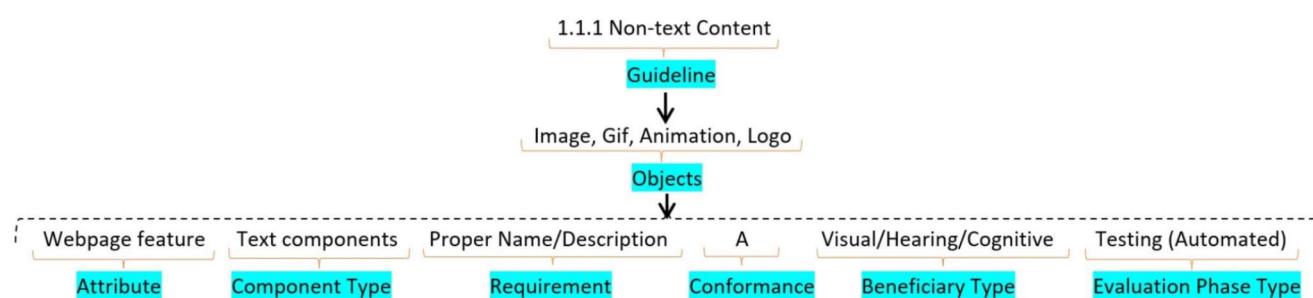


Fig. 7 A simplification process of web content accessibility guideline

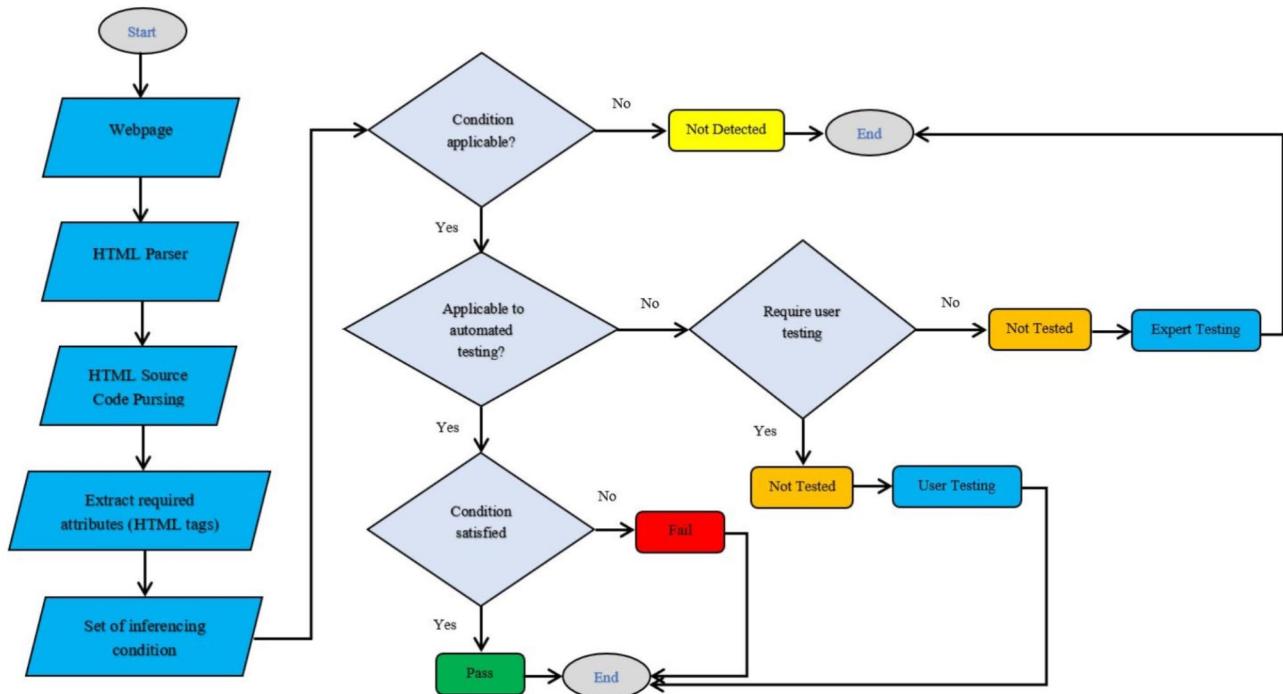


Fig. 8 Work-flow diagram of algorithmic evaluation of accessibility testing

To overcome these difficulties, several studies concluded that algorithmic evaluation of web content is important, especially for accessibility testing of web platforms [1, 16]. Through the algorithmic evaluation process, it is possible to analyze website source code by incorporating every guideline offered by the Web Accessibility Initiative. As WCAG derived most of the prototypes of a website including textual and non-textual content and features, we incorporated two separate algorithms for complexity analysis of textual content and non-textual features that might improve the performance of the algorithm and provide a straight guideline about algorithm specification. To incorporate the textual, and non-textual algorithms, we considered the most effective process to parse the website's HTML code and analyze its features using Artificial Intelligence (AI) techniques that might be effective in assessing each web feature and validating the guidelines and other additional requirements (from user requirements, and expert suggestions) appropriately. Figure 8 shows the workflow diagram of the algorithmic evaluation process for accessibility testing of a particular website. We consider this process is the most convenient process of algorithmic evaluation validating every guideline for every webpage feature and evaluating them in terms of four assessment terminologies such as 'Pass', 'Fail', 'Not tested', and 'Not detected' where Pass refers to those guidelines that have been followed and successfully implemented; Fail refers to those guidelines that have been followed, but wrongly implemented; Not detected refers to those guidelines that should be followed, but not implemented; and

Not tested refers to those guidelines that require the user or expert testing, because the software does not test it.

3.2.4.1 Text complexity To determine the accessibility of webpage content in the context of textual complexity, the potential solution is to derive a text complexity algorithm that evaluates the HTML textual content feature through Natural Language Processing (NLP) to analyze each element associated with textual aspects, validate their accessibility and determine the associated complexity following semantic manner. The WCAG 2.2 states that 35 success criteria are associated with textual elements, which are essential components to determining and ensuring accessibility and reducing the complexity of webpage surfing. The following list of elements are associated with textual aspects that we incorporated in the text complexity algorithm. Besides, for each addressed elements in this section, web content accessibility guidelines (WCAG 2.0) and some other rules or directions about the additional elements are given in Table 10 (Appendix).

[Images]: Image is one of the key elements of web content that is used frequently to convey descriptive information to users. Ensuring proper titles and descriptions of all the images on a webpage such as Image, Gif, Animations,

Logos and decorative images can improve the accessibility of website images.

[Pre-recorded/Live Audio]: In order to share information among users, pre-recorded or live audio content is frequently attached to websites. All audio content, whether pre-recorded or live, like radio webcasts, must include a correct caption and descriptive description to ensure its accessibility.

[Pre-recorded/Live video]: Video is an effective and valuable content of a website to represent information to users. Receiving information through video content is more beneficial than textual content, especially for people with disability. To make these resources more accessible, it is important to define appropriate captions and descriptive descriptions to describe any video content, whether it is live or pre-recorded video, such as video conferencing or live speech.

[Links]: Links are the supplementary resources of a webpage that aim to enhance or extend the knowledge of the web content. A webpage may have several internal and external links to extend its information. Links are not allowed to be broken or unavailable, should be a maximum of 80 characters long, and not be justified to maintain accessibility. The purpose of the link should be properly stated to understand its usefulness, should be distinguishable from the textual content, and must be embedded in 1.5 font.

[Words]: Words are the normal text that is used to represent the content or information of a webpage. To improve the accessibility of the content or represented information, it should be meaningful, and simple to understand. All the words in the content, images, fields, and menu list should be understandable.

[Sentences]: Sentences are the sequential representation of a group of words used to represent the actual meaning or idea. For webpage content, meaningful and simple sentences are essential to improve its readability and accessibility.

[Paragraph]: A paragraph is the combination of multiple sentences to represent complete information of web content to the user. It is a descriptive explanation of the information that is called an extended form of multiple sentences. A paragraph should be meaningful, simple, precise, and contain useful but sufficient information that will help to understand the content.

[Button, icons, fields, Label]: Button, Icons, Fields, and Labels are user interface components. Buttons enable users

to select and proceed with action with a single tap. Icons are generally either graphic files or single images. The information is represented by icons in terms of directions that facilitate the navigation process. Fields are single text boxes that accept input based on its type, including text, numbers, characters, links, etc. Labels are for the representation of information as simple as possible. All of these aspects are crucial for providing a more organized representation of the information and for making website access more enjoyable and comfortable. It is important to provide accurate information and clearly define their function in order to make these features accessible. In addition, the text of these features must be aligned within 80 characters, not be justified, and have proper font size.

[Text]: The webpage text content or textual content should maintain an appropriate color and contrast ratio for different text sizes, such as a normal text size of 4:5:1 ratio, or a large text size of 3:1 ratio (according to WCAG 2.2). Text on a website should have the opportunity to be resized by up to 200% and to change the background and foreground colors, to increase its visual accessibility. To increase its authenticity, repeated and duplicate material must not be added to the webpage. For content representation simplicity, line, paragraph, letter, and word spacing should be kept to 1.5, 2, 0.12, and 0.16, respectively. The CSS pixel width and height should be large enough (at least 320 and 256) to allow content scrolling toward vertical and horizontal alignment.

[Title]: The title is a very important feature to reflect the purpose of the webpage. An accurate, descriptive, and appropriate title is essential to help people to grasp the objective of the webpage.

[Heading and labels]: The secondary essential elements or qualities of the webpage that are most important for accurately representing the information are the headings and labels. These enhance the semantic quality of the web content. A descriptive, meaningful, and appropriate description of headings and labels can improve the accessibility of the webpage content.

[Language]: Language is the unified aspect or feature of a webpage to make it globally accessible to the community. All webpages must be available in an English language page with its native language choice. Besides, the multiple language selection option is an important factor in making the website accessible to users. Moreover, using different languages on the same webpage makes the content robust to grasp and retain its consistency. Thus, keeping a single language for all sections and paragraphs of the content

is another crucial feature of accessibility that must be maintained.

[Idioms]: Idioms are related to complex and unusual words that make the content difficult for the user to understand. Idioms should be avoided in the webpage content.

[Jargons]: Jargon is a complex pattern that makes the content difficult to interpret especially for people with disabilities which reduces the accessibility of the content. Jargon should be avoided in webpage content.

[Abbreviation]: Abbreviations are the shorter form of words or phrases, such as IT, which stands for Information Technology. Although abbreviations might be helpful in certain situations, for accessibility, it is not an appropriate decision to use such short forms as people with disabilities have several issues with the short form of words or phrases. However, in unavoidable situations, it is necessary to provide a broad and expanded form of the used abbreviation to understand its meaning.

[Pronunciation]: Pronunciation is the process of understanding words or sentences without facing any difficulty. To improve pronunciation ability, meaningful words, and sentences are important. Complex and ambiguous words should be avoided in the text content of webpages.

[Reading level]: Reading level refers to the ability to read the content by the user without any difficulty, especially for people with disabilities. Maintaining reading levels is important to improve the accessibility of web content.

[Context-sensitive content]: Several complex words, questions, patterns, and sentences are referred to as context-sensitive content that reduces the accessibility of web content. For these, a proper and detailed explanation could reduce the difficulty of the content understanding.

[Drop-down menu, dialog box, checkbox, combo box]: Drop-down menus provide a list of objects or items to interact with the menu through clicking or cursor hovering. A dialog box is a type of pop-up window that is used to display informational messages including alerts, prompts, and confirmations. Missing or irrelevant information in dialog boxes and drop-down menus may make the website less accessible. Such missing and inappropriate information is not allowed in these elements. A checkbox is a square box that allows it to be ticked or checked for any active action. Combo boxes make it possible to choose an item from a long list of options, making it easier for the user to locate the desired item. Sometimes, difficulties in understanding

the expected input data reduce the accessibility of the content. Therefore, clear instructions about input data should be defined properly such as checking the box, selecting single/multiple objects, etc.

[Search field]: A Search field or search box is mostly used by people with disability who have issues with content navigation or looking for certain information on a webpage. If a webpage lacks organized content, an effective search field can help to improve the navigation. The search field or search box must have an understandable name, a simple design, and appear exactly with the same name and the same way on every page.

[Form]: Forms are important tools on websites for communicating with users for a variety of purposes, including data collection. In order to understand and clarify the instructions regarding the expected input from the user, a web form needs to have a textual format description.

[Error]: If necessary, the error function displays an error message to inform the user about error reason or recommended user action. The defined error message must be appropriate and fully represent the instruction to make the error generation accessible.

[Word length]: Word length indicates how long a word should be. Long words are difficult to pronounce and understand by people with disabilities, thus all words must be shortened as much as possible to make the information accessible to the user. Besides through collecting user opinion or user requirements, it may possible to identify the appropriate length of words that could be helpful for the web developer and researcher.

[Sentence length]: Proper sentence length is an important aspect of making the content accessible and understandable for people with disabilities. A long sentence may not be beneficial to the user for effectively representing the content.

[Paragraph length]: A long paragraph can make the content monotonous and difficult to understand and remember by people with disabilities. An effective and flexible paragraph limit might make the content accessible to users with disabilities.

[Text content length]: When referring to content length, all words, phrases, and paragraphs are included. Long textual content forms difficulty in terms of accessing, searching,

and understanding especially for people with intellectual or cognitive disabilities.

[User information]: In some situations, websites require user personal information such as user name, email address, password, location, a particular interest, etc. These make the website inaccessible as users may not agree to share their information and sometimes users with disabilities don't understand the actual meaning of these requirements.

[CAPTCHA]: Websites with CAPTCHA ("Completely Automated Public Turing test to tell Computers and Humans Apart") have been very common in recent times for security purposes or user behavior understanding. Some users prefer text-based CAPTCHA, while others find image-based CAPTCHA to be more useful. However, few users have issues with CAPTCHA as it requires careful attention. They fail to provide the right response, resulting in repeated attempts or refusals to browse. It frustrates users and reduces their interaction with that specific website.

[Font size]: As a default font size of webpage content, font size must be suitable for every group of people with disability to ensure content accessibility. Mostly, font size 12 is known as an accessible font size for web content, but not for every type of disability such as people with severe vision impairment. Thus, manual font size adjustment is important to ensure content accessibility to every group of people with disability.

[Font family]: As a default font family of webpage content, an appropriate font family is important to make the content accessible for every type of disability. Sans serif is the most frequently and most widely used font family, though it depends on the users' requirements.

[Complexity of text pattern]: Text pattern refers to text representation style such as italic, bold, etc. Inappropriate text patterns can make the content difficult to understand. For some users the 'italic' text pattern is confusing. Therefore, the use of appropriate text patterns may reduce the complexity of the content text and improve its accessibility.

[Website content type]: Content type represents whether a website should contain solely text, images, or video content. A proper content type is important as it helps to represent content and make it more interesting and user-friendly.

3.2.4.2 Non-text complexity

A non-text complexity algorithm could be a potential way for the evaluation of non-textual elements on webpages to determine their accessibility

status. According to WCAG 2.2, 12 success criteria are associated with non-text elements that validate the accessibility of web content. A list of website features related to non-text aspects that we incorporate during our automated web accessibility testing follows next. Also, web content accessibility guidelines (WCAG 2.0) and some other rules or directions about the additional elements are given in Table 10 (Appendix).

[Field, Button, Link]: In terms of accessibility, fields, buttons, and links must have accessible color concerning people with disabilities such as visual impairments including partially blind and color-blind users.

[Image, Logo]: Images and logos are important visual aspects or user interface elements on websites. People with visual and cognitive disabilities may have difficulty accessing these elements if the proper contrast ratio is not maintained. To improve accessibility of these components, the logo should retain a contrast ratio of 7:1, while the suitable contrast ratio for picture content is aligned with a 4:5:1 ratio (according to WCAG 2.2).

[Text in Images]: Text in image refers to the textual content of an image that represents the inherent information of images. To make this textual content accessible, the acceptable contrast ratio for normal text should be 7:1, and the acceptable contrast ratio for large text should be 4:5:1 (according to WCAG).

[Maps, images, Diagram, Data tables, presentation, video]: To improve visual accessibility, appropriate width, and height specification is important for visual content such as Maps, Images, Diagrams, Data tables, Presentations, and video. According to accessibility guidelines, to make these elements widely accessible for everyone, the required width and height of the content should be 320 and 256 CSS pixels, respectively.

[Input size]: Input size specifies the length of the content that is required as input to the input field. In some cases, input size may make content less accessible. For example, a short or long input size may dissatisfy the user. To minimize such issues, the ideal input size is 24 CSS pixels (minimum) AND 44 CSS pixels (maximum).

[Markup language elements]: To retain the webpage structure and content accessible to everyone, all markup language elements for HTML or CSS such as bold tag ``, heading `<h></h>`, paragraph `<p></p>` should define

with start and end tag, and specify the role/property of the elements.

[Loading time]: Webpage loading time is the average time that a webpage takes to appear on the user's screen after searching or browsing through their web address on the search panel. Loading time is important to improve accessibility because if a page has some issues and takes longer than usual to load, user dissatisfaction will arise which may cause the exclusion of that webpage interaction.

[Page length]: Webpage length refers to its display size and content length or navigation time. When a webpage is too long, it can be extremely difficult for people with disabilities to use it because they may have mobility problems or cognitive challenges that make it ineffective for them to browse for a long period. So sufficient material should be correctly organized within a conventional page limit.

[Website availability]: Website availability also referred to as website uptime, is to guarantee that users can browse or access the page whenever they want. In case of failure of website availability, it may reduce the effectiveness of the webpage and users may be less likely to visit that site regularly. Therefore, maintaining website availability is a critical prerequisite for enhancing accessibility.

[Manual text size/font adjustment]: Allowing manual text size/font size adjustment is one of the major factors in reducing accessibility. As preferred text size or font size varies from person to person according to their comfort, allowing users to select several font sizes can enhance the usability and accessibility of a website.

[Manual Color adjustment]: Similar to users manually adjusting font size, offering several color adjustment options to users while navigating a website can increase user satisfaction and accessibility of a webpage. This criterion is especially important for people with color impairment.

[Number of internal and external links]: Internal and external links (hyperlinks) are additional webpage resources that aim to provide more information to users. Unfortunately, most people with disabilities do not consider these additional resources useful and they hardly use these resources. Besides, these resources sometimes make users confused about the actual content and additional content. Thus, a limited number of internal and external links should be added to the websites to avoid these uncertain experiences.

[Number of image content]: To represent webpage information, images are frequently used components. Though

these are useful for sharing content, it sometimes makes it difficult to understand the content objective for people with disabilities. In certain situations, the uses of an excessive number of images cause serious inaccessibility of the content. Thus, a limited number of images should be used on a particular webpage.

[Number of video content]: Sharing information through video content is effective for people without disabilities, but people with impairments may not find it helpful. Users with disabilities may find it difficult to understand the information that is represented within a few minutes. So, the use of video content should be limited on the webpage.

[Accessible color pair]: A set of complementary colors can make colors more accessible. For example, for displaying a webpage banner, a combination of background color and text color may make the content difficult to read or understand. Inappropriate color combinations render the website inaccessible, especially for people with partial visual impairment or colour blindness.

3.2.5 Aspects of accessibility issues and score visualization

After performing the accessibility investigation, we organized/represented the evaluation result by considering several statistics. The major statistics concluded related to the evaluated result in terms of each algorithm, accessibility score for each disability type, overall accessibility score with accessibility status, and arbitrary information of the evaluated webpages.

Regarding the algorithmic evaluation, the first algorithm (Non-Text Complexity Analysis) evaluates the accessibility concerns of the webpage by considering its non-text components. It is able to assess 19 web objects in total, including their functionalities and other aspects. Similar to the first algorithm, the second algorithm (Text-Complexity Analysis) analyzes all of the webpage's text components to determine how complicated or problematic they are from an accessibility standpoint. It is able to assess a total of 12 web objects considering the textual components. The algorithmic evaluation results were arranged into six different aspects considering "Success Criteria", "Conformance Level", "Feedback", "Result", "Impairments Type" and "Improvement Direction". We represent the algorithmic evaluation result into six categories since in terms of accessibility support; most of the developed approaches [16, 22, 47] have no clear indication of the implemented techniques or success criteria and their conformance level. Therefore, users are unable to distinguish between accessibility features that have been implemented and those that are not covered by

the developed approach. Also, without indicating the conformance level of each success criterion introduces difficulties in understanding the importance of a particular success criterion. Thus, to increase the effectiveness of the evaluation report, we have also provided information on the conformance level with reference to the specific success criteria. After that, we offered feedback regarding the evaluation status of each success criterion. Furthermore, in the context of result categorization, almost all the developed tools categorize the accessibility guidelines in terms of several terminologies such as passed, failed, cannot tell, known error, likely error, potential error, error, warning, success, and not applicable. These terminologies sometimes denote an uncertain outcome. For example, failed, cannot tell, error, warning and not applicable terminologies are considered a negative result. It represents that the guideline is either not fulfilled or not identified or difficult to identify. It can also mean that the website has structural defects or programmatic errors in the evaluation tool. Also, for likely error terminology, it is not clear whether it is an identified error or not. So, from these uncertain categorizations, it is difficult to understand the concluded accessibility score that could lead to misleading accessibility representation. Therefore, we concisely categorized all the terminologies or assessment criteria into PASSED, FAILED, NOT DETECTED, and NOT TESTED, referring to each evaluated success criteria to calculate the accessibility score and to appropriately evaluate the accessibility status. We also provide information on each type of impairment related to each success criterion. This information indicates which group of individuals with specific needs these particular success criteria are important to ensure that the web content is accessible to them. In the last, as majority of the developed tools don't indicate which success criteria can be implemented automatically and which require manual or expert investigation as it is not possible to incorporate all the success criteria in an automated manner. Therefore, we offered textual improvement directions that show which criteria the tool successfully validated and which criteria need further verification or expert testing. It is recommended to do additional verification for FAILED success criteria and expert testing for NOT TESTED criteria.

Regarding the accessibility score for each disability type, none of the selected existing tools provide the percentage of accessibility for each disability type, such as visual impairment, cognitive impairment, etc. It may be challenging for web practitioners to comprehend which types of disabilities were prioritized during the development of that particular website and which disability type needs to be prioritized for future improvement. Providing accessibility percentages for each type of disability can give a better understanding of how accessible a certain website is for a specific group of people. Furthermore, an overall accessibility score with

accessibility status has also been provided. Besides, we summarized some arbitrary information that helps to understand some basic information about the tested webpage such as page URLs, page title, total number of checked HTML elements, page size or length, and page loading time. These statistics help a wide array of people (i.e., end users, designers, developers, practitioners, etc.) to understand the overall accessibility status of the tested webpage.

4 Proposed automated accessibility testing Framework validation

To validate the proposed framework, we performed a comparative analysis with similar existing models considering several functional properties, such as updated guidelines, user and expert criteria consideration, textual/non-textual component analysis, evaluated and not evaluation checkpoints feedback, overall accessibility score computation and accessibility score computation for different disability types. Regarding comparative assessment, we compared our proposed framework with the existing six models that have been mentioned in state-of-the-art literature in terms of several functionalities as mentioned in Table 9.

Table 9 depicts that for information regarding the updated version of accessibility guidelines, one model considered the updated guideline in their model (Boyalakuntla et al. [25]), and other models have no such concern. Concerning user and expert requirements and suggestions, none of the compared tools were found to address such concerns in their evaluation process. Besides, for textual and non-textual component analysis, none of the models were found with such concern in their evaluation process. For evaluated and not evaluated guideline information, one model (Pelzetter [22]) provides feedback about the evaluated and not evaluated checkpoints feedback. Even though other models did not compute the total accessibility score, two models, (Pelzetter [47] and Hilera et al. [1]) were found to be concerned about the overall accessibility score into account. Lastly, concentrating on disability types in the computation of the accessibility score, none of the models generate an accessibility score for each type of disability. Out of all these factors, the proposed framework takes into account every element addressed in Sect. 3 to enhance the accessibility assessment or evaluation outcome of a webpage. Therefore, our hypothesis is that the proposed framework, in comparison to other comparable models, can offer a comprehensive and up-to-date view of webpage accessibility.

Table 9 Comparative assessment results considering functional properties with existing models

References	Assessment Functional Properties						
Heading level	WCAG 2.2	User requirements consideration	Expert suggestion consideration	Textual/non-textual component analysis	Evaluated and not Evaluated Checkpoints feedback	Overall Accessibility Score	Accessibility Score for Disability type
Jens Pelzetter (2018) [47]	X (No)	X (No)	X (No)	X (No)	X (No)	✓(Yes)	X (No)
Jens Pelzetter, (2020) [22]	X (No)	X (No)	X (No)	X (No)	✓(Yes)	X (No)	X (No)
Michailidou et al. (2021) [16]	X (No)	X (No)	X (No)	X (No)	X (No)	X (No)	X (No)
Boyalakuntla et al. (2021) [25]	✓(Yes)	X (No)	X (No)	X (No)	X (No)	X (No)	X (No)
Hilera et al. (2016) [1]	X (No)	X (No)	X (No)	X (No)	X (No)	✓(Yes)	X (No)
Ingavélez-Guerra et al. (2018) [27]	X (No)	X (No)	X (No)	X (No)	X (No)	X (No)	✓(Yes)
Proposed Model	✓(Yes)	✓(Yes)	✓(Yes)	✓(Yes)	✓(Yes)	✓(Yes)	✓(Yes)

5 Discussion

Several recent analyses of website accessibility have found that the proportion of inaccessible websites is growing rapidly [7, 40], which negatively impacts people with disabilities in their access to digital resources. In recent years, this has drawn the attention of researchers interested in finding ways to improve this problem so that persons with impairments can gain better access options. Among several possible solutions, the most effective technique to demonstrate the limitations of a developed website is by automatically reviewing the website to determine its accessibility status in terms of several factors like accessibility score, accessibility concerns, etc. Therefore, in this research work, our prime aim is to focus on an automated accessibility testing framework in order to improve the effectiveness of accessibility testing. We aim to consider automated techniques as automated techniques are more effective than the hybrid approach when it comes to minimizing time and cost. For example, in a hybrid approach, user and expert testing have been incorporated where user testing involves a group of users accessing the website and testing accessibility criteria from their knowledge and experience. Besides, expert testing involves experts or professionals from accessibility domain who check manually all the conformance levels and guidelines. However, the user testing process requires some pre-evaluation setup which requires additional cost and time [48]. For example, user selection (in terms of several criteria like their ability, cognitive functionality, knowledge about accessibility criteria, available time, etc.); selection of required assistive technology (if it is applicable), testing tools or questionnaire preparation, setting up the testing environment (online or offline), prepare the guidelines or instruction about how to conduct the test or how to evaluate the website, etc. Also, in expert evaluation, to get the proper and acceptable outcome, potential expert selection with adequate knowledge and experience is a difficult task and sometimes they are not available for conducting the testing. Few experts might have domain-specific knowledge

such as accessibility for mobile applications or accessibility criteria for software only. Thus, finding experts who have adequate knowledge about the guidelines of the target domain (mobile/software/web platform) is challenging [49]. Other issues related to promptly auditing or verifying the evaluated result could increase the cost of the entire evaluation process [50]. For example, as the current web platform is changing dynamically and to keep this dynamic domain accessible, proper auditing is important and should be conducted routinely. Thus, organizing the hybrid testing itself is time-consuming, and costly. According to the scenario and our observation, we agree with Keselj et al. [51] in emphasizing the importance of automated testing for identifying the accessibility aspects of web platforms. Additionally, with the help of an automated system, it is feasible to evaluate many websites in a short time with a minimal cost which is nearly impossible with hybrid testing.

However, in order to address the current limitations of automated accessibility testing tools, our findings are aligned with appropriate guideline selection, user and expert suggestion incorporation, guideline modelling, semantic improvement, and proper accessibility result formulation. All these aspects have been addressed in the proposed accessibility aspects framework, shown in Fig. 4. To implement this framework, adequate knowledge of programming, sufficient knowledge about Web Content Accessibility Guidelines (WCAG), the user and expert suggestions, and guideline simplification were required. Besides, to implement this framework, we face challenges with the classification or distribution of the guidelines in terms of semantic aspects and non-semantic aspects. Therefore, we simplified the guidelines that can be found in [52], which presented complete information with simplified information of WCAG 2.2 addressing several aspects such as web objects, guidelines, attributes, conformance level, requirements, beneficial type (several disabilities), and evaluation phase (whether can be evaluate automated or require additional checking). Furthermore, implementing the algorithms for semantic and non-semantic aspects separately required

adequate knowledge of web programming and natural language processing techniques. Also, determining the proper visualization criteria focusing on every disability type and ensurement of appropriate assessment terms was challenging while we incorporated this framework which required additional concern and expertise on accessibility perspective in detail. However, the proposed framework has some limitations, such as the proposed model only considers WCAG that can be extended incorporating other guidelines. Besides, the proposed model considers fewer criteria from 2 experts and 5 user suggestions, which can be extended by considering other criteria by incorporating more user and expert involvement. By addressing these limitations, the effectiveness of the proposed model can be improved further.

6 Conclusion

In this study, we looked at earlier research in the area of developing accessibility evaluation tools to assess their strengths, limitations, and weaknesses. According to the observation of earlier research and by addressing the identifying issues, we proposed an accessibility evaluation framework considering several aspects from the coding to the visualization part which led web practitioners to understand the accessibility evaluation process and how it could facilitate to improvement of the accessibility evaluation result. We structured the proposed accessibility framework according to several aspects: guideline selection, user and expert suggestion consideration, guideline visualization, listing several website features that require special focus during tool development, and acceptable accessibility issue identification and visualization process; these represent the effectiveness of the proposed approach to facilitate the evaluation process and improve its effectiveness. The proposed framework has the potential to overcome the limitations of current accessibility evaluation tools. Additionally, none of the studies found in the literature addressed such aspects for web accessibility evaluation, which completely represent the novelty of the proposed framework. Moreover, the proposed framework is part of a web accessibility testing software development project. As an initial effort, we proposed the accessibility evaluation framework to address the potential accessibility criteria that could help to facilitate the evaluation process. Our next step or future work is aligned with the experimentation of the proposed framework in practical cases to evaluate and validate the accessibility of web pages.

7 Appendix

Table 10 Directions/guidelines/rules for each addressed element in the proposed framework

Objects	Related guidelines/Improvement directions
Images	(1.1.1) Non-text Content; (1.4.5) Images of Text [WCAG 2.2]
Pre-recorded/Live Audio	(1.2.1) Audio-only and Video-only; (1.2.3) Audio Description or Media Alternative (Prerecorded); (1.2.4) Captions (Live) [WCAG 2.2]
Pre-recorded/Live Video	(1.2.1) Audio-only and Video-only; (1.2.2) Captions (Prerecorded); (1.2.4) Captions (Live) [WCAG 2.2]
Links	(1.3.6) Identify Purpose; (1.4.1) Use of Color; (1.4.8) Visual Presentation; (2.4.4) Link Purpose (In Context); (2.4.9) Link Purpose (Link Only), (4.1.2) Name, Role [WCAG 2.2]
Words	(1.3.2) Meaningful Sequence; (2.4.1) Bypass Blocks; (3.1.3) Unusual Words [WCAG 2.2]
Sentences	(1.3.2) Meaningful Sequence; (2.4.1) Bypass Blocks; (3.1.3) Unusual Words [WCAG 2.2]
Paragraph	(1.3.2) Meaningful Sequence; (1.4.1) Use of Color; (2.4.1) Bypass Blocks; (3.1.3) Unusual Words [WCAG 2.2]
Button, Icons, Fields, Label	(1.3.3) Sensory Characteristics; (1.3.5) Identify Input Purpose (placeholder); (1.3.6) Identify Purpose (Buttons); (2.1.1) Keyboard; (2.1.4) Character Key Shortcuts; (3.3.2) Labels or Instructions [WCAG 2.2]
Text	(1.3.2) Meaningful Sequence; (2.4.1) Bypass Blocks; (3.1.3) Unusual Words [WCAG 2.2]
Title	(1.3.2) Meaningful Sequence; (1.4.12) Text Spacing; (2.4.1) Bypass Blocks; (2.4.2) Page Titled [WCAG 2.2]
Heading and Labels	(1.4.1) Use of Color; (2.4.6) Headings and Labels; (2.4.10) Section Headings [WCAG 2.2]
Language	(3.1.1) Language of Page [WCAG 2.2]
Idioms	(3.1.3) Unusual Words [WCAG 2.2]
Jargons	(3.1.3) Unusual Words [WCAG 2.2]
Pronunciation	(3.1.6) Pronunciation [WCAG 2.2]
Reading level	(3.1.5) Reading Level [WCAG 2.2]
Context-sensitive content	(3.1.3) Unusual Words [WCAG 2.2]
Drop-down menu, dialog box, checkbox, combo box	(3.2.1) On Focus; (3.2.2) On Input [WCAG 2.2]
Search field	(2.4.5) Multiple Ways; (3.2.3) Consistent Navigation; (3.2.4) Consistent Identification; (3.3.2) Labels or Instructions [WCAG 2.2]
Form	(3.2.5) Change on Request; (4.1.3) Status Messages [WCAG 2.2]
Error	(3.3.1) Error Identification; (3.3.3) Error Suggestion [WCAG 2.2]
Word length	(1.3.3) Sensory Characteristics; (1.3.4) Orientation [WCAG 2.2]
Sentence length	(1.3.3) Sensory Characteristics; (1.3.4) Orientation [WCAG 2.2]
Paragraph length	(1.3.3) Sensory Characteristics; (1.3.4) Orientation [WCAG 2.2]

Table 10 Directions/guidelines/rules for each addressed element in the proposed framework

Objects	Related guidelines/Improvement directions
Text content length	(1.3.3) Sensory Characteristics; (1.3.4) Orientation [WCAG 2.2]
User Information	No ‘Username’ and ‘Password’
CAPTCHA	(1.1.1) Non-text Content [WCAG 2.2]
Font size	(1.4.12) Text Spacing [WCAG 2.2]
Font family	Accessible font-family are ‘Tahoma’ or ‘Calibri’ or ‘Helvetica’ or ‘Arial’ or ‘Verdana’ or ‘Times New Roman’
Complexity of text pattern	No or or <i> or or <mark> or <sub> or <sup> elements/patterns
Website content type	Webpage content type should be the combination of paragraph, image and video content
Field, Button, Link	(1.3.1) Info and Relationships; (1.4.8) Visual Presentation; (2.4.5) Multiple Ways; (3.3.2) Labels or Instructions; (4.1.2) Name, Role, Value [WCAG 2.2]
Image, Logo	(1.3.3) Sensory Characteristics [WCAG 2.2]
Text in Images	(1.3.3) Sensory Characteristics [WCAG 2.2]
Maps, Images, Diagram, Data tables, Presentation, video	(1.4.8) Visual Presentation; (1.4.10) Reflow [WCAG 2.2]
Input size	(1.3.3) Sensory Characteristics; (1.3.5) Identify Input Purpose [WCAG 2.2]
Markup language elements	(1.3.6) Identify Purpose [WCAG 2.2]
Loading time	Webpage loading time should be ≤ 0.3 s
Page length	Webpage length should be ≤ 14 KB
Website availability	Webpage must be available or not down
Manual text size/ font adjustment	Webpage should have manual text and font size adjustment option
Manual Color adjustment	Webpage should have manual color adjustment option
Number of internal and external links	Webpage number of internal and external or hyperlinks must be ≤ 50
Number of image content	Webpage images number must be ≤ 10
Number of video content	Webpage number of audios should be ≥ 1 to ≤ 2 and videos should be ≥ 1 to ≤ 2
Accessible color pair	Webpage color pair should be accessible except [‘red’, ‘green’] and [‘red’, ‘black’] pair/combination

Acknowledgements The authors would like to thank all the authors for their contributions.

Author contributions Jinat Ara and Cecilia Sik-Lanyi conceptualized the study. Jinat Ara undertook the investigation, conducted the past literature analyses, managed the data, and developed the proposed framework with support from Cecilia Sik-Lanyi, and wrote the first draft version of the paper. Cecilia Sik-Lanyi and Arpad Kelemen reviewed the first draft version of the paper and provided their valuable direction and suggestions. According to their suggestions, Jinat Ara prepared the final version of the paper. Finally, Arpad Kelemen and Tibor Guzsvinecz reviewed the final version of the paper. All authors critically inputted their contributions, reviewed, and agreed on the fi-

nal version. All authors have read and agreed to the published version of the manuscript.

Funding The authors declare that this article is their research and was not financially supported.

Open access funding provided by University of Pannonia.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

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. Hilera, J.R., Otón, S., Martin-Amor, C., Timbi-Sisalima, C.: Towards a service-based architecture for web accessibility federated evaluation. In *9th International Conference on Advances in Computer-Human Interactions (ACHI'16)* (pp. 6–10). (2016), April
2. Giraud, S., Thérouanne, P., Steiner, D.D.: Web accessibility: Filtering redundant and irrelevant information improves website usability for blind users. *Int. J. Hum. Comput. Stud.* **111**, 23–35 (2018)
3. Acosta-Vargas, P., Salvador-Ullauri, L.A., Luján-Mora, S.: A heuristic method to evaluate web accessibility for users with low vision. *IEEE Access*, **7**, 125634–125648. (2019)
4. Li, L., Wang, C., Song, S., Yu, Z., Zhou, F., Bu, J.: A task assignment strategy for crowdsourcing-based web accessibility evaluation system. In *Proceedings of the 14th International Web for All Conference* (pp. 1–4). (2017), April
5. Ismail, A., Kuppusamy, K.S.: Web accessibility investigation and identification of major issues of higher education websites with statistical measures: A case study of college websites. *J. King Saud University-Computer Inform. Sci.* (2019)
6. Rashida, M., Islam, K., Kayes, A.S.M., Hammoudeh, M., Arefin, M.S., Habib, M.A.: Towards developing a framework to analyze the qualities of the university websites. *Computers.* **10**(5), 57 (2021)
7. Ara, J., Sik-Lanyi, C.: Investigation of COVID-19 Vaccine Information websites across Europe and Asia using automated accessibility protocols. *Int. J. Environ. Res. Public Health.* **19**(5), 2867 (2022)
8. Kim, H.K., Park, J.: Examination of the Protection offered by current accessibility acts and guidelines to people with disabilities in using Information Technology devices. *Electronics.* **9**(5), 742 (2020). <https://doi.org/10.3390/electronics9050742>

9. Manca, M., Palumbo, V., Paternò, F., Santoro, C.: The Transparency of Automatic Web Accessibility Evaluation Tools: Design Criteria, State of the Art, and User Perception. *ACM Transactions on Accessible Computing* (2022)
10. Mazalu, R., Cechich, A.: Web accessibility Assessment through Multi-agent Support for visually impaired users. *Int. J. Coop. Inform. Syst.* **29**, 03 (2020)
11. Alsaeedi, A.: Comparing web accessibility evaluation tools and evaluating the accessibility of webpages: Proposed frameworks. *Information*. **11**(1), 40 (2020)
12. Schiavone, A.G., Paternò, F.: An extensible environment for guideline-based accessibility evaluation of dynamic web applications. *Univ. Access Inf. Soc.* **14**(1), 111–132 (2015)
13. 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)
14. Lee-Geiller, S., Lee, T.D.: Using government websites to enhance democratic E-governance: A conceptual model for evaluation. *Government Inform. Q.* **36**(2), 208–225 (2019)
15. Bonacin, R., Reis, J.C.D., de Araujo, R.J.: An ontology-based framework for improving color vision deficiency accessibility. *Universal Access in the Information Society*, 1–26. (2021)
16. Michailidou, E., Eraslan, S., Yesilada, Y., Harper, S.: Automated prediction of visual complexity of web pages: Tools and evaluations. *Int. J. Hum. Comput. Stud.* **145**, 102523 (2021)
17. Robal, T., Marenkov, J., Kalja, A.: Ontology design for automatic evaluation of web user interface usability. In *2017 Portland International Conference on Management of Engineering and Technology (PICMET)* (pp. 1–8). IEEE, (2017), July
18. Kaur, N., Kumar, V.: Framework for Covering the Limitations of Web Accessibility Improvement Tools. *FP-International Journal of Computer Science Research (IJCSR)*, Volume 2, Issue 1, Pages 27–31, March (2015)
19. Duarte, C., Matos, I., Carriço, L.: Semantic content analysis supporting web accessibility evaluation. In *Proceedings of the 15th International Web for All Conference* (pp. 1–4). (2018), April
20. Ojha, P.K., Ismail, A., Srinivasan, K.K.: Perusal of readability with focus on web content understandability. *J. King Saud University-Computer Inform. Sci.* **33**(1), 1–10 (2021)
21. Broccia, G., Manca, M., Paternò, F., Pulina, F.: Flexible automatic support for web accessibility validation. *Proc. ACM Hum Comput Interact.* **4**(EICS), 1–24 (2020)
22. Pelzetter, J.: A declarative model for accessibility requirements. In *Proceedings of the 17th International Web for All Conference* (pp. 1–10). (2020), April
23. Moher, D.: The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Syst. Reviews.* **10**(1), 1–11 (2021)
24. Mohamad, Y., Velasco, C.A., Kaklanis, N., Tzovaras, D., Paternò, F.: A holistic decision support environment for web accessibility. In *International Conference on Computers Helping People with Special Needs* (pp. 3–7). Springer, Cham. (2018), July
25. Boyalakuntla, K., Venigalla, A.S.M., Chimalakonda, S.: WAccess—A Web Accessibility Tool based on the latest WCAG 2.2 guidelines. *arXiv preprint arXiv:2107.06799*. (2021)
26. Shrestha, R.: A Neural Network Model and Framework for an Automatic Evaluation of Image Descriptions based on NCAM Image Accessibility Guidelines. In *2021 4th Artificial Intelligence and Cloud Computing Conference* (pp. 68–73). (2021), December
27. Ingavelez-Guerra, P., Robles-Bykbaev, V., Oton, S., Vera-Rea, P., Galan-Men, J., Ulloa-Amaya, M., Hilera, J.R.: A proposal based on knowledge modeling and ontologies to support the accessibility evaluation process of learning objects. In *2018 Congreso Argentino de Ciencias de la Informática y Desarrollos de Investigación (CACIDI)* (pp. 1–5). IEEE, (2018), November
28. Song, S., Bu, J., Artmeier, A., Shi, K., Wang, Y., Yu, Z., Wang, C.: Crowdsourcing-based Web accessibility evaluation with golden maximum likelihood inference. *Proceedings of the ACM on Human-Computer Interaction*, 2 (CSCW), 1–21. (2018)
29. Li, L., Bu, J., Wang, C., Yu, Z., Wang, W., Wu, Y., Zhou, Q.: CrowdAE: A crowdsourcing system with human inspection quality enhancement for web accessibility evaluation. In: *International Conference on Computers Helping People with Special Needs* (pp. 27–30). Springer, Cham
30. Alahmadi, T.: A multi-method evaluation of university website accessibility: Foregrounding user-centred design, mining source code and using a quantitative metric. In *Proceedings of the 14th International Web for All Conference* (pp. 1–2). (2017), April
31. Song, S., Bu, J., Wang, Y., Yu, Z., Artmeier, A., Dai, L., Wang, C.: Web accessibility evaluation in a crowdsourcing-based system with expertise-based decision strategy. In *Proceedings of the 15th International Web for All Conference* (pp. 1–4). (2018), April
32. Hambley, A.: Empirical web accessibility evaluation for blind web users. *ACM SIGACCESS Accessibility Comput.* **12**9, 1–5 (2021)
33. 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. *Universal Access in the Information Society*, 20 (3), 595–615. (2021)
34. Hassan, A., El Abd, R.: Social Networks website accessibility for people with visual disability: An experimental study. *JRL Fac. Commer. Sci. Res.* **54**, 2 (2017)
35. Ara, J., Sik-Lanyi, C., Kelemen, A.: An Integrated Variable-Magnitude Approach for Accessibility Evaluation of Healthcare Institute Web Pages. *Appl. Sci.* **13**(2), 932 (2023)
36. Rodrigues, S.S., Scuracchio, P.E., de Fortes, M.: A support to evaluate web accessibility and usability issues for older adults. In *Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion* (pp. 97–103). (2018), June
37. Ara, J., Sik-Lanyi, C., Kelemen, A.: Accessibility engineering in web evaluation process: a systematic literature review. *Universal Access in the Information Society*, (pp. 1–34). (2023)
38. Eraslan, S., Yaneva, V., Yesilada, Y., Harper, S.: Web users with autism: Eye tracking evidence for differences. *Behav. Inform. Technol.* **38**(7), 678–700 (2019)
39. Ara, J., Hasan, M.T., Omar, A., A., Bhuiyan, H.: Understanding customer sentiment: Lexical analysis of restaurant reviews. In *2020 IEEE Region 10 Symposium (TENSYMP)* (pp. 295–299). IEEE, (2020), June
40. Karaim, N.A., Inal, Y.: Usability and accessibility evaluation of Libyan government websites. *Univ. Access Inf. Soc.* **18**(1), 207–216 (2019)
41. Leporini, B., Buzzi, M., Hersh, M.: Video conferencing tools: Comparative study of the experiences of screen reader users and the development of more inclusive design guidelines. *ACM Trans. Accessible Comput.* **16**(1), 1–36 (2023)
42. Text simplification guidelines: *Inclusive Design Research Centre, OCAD University, Toronto, Ontario, Canada (1998)*. <https://snow.idrc.ocadu.ca/accessible-media-and-documents/text-simplification-guidelines/>
43. Gaggi, O., Pederiva, V.: WCAG4All, a tool for making web accessibility rules accessible. In *2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)* (pp. 1–6). IEEE, (2021), January
44. Understanding the W3C's Accessibility Conformance Testing (ACT), Requirements: *Bureau of Internet Accessibility, East Greenwich, US*, (2023). <https://www.boia.org/blog/>

- [understanding-the-w3cs-accessibility-conformance-testing-act-requirements](#)
- 45. Mens, K., Kellens, A.: Mining Source Code for Design Regularities. *Working Session on Industrial Realities of Program Comprehension (IRPC 2008)*. (2008)
 - 46. Sharma, A., Revathi, M.: Automated API testing. In *2018 3rd International Conference on Inventive Computation Technologies (ICICT)* (pp. 788–791). IEEE. (2018), November
 - 47. Pelzetter, J.: Using Ontologies as a Foundation for Web Accessibility Tools. In *Proceedings of the 15th International Web for All Conference* (pp. 1–2). (2018), April
 - 48. Baazeem, I.S., Al-Khalifa, H.S.: Advancements in web accessibility evaluation methods: how far are we? In *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services* (pp. 1–5). (2015), December
 - 49. Tsaktsiras, K., Katsanos, C.: ESALP 2.0: Educational System to Support Learning of Web Content Accessibility Guidelines by Greek Web Practitioners. In *Proceedings of the 2nd International Conference of the ACM Greek SIGCHI Chapter* (pp. 1–6). (2023), September
 - 50. Shelly, C.C., Barta, M.: Application of traditional software testing methodologies to web accessibility. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)* (pp. 1–4). (2010), April
 - 51. Keselj, A., Milicevic, M., Zubrinic, K., Car, Z.: The Application of Deep Learning for the Evaluation of User Interfaces. *Sensors*, 22(23), 9336. (2022)
 - 52. Ara, J., Sik-Lanyi, C.: AccGuideLiner: Towards a Modelling Approach of Web Accessibility Requirements following WCAG 2.2. In *2023 IEEE International Conference on Smart Information Systems and Technologies (SIST)* (pp. 10–15). IEEE. (2023), May

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