Comparing Web Accessibility Evaluation Tools and Evaluating the Accessibility of Webpages: Proposed Frameworks

Abdullah Alsaeedi [ID]

Department of Computer Science, College of Computer Science and Engineering, Taibah University, Medina 42353, Saudi Arabia; aasaeedi@taibahu.edu.sa

Abstract: With the growth of e-services in the past two decades, the concept of web accessibility has been given attention to ensure that every individual can benefit from these services without any barriers. Web accessibility is considered one of the main factors that should be taken into consideration while developing webpages. Web Content Accessibility Guidelines 2.0 (WCAG 2.0) have been developed to guide web developers to ensure that web contents are accessible for all users, especially disabled users. Many automatic tools have been developed to check the compliance of websites with accessibility guidelines such as WCAG 2.0 and to help web developers and content creators with designing webpages without barriers for disabled people. Despite the popularity of accessibility evaluation tools in practice, there is no systematic way to compare the performance of web accessibility evaluators. This paper first presents two novel frameworks. The first one is proposed to compare the performance of web accessibility evaluation tools in detecting web accessibility issues based on WCAG 2.0. The second framework is utilized to evaluate webpages in meeting these guidelines. Six homepages of Saudi universities were chosen as case studies to substantiate the concept of the proposed frameworks. Furthermore, two popular web accessibility evaluators, Wave and SiteImprove, are selected to compare their performance. The outcomes of studies conducted using the first proposed framework showed that SiteImprove outperformed WAVE. According to the outcomes of the studies conducted, we can conclude that web administrators would benefit from the first framework in selecting an appropriate tool based on its performance to evaluate their websites based on accessibility criteria and guidelines. Moreover, the findings of the studies conducted using the second proposed framework showed that the homepage of Taibah University is more accessible than the homepages of other Saudi universities. Based on the findings of this study, the second framework can be used by web administrators and developers to measure the accessibility of their websites. This paper also discusses the most common accessibility issues reported by WAVE and SiteImprove.

Keywords: accessibility; government websites; accessibility evaluation; WCAG 2.0; web accessibility evaluation; web accessibility evaluation tools; automatic tools comparisons

1. Introduction

Web Content Accessibility Guidelines 2.0 (WCAG 2.0) [1] were developed to provide recommendations and guidance for creating accessible web content to meet the needs of different disabled users. Some countries have adapted these guidelines and used them as a law like section 508 in the United States [2]. The Saudi government established a program called "Yesser" to focus on digital transformation and the provision of e-services. One of the aspects that Yesser covers is creating accessible web content according to W3C guidelines [3].

The evaluation of a website in terms of web accessibility is a hard task [4]. In the literature, there are many approaches to evaluating the accessibility of webpages. One of the most common approaches is to examine the web accessibility of a webpage using automatic evaluation tools. These tools can also be called web accessibility checkers, and the terms are used interchangeably in this paper. Furthermore, with the development of WCAG 2.0, various automated web accessibility evaluation tools have been widely used to determine to what extent a specific webpage meets accessibility guidelines, especially WCAG 2.0. The widespread utilization of automated web accessibility checkers is due to the fact that WCAG 2.0 was designed to be more testable and measurable than WCAG 1.0. Many organizations rely on automated web accessibility evaluation tools as the main indicator of their accessibility level in the absence of expert evaluators [5].

Web accessibility checkers have been used to evaluate the web accessibility in different domains such as e-learning [6], e-Commerce [7] and banking [8]. Alshamari [7] evaluated the accessibility of three well-known Arab e-commerce websites using AChecker, TAW, Eval Access, MAUVE and FAE tools. Yakup and Kemal [9] evaluated the web accessibility of 25 official government websites in Turkey using AChecker, eXaminator, TAW, Total Validator, WAVE, Web Accessibility Assessment Tool, Eval Access, Cynthia Says, MAGENTA, HERA, Amp and Sort Site. Solomon and Ibrahim [10] used TAW and site analyser to report accessibility issues of Nigerian e-government websites. Basel and Faouzi [11] evaluated the homepages of 21 e-government websites using the TAW tool.

It is vital to evaluate websites using web accessibility evaluation tools, as they play an important role in assisting web developers with designing and developing more accessible websites. Despite the popularity of automated web accessibility evaluation tools in practice, few studies focus on comparing the performance and quality of web accessibility evaluation tools systematically. It is important to establish a systematic method of tool performance comparisons to help webmasters to select an appropriate evaluation tool for checking the compliance of their websites with guidelines. Similar to Brajnik [12], one of the main motivations of this paper is to help web developers compare web accessibility evaluation tools and select an appropriate tool. In order to establish a systematic method for comparing the performance of accessibility evaluation tools, this study proposed a framework to compare the performance of web-accessibility checkers. The comparison of tools' performances will be based on the idea of measuring how effective these tools are in detecting web accessibility issues compared to other tools, taking WCAG 2.0 guidelines into the consideration. Hence, a specific metric called a coverage error ratio (CER) metric is proposed.

Besides that, measuring the accessibility of given webpages compared to others based on WCAG 2.0 guidelines is necessary, as web developers require to measure whether the new version of webpages is more accessible than the old ones. For achieving this, another framework is proposed to evaluate how accessible webpages that meet these guidelines are. Therefore, a specific metric called the web accessibility accuracy (WAA) metric is proposed.

The organization of the paper is as follows. In Section 2, we present the background of WCAG 2.0. Section 3 describes the related works, including web accessibility evaluation approaches, the previous works that compare the performance of web accessibility tools, and relevant studies that evaluate the accessibility of websites in Saudi Arabia. Section 4 presents the proposed frameworks for comparing the performance of automatic web accessibility evaluation tools and measuring the accessibility of webpages. Section 5 describes the methodology that we followed in carrying out the study. Section 6 presents the results and discussion. Section 7 discusses the findings of the study and future works.

2. Background

WCAG 2.0 was developed to cover recommendations that make web content more accessible [1], taking into account various web technologies [1]. It comprises 12 guidelines related to four main web accessibility principles: perceivable, operable, understandable, and robust. Each guideline comprises various success criteria (SCs).

The guidelines are considered a framework that guides developers and webmasters aiming to make content easily accessible for disabled users. It is important to comply with these guidelines to allow elderly and disabled users to access contents without any barriers. However, it is difficult to measure and test whether the contents comply with the guidelines or not. Therefore, SC were proposed to be testable manually or automatically using web accessibility checkers. WCAG 2.0 [1] organized SC into levels of conformance: the minimum level of conformance (denoted by A) covering 25 SC, the intermediate level of conformance (denoted by AA) covering 13 SC as well as all the criteria in level A, and the highest level of conformance (AAA) covering all the criteria in level AA and 23 additional SC. In other words, each of the SC belongs to a level of conformance. A webpage is said to meet a specific level of conformance if it meets all the SC in that level and the preceding level.

The next two subsections describes two of well-known accessibility tools that will be used as case study to proof the concept of the proposed framework.

2.1. WAVE

WAVE is an automatic tool developed by WebAIM that allow users to enter the web address of a current site. It aims to help web developers check the accessibility of a given webpage to make it more accessible [13]. It adds icons to a webpage that allow users and experts to check potential accessibility issues. Red icons refer to accessibility errors, yellow icons indicate alerts, green icons indicate accessibility features, and all the light blue icons indicate structural, semantic, or navigational elements.

2.2. SiteImprove

SiteImprove is an online service that allows webmasters to check the web accessibility of a webpage with respect to WCAG 2.0. The browser extension of SiteImprove can be activated to automatically analyze the webpages for accessibility violations regarding the A, AA, or AAA level of the WCAG standards. It allow users to choose the conformance level, either A, AA, or AAA, and distributes the reported errors into different responsibilities. Each reported accessibility error is associated with a direct link to the corresponding WCAG manual to obtain a more detailed explanation of the reasons for the existence of errors. Furthermore, the reports produced by SiteImprove comes with suggestions to fix accessibility errors to gain compliance with WCAG 2.0. Similar to WAVE, SiteImprove has the ability to highlight the location of errors on the site itself and to point out the snippet source code in the browser's developer tools.

3. Related Works

3.1. Web Accessibility Evaluation Approaches

Basically, there are four distinct approaches that are widely used to evaluate websites' accessibility. The first approach is the automatic approach that runs accessibility evaluation tools on the website to gather accessibility violations against predefined guidelines. Web accessibility evaluation tools (web accessibility checkers) can be defined as software programs that help web administrators determine whether a website meets web accessibility guidelines [14]. There is a list of web accessibility tools [15]. These tools can be categorized into two groups [16]: general and specific tools. The general tools are those that evaluate almost all guidelines, such as TAW 3.0, WAVE, SiteImprove, and AChecker. The specific tools are those that evaluate specific web accessibility aspects, such as Contrast Checker that assesses the color contrast. A number of disadvantages related to these tools have been highlighted in the literature. One of the most common failings of web accessibility evaluation tools is the difficulty in interpreting results [17]. Many studies, such as those conducted by Abanumy et al. [18], Rana et al. [19], Al-Khalifa et al. [20], Alahmadi and Drew [21], and Alshamari [7], employed automatic tools to check the accessibility of websites.

The second approach is the manual evaluation using human experts to examine webpages to identify violations of accessibility guidelines. This approach is introduced to mitigate limitations of automated web accessibility evaluation tools as they cannot determine conformance to all accessibility guidelines. One of these limitations is that some guidelines are subjective and therefore human experts are required to examine websites against such subjective guidelines. Brajnik et al. [22] demonstrated the ability of expert users to detect accessibility violations with more reliability. However, Hong et al. [23] conducted a study to compare web accessibility evaluation using human experts and automated software separately. The outcomes of their study showed that automated software tools have the ability to discover more accessibility errors than human experts. Furthermore, web accessibility evaluation using experts may include bias while finding accessibility barriers and violations [24]. Despite the effectiveness of this approach compared to the automated evaluation in detecting comprehensive accessibility violations, it is time consuming [25]. Moreover, relaying on experts' evaluation is more very costly and effective than user testing, such as testing with the aid of disabled users [25].

The third approach is testing with the aid of users such as disabled users. It is also called user testing to identify accessibility issues while disabled users are interacting with the content of webpages [26]. An example of a study in which disabled users were asked to evaluate websites' accessibility was conducted by Petrie and Kheir [27]. Six blind and six sighted people were involved in assessing the accessibility of two websites, and empirical data were collected through usability testing. This approach is effective since the reliance on software tools or human experts may fail to detect issues with the implementation of accessibility features. It is considered an ultimate approach for evaluating the accessibility of webpages [28]. Despite the effectiveness of this approach, recruiting disabled people is a very difficult task [29].

The fourth approach is a hybrid approach that combines automated and manual evaluation (human experts or with the aid of disabled people). An example of this approach was utilized by Kumar and Owston [30] to evaluate the accessibility of e-learning technologies using automated tools and students with a learning disability. Abdul Latif and Masrek [31] recommended combining automatic tools and disabled users to detect accessibility violations. Basel and Faouzi [11] recommended including expert users, disabled users, automated tools, webmasters and web developers to assess e-government websites. Al-Khalifa [32,33] used the WAVE checker toolbar alongside the manual evaluation of 36 Saudi Arabian e-government websites to detect the most common accessibility errors. Al-Khalifa [34] used the combination of an automatic tool, manual evaluation using experts, user evaluation, and surveys for web administrators and developers. Al-Khalifa [34] stated that user testing is the most precise method to evaluate the accessibility of websites, but it requires time and experienced testers, causing this type of evaluation to be more challengeable. Khan and Buragga [35] checked the accessibility of the websites of Saudi Railways and Saudi Post using non-experienced evaluators and automatic tools, namely Eval Access 2.0 and Cynthia Says tools. The main finding of their study is that the manual inspection of accessibility attained a very similar results to those obtained using the online tools. An interesting study was conducted by Alotaibi [36] that combined manual evaluation, automatic evaluation, and testing with the aid of disabled users to evaluate the accessibility of the Blackboard e-learning system. It is vital to highlight that manual evaluation by accessibility experts or with the aid of disabled users is a time consuming and complex task [11].

The performance and suitability of accessibility evaluation tools can be assessed and compared in two ways [5]: by selecting a representative sample of websites or using test suites. The latter contain a number of tests to assess tools with respect to a specific SC. Tools will be considered efficient based on several metrics, such as correctness, coverage, and completeness. The former way focuses on selecting real websites that contain known violations is another approach to assess the ability of specific evaluation tools to detect these violations. In this paper, we selected real webpages to compare the performance of web accessibility evaluation tools and to measure their accessibility using specific proposed metrics.

3.2. Comparing the Performance of Web Accessibility Evaluation Tools

Brajnik [12] proposed a method to compare a pair of tools based on measuring the correctness, specificity, and completeness with respect to the accessibility guidelines. The term "false positive" is introduced to denote the number of reported accessibility issues that are not true while the term "false negatives" denotes the number of true accessibility problems that are not detected by a given tool. The completeness aimed to count the number of accessibility violations that are detected by the tool and correctly reported to the users or web developers. The completeness computed the effectiveness of a tool in minimizing true negatives. It is difficult to characterize this practically as it requires all true accessibility issues in advance. The correctness is the proportion of accessibility issues that are reported that are true. The specificity of a tool is the number of potential accessibility problems such as warning sand suggestions that can described by the tool after detecting them.

To measure and compare the effectiveness of accessibility checkers using the Brajnik's [12] method, real websites were chosen as case studies. Well-known accessible websites were chosen to stress the tools to generate false positives. A number of inaccessible websites were also selected to measure the ability of the tools to avoid generating false negatives. Issues were classified manually as false positives or false negatives for both tools. The human inspectors classify an issue as false positive for any tool if the issue is irrelevant or wrongly reported. An issue that is reported by a tool A and classified as false positive will be used as a reference for a tool B, so the issue will be classified as false negative if the tool B fails to detect the issue. Despite the novelty of Brajnik's method [12], it is limited as it only compares a pair of tools and issues are classified manually.

Vigo et al. [5] focused on three metrics: coverage, completeness, and correctness. The following variables were considered while computing coverage, completeness, and correctness. True positives are actual problems found by the tool. False positives are mistakenly flagged accessibility issues. False negatives are the issues that the tool did not catch and are therefore missed. Coverage computes the number of criteria that are violated at least once, completeness measures the proportion of true violations to the total number of violations reported by user experts, and correctness measures the effectiveness of tools in reducing the number of wrongly determined accessibility violations. The effectiveness of six web accessibility evaluation tools were measured in terms of coverage, completeness, and correctness. In terms of coverage, the TAW tool was the best tool and it covered only 50% of the SC. The TAW tool showed superiority among other tools in terms of completeness and achieved a value of 38%. In terms of correctness, Deque attained the highest correctness score of 96%.

3.3. Metrics for Web Accessibility Evaluation

The first attempt to quantitatively measuring the web accessibility was proposed by Sullivan and Matson [37]. In their study, eight checkpoints of Priority 1 from WCAG 1.0 were selected to test websites. The failure rate (FR) measure the ratio of actual accessibility errors to the potential accessibility errors for a given webpage (p). The total number of accessibility errors is denoted as B_p and the numbers of potential accessibility errors is denoted as N_p.

$$\text{failure rate } (Fr_p) = \frac{B_p}{N_p} \tag{1}$$

Parmanto [38] developed a metric based on WCAG checkpoints to automatically test the accessibility using automated accessibility evaluation tools. Besides, the metric is proposed to satisfy several requirements. The first requirement is that measuring the accessibility using a quantitative score to represent the range of accessibility from perfectly accessible to completely inaccessible. The quantitative score would help web developers to assess the changes in terms of accessibility made overtime and to compare between websites. The second requirement is that the metric score would allow measuring the rate of change of web accessibility over time. The third requirement is the fairness of metric by taking the size of websites into the account, and one may use several webpages

to ensure having webpages with different sizes and complexities. V denotes the total violation of a webpage

$$WAB_{score} = \frac{\sum\limits_{p=1}^{NP}\sum\limits_{j=1}^{V}\left(\frac{B_{pj}}{P_{pj}}\right)(W_i)}{NP} \tag{2}$$

A high score of WAB_{score} denotes more accessibility barriers for disable users, while lower score of WAB_{score} means a fewer accessibility barriers for disable users. That is, a lower score of WAB_{score} means a website takes more accessibility criteria into accounts. The total webpages of a website is denoted by NP. The B_{pj} denotes the number of violations and the P_{pj} denote the number potential violations. W_i denoted weight of violations in inverse proportion to WCAG priority level.

Another metric was proposed in Unified Web Evaluation Methodology (UWEM) project [39]. It has been developed to ensure the accessibility evaluation are compatible with W3C/WAI accessibility guidelines. The metric computes the probability of detecting barriers that preventing users from accomplishing a task. The metric takes potential errors into the consideration.

$$UWEM = \frac{1}{2}\sum_{j=1}^{n}1 - \prod_{b}\left(1 - \frac{B_{pj}}{P_{pj}}W_b\right) \tag{3}$$

Buhler et al. [40] proposed a modification on the UWEM metric in order to take complexity and scalability properties into the account while evaluating web accessibility of given websites. The C_{pb} was proposed for computing the complexity of p with respect to a barrier type b. The idea behind aggregating the computation of C_{pb} is to include the ratio of potential and actual barriers and the ratio of all failures to the number of failures for one barrier. For a disability group u, the S_{ub} represents the severity of a barrier type b. The disabled users should be involved to rate the S_{ub} for each barrier type.

$$A_3(p,u) = 1 - \prod_{b}(1 - S_{ub})^{(C_{pb})} \tag{4}$$

$$C_{pb} = \frac{B_{pb}}{N_{pb}} + \frac{B_{pb}}{B_p} \tag{5}$$

Song et al. [41] proposed a metric called the Reliability Aware Web Accessibility Experience Metric (RA-WAEM), where disable users are involved in sharing their experience for assessing the severity of accessibility barriers. An evaluation has been performed in [41] on a collected dataset and showed that RA-WAEM performed better than the state-of-the art metrics in reflecting the user experience of disabled people while evaluating the web accessibility. The computation of the RA-WAEM metric begins by selecting properties of websites. After that, the selected websites were evaluated using an evaluation system to form the pass rate matrix based on number of checkpoints (m). The accessibility score q_i of the website is computed as follow as shown in Equation (6). The checkpoints is denoted as m. The $n \times m$ pass rate matrix P is obtained with respect to the number of n websites and m checkpoints. The pass rate $P_{(i,j)}$ of a checkpoint j for a given website i is the number of webpages that pass this checkpoint divided by the number of webpages which contains possible violations corresponding to this checkpoint. Additionally, the checkpoint weights $w = (w_1, w_2, w_3 \cdots w_m)^T$

$$q_i = \sum_{j=1}^{m} P_{i,j}w_j \tag{6}$$

3.4. Accessibility Evaluation of Saudi Government Websites

In 2005, Abanumy et al. [18] used the Bobby evaluation tool to evaluate the accessibility of government websites of Saudi Arabia and Oman. The conducted accessibility evaluation showed that none of the websites conform to all priority 1 checkpoints. The authors claimed that both countries

require substantial efforts to meet W3C WCAG and greater awareness of the importance of accessibility. Furthermore, they suggested that, the accessibility policies in these countries should be reviewed to have accessible e-government websites.

Al-Faries et al. [42] investigated the accessibility of top e-government services in Saudi Arabia with respect to WCAG 2.0. Four evaluators were recruited to identify violations for each guideline per principle. For the perceivable principle, the most commonly violated guideline was guideline 1.1 relates to ensuring that a text alternative is provided for all non-text content. To a lesser extent, guideline 1.3 and guideline 1.4 were also violated. For the operable principle, the most frequently violated guideline was guideline 2.1 that was introduced to ensure that all functionalities are accessible from a keyboard. More precisely, both SC 2.1.1 and SC 2.1.2 tend to be the rarely satisfied SC in guideline 2.1. One of the most commonly violated guidelines was guideline 2.4 that was introduced to help users find contents, navigate, and determine where they are. In terms of the understandable principle, guideline 3.1 was also violated, where the most commonly violated SC was SC 3.1.6. Moreover, SC 3.2.5 was the SC with the highest violation in guideline 3.2. For the robust principle, SC 4.1.1 and SC 4.1.2 were violated in 85% and 70% of all services, respectively. According to Al-Faries et al. [42], the robust principle was considered as the most common violated principle in the top e-government services in Saudi Arabia. The authors highly recommended that web developers should follow accessibility guidelines to ensure that e-government services are accessible for all users, especially disabled users.

Al-Khalifa [33] evaluated the accessibility of 36 government websites in Saudi Arabia according to WCAG 2.0. The homepages of these websites were selected to be manually evaluated for all accessibility conformance levels with the aid of the WAVE checker toolbar. The failed SC and the number of violations were recorded. In terms of the failed SC in level A, the most commonly violated SC for guideline 1.1 was SC 1.1.1 that relates to ensuring the existence of an appropriate text alternative for all non-text content. With regard to guideline 1.3, SC 1.3.1 and SC 1.3.2 were the most commonly violated SC. SC 2.4.1 and SC 4.1.1 were violated widely with respect to the 2.4 and 4.1 guidelines, respectively. In terms of the failed SC in level AA conformance, SC 3.1.2, SC 1.4.5, and SC 1.4.4 were violated widely. The most common failed SCs in level AAA conformance were SC 1.4.6, SC 1.4.9, SC 2.4.9, and SC 2.4.10.

Mukhtar et al. [19] evaluated the accessibility of 21 Saudi Arabian government universities. A total validator tool was used to evaluate the compliance with accessibility standards. Among these 21 websites, two university websites passed WCAG 1.0 criteria and none of them met WCAG 2.0 criteria. The study showed that alternative text for images and buttons were the most frequent failures. Moreover, the author stated that 16 university websites had many accessibility failures due to missing the alt attribute in the image tag. To sum up, the study showed that 80% of websites did not meet the level A accessibility level. The authors stated that web developers and designers lacked awareness of the importance of website accessibility standards. They also analyzed the functional accessibility of Saudi university websites in terms of the following aspects: navigation and orientation, text equivalent, styling, and HTML slandered. They computed the average error for each aspect and obtained averages of 24.30%, 29.15%, 38.02%, and 8.53% for navigation and orientation, text equivalent, styling, and HTML slandered, respectively.

Uthman et al. [43] evaluated the web accessibility of the LMS Blackboard at King Saud University. The study was based on a questionnaire that was prepared to evaluate the ease of use, design user interface, navigational features, and accessibility of the contents. The study showed that Blackboard is usable and accessible by teachers in terms of delivering course contents. The authors recommended increasing the accessibility and usability by offering courses in Arabic and English [43].

## 4. The Proposed Frameworks for Comparing Tool and Web Accessibility

The next sections present two frameworks. The first is proposed to compare the performance of automatic web accessibility tools and the second is introduced to evaluate various webpages in terms of their accessibility.

4.1. Study 1: Framework for Comparing the Performance of Web Accessibility Tools

In this section, a framework for comparing the performance of web accessibility tools is proposed. It relies on collecting accessibility errors using a number of tools. In this study, a web accessibility error denotes a contradiction that may violate one or more WCAG 2.0 criteria. Let us assume that there are a number of tools to measure web accessibility for a given webpage. Therefore, with respect to various accessibility tools, it can be said that a specific tool is considered the one that performs the best if it has the ability to detect more web accessibility errors accurately. To achieve this, a coverage error ratio (CER) metric is proposed to be computed for each tool and a given webpage. The performance of web accessibility tools can be measured by computing a CER score for each tool. By comparing the attained CER scores for tools, the tool with the highest CER score can be considered the one with the best performance.

$$CER = \frac{Number\ of\ Errors\ detected\ by\ a\ given\ tool\ (t)}{The\ total\ number\ of\ Errors\ detected\ by\ all\ tools} \qquad (7)$$

Figure 1 illustrates the general framework for comparing the performance of web accessibility tools with respect to WCAG 2.0 criteria for a given webpage. In general, this framework relies on a number of webpages and different web accessibility checkers. Therefore, it begins by providing numbers of webpages and various accessibility tools. Subsequently, for the given webpage, the web accessibility is evaluated using tools such as Wave, Achecker, SiteImprove, and so on. Once the webpages have been evaluated using the tools, accessibility errors are collected for each web accessibility checker for each webpage separately.
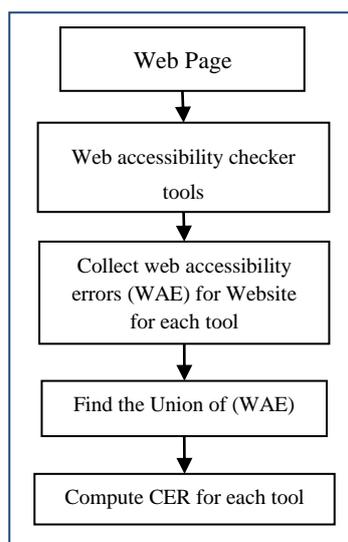
Figure 1. Evaluation framework for tools comparison.

The union set of errors is gathered manually by analyzing the reported errors obtained from various tools. The aim of this stage is to collect distinctive errors without redundancies. This phase requires non-expert users to map errors generated by different tools. For instance, SiteImprove showed this error message "Image link has no alternative text", which corresponds to the following message "Linked image missing alternative text" generated by WAVE. This mapping process is required to collect the union errors generated by different tools. The union set of errors will be used as a reference to measure the effectiveness of each tool in finding these distinctive errors. Subsequently, the CER scores are computed, as described in Equation (1). This ratio represents the proportion of errors detected by the given tool divided by the union set of errors detected by all the tools.

Equation (7) can be rewritten using the following equation, where $de_{(t_x,w)}$ denotes the number of accessibility errors detected using the tool t_x for a webpage w and $\bigcup_{t \in T} de_{(t,w)}$ denotes the number

of errors in the union set of errors detected by all the web accessibility evaluation tools for the given website w.

$$CER(t_x, w) = \frac{de_{(t_x, w)}}{\bigcup_{t \in T} de_{(t,w)}} \tag{8}$$

The procedures of the comparison process to assess the performance of various tools is described in Algorithm 1. The comparison starts by selecting a number of web accessibility tools and webpages, as shown in lines (1) and (2). Then, an iteration over the webpages set is performed to collect accessibility errors using various tools. Following that, the Errors Map is initialized to record the detected errors via tools as illustrated in line (4). The UnionErrors set is defined to store the union set of errors detected by all tools for a given webpage. The next step is to iterate over tools to collect accessibility errors. The CollectErrors (t,w) function is responsible for collecting accessibility errors given the tool (t) and webpage (w). The tool and the detected errors will be recorded as a pair (t, detectedErrors) with the Errors as shown in line 9. Subsequently, the union errors are updated for the current webpage until there is no tool to select.

Algorithm 1: Computing CER score for each tool.

Input : *WebPages*, *Tools*

```
/* WebPages are a list of webpages under evaluation and a set of web accessibility checker
   tools Tools                                                                            */
```

Result: *Scores* Map CER scores for tools to each webpage

Data: *Errors* Maps detected errors to tools,*CERScore* is coverage error ratio for each tool

$Tools \leftarrow \{tool_1, tool_2, \cdots, tool_n\}$;

$WebPages \leftarrow \{Webpage_1, Webpage_1, \cdots, Webpage_m\}$;

for $w \in WebPages$ **do**

    $Errors \leftarrow \{\}$;

    $UnionErrors \leftarrow \{\}$;

    for $t \in Tools$ **do**

        $DetectedErrors \leftarrow \{\}$;

        $DetectedErrors \leftarrow CollectErrors\ (t, w)$;

        $Errors \leftarrow Errors \cup\ (t, DetectedErrors)$;

        $UnionErrors \leftarrow \{UnionErrors \cup DetectedErrors\ \}$;

    end

    for $t \in Tools$ **do**

        $CERScore \leftarrow ComputeCER\ (Errors(t), UnionErrors)\ Scores \leftarrow Scores \cup\ (w, (t, CERScore))$;

    end

end

return *Scores*

Once the union errors have been collected, another iteration over the tool set is performed to compute the CER score for each tool and the current webpage under analysis, as illustrated in lines (12)–(14). Moreover, Errors (t) is a function that returns a list of errors detected by a tool t. The computeCER (Errors(t), UnionErrors) function computes the CER score based on errors detected by a given tool and the UnionErrors set. Then, a pair of (t, CERScore) is mapped to the current webpage (w) as illustrated in line (14). The comparison process is carried out for the next webpage until no further webpages can be selected to compute the CER scores.

Our proposed CER metric in this paper is focused on false negatives and benefits from all issues reported by the tools to select the reference issues stored in the union set. The set of reference issues contains all error reported by all tools. Furthermore, all reported issues are assumed to be true and the effectiveness of the tool will be measured based on its ability to detect more issues from the reference list.

4.2. Study 2: Framework for Evaluating Webpages in Terms of Web Accessibility

In this section, a framework for evaluating various webpages in terms of web accessibility is proposed. The aim of this framework is to establish a systematic approach to measure the accessibility level of webpages. Moreover, it can be considered as a performance indicator to compare different webpages in terms of accessibility. For instance, the Ministry of Education can utilize this approach to determine which universities' webpages meet WCAG 2.0 criteria. It can be used to compare two versions of webpages for the same sites. The key benefit of this approach is that is helps the web developers decide whether the new version of a webpage is more accessible than the old version.

To determine the most accessible webpage among a number of webpages, web designers can rely on a number of evaluation tools to measure the web accessibility of webpages. Let us assume that there are a number of webpages of which the accessibility needs to be measured with respect to WCAG 2.0 and there are various accessibility checkers to detect all violations, we can say that a webpage h is the most accessible webpage in considering WCAG 2.0 compared to other webpages if it violated fewer of those guidelines. In other words, a webpage with fewer accessibility violations than other webpages is the most accessible webpage in terms of web accessibility.

To achieve this, the web accessibility accuracy (WAA) metric is proposed to compute it for each tool and a given webpage, as shown in Equation (9). The accuracy of web accessibility can be measured by computing a WAA score for each webpage. By comparing the attained WAA scores, the webpage with the highest WAA score can be considered the most accessible one. In other words, a webpage with the highest WAA score among a list of webpages is the most accessible one according to WCAG 2.0.

$$WAA = 1 - \frac{Number\ of\ Errors\ in\ the\ UnionErrors\ set\ for\ the\ given\ webpage}{Number\ of\ Errors\ in\ the\ UnionErrors\ set\ for\ all\ webpages} \tag{9}$$

Figure 2 shows the framework for evaluating the web accessibility for any webpage with respect to WCAG 2.0. In general, the evaluation relies on gathering web accessibility errors that violated WCAG 2.0 criteria based on multiple known web accessibility checkers. Then, a WAA metric is proposed to evaluate the web accessibility of a given webpage. Similar to the previous comparison framework described in Section 4.1, the proposed framework in this section relies on webpages and different web accessibility checkers. Therefore, it begins by providing various webpages and various accessibility tools. It aims to measure and compare the web accessibility of webpages. First, the web accessibility for the given webpage is measured using tools such as Wave, SiteImprove, and so on. Second, for each accessibility checker tool, violations with respect to WCAG 2.0 are gathered. Third, the union of accessibility errors detected by various tools is collected separately for each webpage. The aim of this step is to find distinctive errors (a reference set of accessibility issues) collected without any redundancies using different tools for the webpage under assessment. The union set of errors will be used to assess the accessibility of the webpage with respect to WCAG 2.0. subsequently, the WAA is computed, as described in Equation (9). This ratio represents the proportions of union errors detected by all tools for the current webpage divided by the total number of errors in all union sets for all webpages involved in the comparison process.

Equation (9) can be rewritten using the following equation, where $\bigcup_{t \in T} de_{(t, w_y)}$ denotes the number of errors in the union set of errors detected by all web accessibility evaluation tools for the given website W_y and $\Sigma_{w \in W} \bigcup_{t \in T} de_{(t, w)}$ denotes the number of errors in all union sets of errors detected by all web accessibility evaluation tools for all webpages W.

$$WAA(t, w_y) = 1 - \left(\frac{\bigcup_{t \in T} de_{(t, w_y)}}{\Sigma_{w \in W} \bigcup_{t \in T} de_{(t, w)}} \right) \tag{10}$$
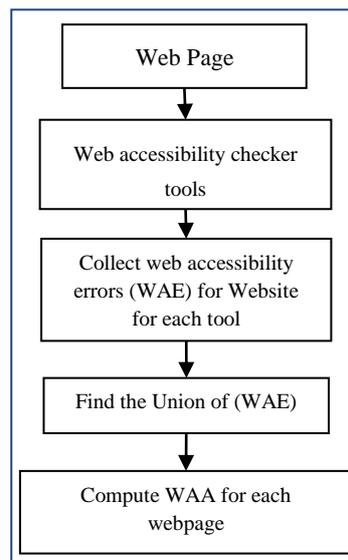
Figure 2. Evaluation framework for webpage accessibility comparison.

The computation of web accessibility accuracies of various webpages is presented in Algorithm 2. The evaluation framework is provided with distinctive webpages and different web accessibility checkers, as shown in lines (1) and (2). The aim of entering both pices of information is to measure the web accessibility of the webpages. Then, for each webpage, the union of web accessibility errors is collected for all tools, as illustrated in lines (6)–(10). Subsequently, the AllUnionErrors Map is updated by recording the collection of union errors for each webpage.

Algorithm 2: Computing WAA score for each webpage.

Input : *WebPages*, *Tools*
/* *WebPages* are a list of webpages under evaluation and a set of web accessibility checker
   tools *Tools*   */
Data: *WAA* score is web accessibility accuracy for each webpage
Result: *WAAScores* Map *WAA* score to each webpage
$Tools \leftarrow \{tool_1, tool_2, \cdots, tool_n\}$;
$WebPages \leftarrow \{Webpage_1, Webpage_1, \cdots, Webpage_m\}$;
$AllUnionErrors \leftarrow \{\}$;

for $w \in WebPages$ **do**
    $CurrentUnionErrors \leftarrow \{\}$;
    for $t \in Tools$ **do**
        $DetectedErrors \leftarrow \{\}$;
        $DetectedErrors \leftarrow CollectErrors\ (t, w)$;
        $CurrentUnionErrors \leftarrow \{CurrentUnionErrors \cup DetectedErrors\ \}$;
    end
    $AllUnionErrors \leftarrow \{AllUnionErrors \cup (w, CurrentUnionErrors)\ \}$;
end
for $w \in WebPages$ **do**
    $WAA \leftarrow ComputeWAA\ (AllUnionErrors\ (w), GetAllUnionErrors\ (AllUnionErrors))$;
    $WAAScores \leftarrow WAAScores \cup (w, WAA)$;
end
return WAAScores

The second iteration shown in lines (13)–(16) aims to compute the WAA score, as described in Equation (10). The AllUnionErrors(w) function returns the set of union errors detected by all tools for the webpage of interest. Furthermore, the GetAllUnionErrors(AllUnionErrors) function is responsible for returning all the union errors for all webpages. The computeWAA function computes the WAA

score by dividing the number of errors in the union set for the homepage under assessment by the total number of errors in all the union sets for all the webpages. Then, a pair (w, WAA) is added to the WAAScores Map to store the WAA score for the current homepage under assessment as illustrated in line (15).

5. Experimental Methodology

This section presents the methodology of our experiments. The conducted experiments are the proof of concept for the proposed frameworks. Therefore, six homepages of Saudi public universities were selected in our experiments as case studies. We followed the methodology used by Al-Khalifa [32,33], Alshamari [7] and Rana et al. [19] to evaluate only the homepages. The reason behind selecting the homepages only is that they are indicators for other webpages and the starting points for visitors. Moreover, two distinctive tools were selected to measure their performance using Algorithm 1 described in Section 4.2. We also evaluate the web accessibility of the six homepages using Algorithm 2.

The maturity level indicator of the electronic transformation of government core services is computed by Yesser for all Saudi public institutions. According to the service maturity indicators report generated by Yesser, the education and research sector in KSA has been classified into three categories according to their performance in providing e-services. The green (excellent) category includes universities and research centers that the performance ratio ranged from 85% to 100%. The yellow (average) category consists of educational institutions of which the ratio varied between 60% and 84%. The red (poor) category includes institutions of which the performance ratio ranged from 0% to 59%. Table 1 summarizes the maturity level indicator for various university websites. We selected these six universities in our experiments as they are distinctive samples from each category defined by Yesser. In other words, we have relied on the maturity level indicator to select these six homepages. This indicator works as key performance indicator and has been computed and publicly published by Yesser to measure the maturity level of electronic transformation of services provided by Saudi universities. In this way, we selected two homepages from each categories (excellence–average–poor).

Table 1. Maturity level indicator for e-services provided by various universities.

University	Maturity Level Indicator	Category
Taibah	12.82	red
Najran	37.21	red
King Fahad	62.16	Yellow
King Khaled	69.77	Yellow
Prince Sattam	85.71	Green
King Saud	93.75	Green

Moreover, two web accessibility checkers were selected, namely SiteImprove and Wave, in the conducted experiments. The reason for selecting these checkers is that they are free, open source, and descriptive tools where accessibility issues are described alongside with relevant violated SCs. Both tools allow evaluators to navigate accessibility issues on the webpages and the source codes as well.

6. Experimental Results and Discussion

6.1. Study 1: Result of Comparing Web-Accessibility Tools

The bar plots of the CER scores computed for all the university homepages are shown in Figure 3. It is apparent from Figure 3 that SiteImprove outperformed Wave in five homepages. However, Wave detected more accessibility errors than SiteImprove in the Taibah University homepage. This is due to the fact that the Wave checker detected ten empty links, and these were not discovered by SiteImprove.
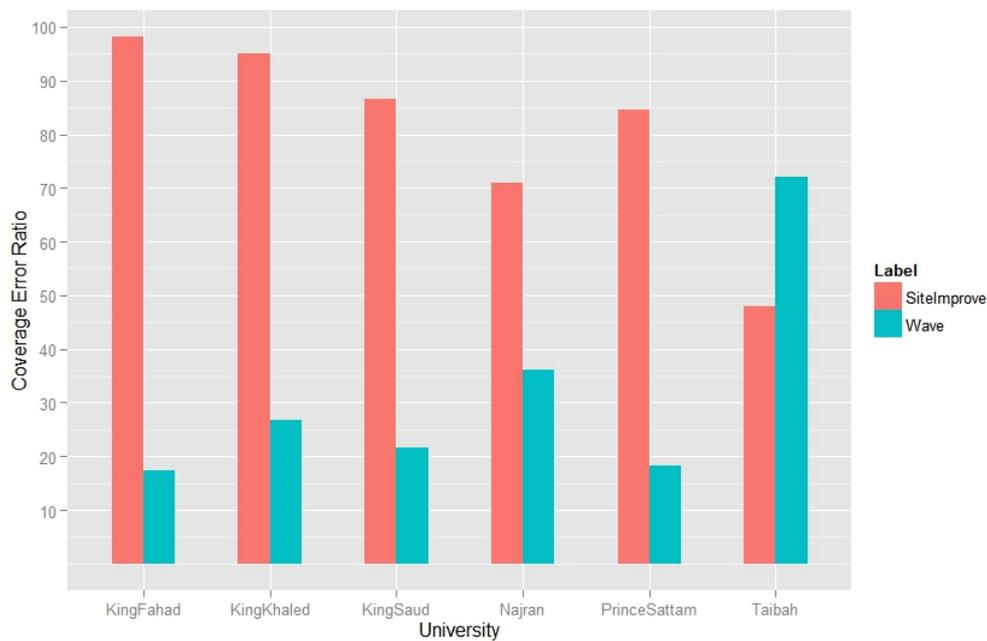
Figure 3. Coverage error ratio computed for six universities' homepages.

6.2. Study 2: Result of Evaluating Webpages in Terms of Web Accessibility

The bar plots illustrated in Figure 4 summarize the WAA scores attained for each homepage. From this figure, it is obvious that the homepage of Taibah University is more accessible than other homepages as its WAA score is 96.16%, which is the highest score. This means that the homepage of Taibah University violates fewer guidelines in WCAG 2.0 than other homepages. This is an interesting outcome as Taibah University was classified in Yesser in the red (poor) category in providing e-services. This indicates that the IT center is aware of WCAG 2.0. Conversely, the homepages of Prince Sattam University and King Saud University attained lower WAA scores than Taibah University even though they belong to the green (excellent) category in Yesser.
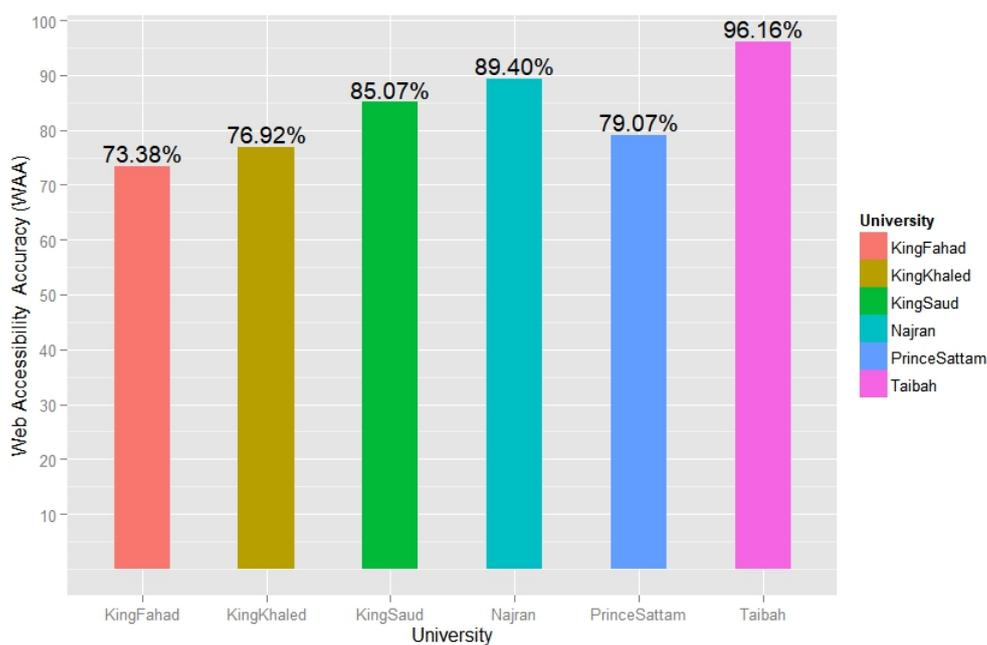

Figure 4. Web accessibility accuracy computed for six universities homepages.

6.3. Accessibility Issues in Saudi Universities Homepages

Table 2 summarizes the accessibility errors detected by WAVE and SiteImprove for the Taibah University homepage. The corresponding SC for the reported accessibility issues are shown in Table 2. One of the main advantages of WAVE and SiteImprove is that they describe accessibility errors with the related WCAG 2.0 criteria and provide guidelines to fix errors. It is clear that ten links do not contain descriptive texts, causing difficulties in describing different links for disabled users who use a screen reader, Braille, or text. SiteImprove detected two non-distinguishable links, which means the same link text is used for multiple links navigating to different destinations on the current webpage.

Table 2. Accessibility errors detected for Taibah University.

University	Wave			SiteImprove		
	Errors Message	Occurrences No.	Corresp. Criteria	Errors Message	Occurrences No.	Corresp. Criteria
	Empty button	3	1.1.1 2.4.4	Not Detected	0	-
	Not Detected	0	-	Image with no alt attribute	2	1.1.1
Taibah University	Linked image missing alternative text	2	1.1.1 2.4.4	Image link has no alternative text	2	2.4.4
	Empty link	10	2.4.4	Not Detected	0	-
	Not Detected	0	-	Non-distinguishable links	2	2.4.4
	Not Detected	0	-	There is no top level heading (H1) on the page	1	1.3.1
	Not Detected	0	-	Naming generic landmarks	2	1.3.1
	Missing form label	3	1.1.1 1.3.1 2.4.6 3.3.2	Input field is missing a description	3	1.3.1 3.3.2 4.1.2

Table 3 shows the summary accessibility errors obtained for the Prince Sattam University homepage. It is obvious that SiteImprove detected more errors than WAVE. The number of elements that are not highlighted on focus is 87. Errors of this type cause a difficulty for keyboard users to highlight focused elements in a webpage, which aim to tell users where they are on the page. It is vital to mention that WAVE failed to detect errors of this type. Similar to the homepage of Taibah University, the homepage of Prince Sattam University has 21 empty links. WAVE describes this type of error as "A link contains no text". One of the main findings in the accessibility issues on the Prince Sattam University homepage is that multiple links should be combined. These errors occur for adjacent links pointing to the same destination in the case where one has a textual hyperlink and the other is associated with an iconic representation of the same link.

Table 4 reports the accessibility errors detected by WAVE and SiteImprove for the King Khaled University homepage. Fifteen empty heading errors were detected by both tools, which means there are 15 heading tags, but the text is empty. However, both tools expressed these errors using different warning messages. These errors violate three criteria, 1.3.1, 2.4.1, and 2.4.6, according to the SiteImprove checker. The number of elements that are not highlighted on focus in the homepage of King Khaled University is 96. Furthermore, 11 images do not have an alt attribute. It is noted that each accessibility checker tool describes the errors in their own way. Moreover, there is a difference between the two web accessibility checker tools in the way of describing the violated criteria related to each error.

Table 3. Accessibility errors detected for Prince Sattam University.

University	Wave			SiteImprove		
	Errors Message	Occurrences No.	Corresp. Criteria	Errors Message	Occurrences No.	Corresp. Criteria
Prince Sattam University	Linked image missing alternative text	1	1.1.1 2.4.4	Image link has no alternative text	1	2.4.4
	Empty link	21	2.4.4	Not Detected	0	-
	Broken ARIA reference	3	1.3.1	Incorrect reference for 'aria-labelledby'	3	1.1.1 1.3.1 2.4.4 3.3.2 4.1.2
	Not Detected	0	-	Links should be combined	8	1.1.1
	Not Detected	0	-	Link refers to a non-existing element	1	1.3.1
	Not Detected	0	-	Use unique identification for elements	4	1.3.1 4.1.1
	Not Detected	0	-	Non-distinguishable links	11	2.4.4
	Not Detected	0	-	Element not highlighted on focus	87	2.4.7

Table 4. Accessibility errors detected for King Khaled University.

University	Wave			SiteImprove		
	Errors Message	Occurrences No.	Corresp. Criteria	Errors Message	Occurrences No.	corresp. Criteria
King khaled University	Missing alternative text	9	1.1.1	Image with no alt attribute	11	1.1.1
	Linked image missing alternative text	8	1.1.1 2.4.4	Image link has no alternative text	8	2.4.4
	Empty heading	15	1.3.1 2.4.6 2.4.1	The heading is missing text	15	1.3.1
	Empty button	4	1.1.1 2.4.4	Not Detected	0	-
	Empty link	4	2.4.4	Not Detected	0	-
	Not Detected	0	-	Presentational attributes used	1	1.3.1
	Not Detected	0	-	There is no top level heading (H1) on the page	1	1.3.1
	Not Detected	0	-	Element not highlighted on focus	96	2.4.7
	Not Detected	0	-	Non-distinguishable links	10	2.4.4

Table 5 summarizes the accessibility errors detected by WAVE and SiteImprove for the King Saud University homepage. It is apparent that 13 links do not contain text and are considered by Wave as accessibility errors. The conducted experiments show that this type of error is common in all the homepages except the homepage of King Fahad University. As Table 5 shows, the number of elements that are not highlighted on focus is 57, and these are considered as accessibility errors that are detected by the SiteImprove checker.

The accessibility errors determined by WAVE and SiteImprove for the homepage of King Fahad University are illustrated in Table 6. It is clear that 15 images do not have the correct alternative text. Similar to other universities in the conducted study, the number of elements that are not highlighted on focus is 81. The number of images with no alt attribute detected by SiteImprove is 24, whereas WAVE detected only four. There are ten accessibility issues related to use of presentational attributes, in which

attributes such as 'border' and 'align', are used in the HTML tags and these attributes should be used CSS for these attributes.

Table 5. Accessibility errors detected for King Saud University.

University	Wave			SiteImprove		
	Errors Message	**Occurrences No.**	**Corresp. Criteria**	**Errors Message**	**Occurrences No.**	**Corresp. Criteria**
	Linked image missing alternative text	6	1.1.1 2.4.4	Image link has no alternative text	6	2.4.4
	Missing alternative text	1	1.1.1	Image with no alt attribute	7	1.1.1
	Empty heading	1	1.3.1 2.4.6 2.4.1	The heading is missing text	1	1.3.1
King Saud University	Empty link	13	2.4.4	Not Detected	0	-
	Not Detected	0	-	Non-distinguishable links	9	2.4.4
	Not Detected	0	-	'Select box' without a description	1	1.3.1
	Not Detected	0	-	Presentational attributes used	1	1.3.1
	Not Detected	0	-	There is no top level heading (H1) on the page	1	1.3.1
	Not Detected	0	-	The iFrame is missing a description	1	1.3.1
	Not Detected	0	-	Element not highlighted on focus	57	2.4.7

Table 6. Accessibility errors detected for King Fahad University.

University	Wave			SiteImprove		
	Errors Message	**Occurrences No.**	**Corresp. Criteria**	**Errors Message**	**Occurrences No.**	**Corresp. Criteria**
	Missing alternative text	4	1.1.1	Image with no alt attribute	4	1.1.1
	Linked image missing alternative text	22	1.1.1	Image link has no alternative text	22	2.4.4
	Missing form label	1	3.3.2	Input field is missing a description	1	1.3.1
King Fahad University	Empty heading	1	1.3.1 2.4.6 2.4.1	Not Detected	-	-
	Empty link	1	2.4.4	Not Detected	-	-
	Empty table header	1	1.3.1	Not Detected	-	-
	Not Detected	0	-	Image with no alt attribute	20	1.1.1
	Not Detected	0	-	The image does not have the correct alternative	15	1.1.1
	Not Detected	0	-	Presentational attributes used	10	1.3.1
	Not Detected	0	-	Non-distinguishable links	11	2.4.4
	Not Detected	0	-	Element not highlighted on focus	81	2.4.7

Table 7 summarizes the accessibility errors detected by WAVE and SiteImprove for the Najran University homepage. It is clear that ten links do not contain text, as reported only by WAVE. As can be seen in Table 7, a number of text hyperlinks are not distinguishable (used the same link text) as they are pointing to different destinations. There are nine instances of select box without a descriptive title, and this should be fixed to allow users utilizing assistive technologies to know what the select box menu is for.

Table 7. Accessibility errors detected for Najran University.

University	Wave			SiteImprove		
	Errors Message	Occurrences No.	Corresp. Criteria	Errors Message	Occurrences No.	Corresp. Criteria
Najran university	Missing alternative text	1	1.1.1	Image with no alt attribute	1	1.1.1
	Linked image missing alternative text	3	1.1.1	Image link has no alternative text	3	2.4.4
	Missing form label	10	2.4.4	Not Detected	0	-
	Empty heading	1	1.3.1 2.4.6 2.4.1	Heading is missing text	1	1.3.1
	Empty link	10	2.4.4	Not Detected	0	-
	not detected	0	-	Alternative text for image is identical to link text	6	1.1.1
	not detected	0	-	Image link is missing alternative text	1	2.4.4
	not detected	0	-	Link text used for multiple different destinations	11	2.4.4
	not detected	0	-	Select box has no description	9	3.3.2
	not detected	0	-	Element ID is not unique	2	4.1.1
	not detected	0	-	Missing button in form	2	3.3.2
	not detected	0	-	HTML is used to format content	3	1.3.1
	not detected	0	-	Input field has no description	1	1.3.1
	not detected	0	-	Generic landmarks are not named	4	1.3.1
	not detected	0	-	iFrame is missing a description	1	4.1.2
	not detected	0	-	Invalid WAI-ARIA role, state or property	2	4.1.2
	not detected	0	-	Adjacent links used for same destination	2	1.1.1

6.4. General Finding

Among the six university homepages evaluated, all failed to add alternative text for image links. The homepage of King Fahad University has 22 image links without alternative texts, reaching a higher number for this type of failure than other homepages that were included in the study. In 2010, Al-Khalifa [32] stated that one of the three most common accessibility errors encountered in governmental homepages was that they did not add a text alternative for non-text elements. This may be attributed to the web developers' lack of knowledge of the importance of alternative texts for images [19].

Furthermore, all the university homepages had accessibility issues in that they did not add text for links describing the functionality or the target of links. The homepage of Prince Sattam University has 21 empty links, more than any other homepage. The importance of adding descriptive text for links is to aid people using a screen reader, Braille, or a text browser to distinguish different links [19]. Moreover, all the university homepages failed to distinguish between links in the same webpage, as the same link texts are used.

Four university homepages failed to meet SC 2.4.7 that focuses on highlighting the components while the user uses keyboard navigation. It is important to mention that WAVE is not able to detect this type of accessibility issue. The result of accessibility evaluation showed that the homepages of Prince Sattam University, King Khalid University, King Saud University, and King Fahad University had 87,

96, 57, and 81 issues, respectively related to elements that are not highlighted on focus (no keyboard accessibility). Various studies [32,33,42] showed that this type of accessibility issue is considered as one of the most common accessibility violations found in Saudi e-government websites. We recommend that web developers should ensure that elements receiving keyboard focus are highlighted on focus.

It is vital to find the common accessibility issues with respect to the four main principles. With regard to the operable principle, Al-Faries et al. [42] stated that the most common violated guideline is 2.4. Furthermore, the conducted experiment showed that guideline 2.4 is the most common violated guideline, especially the 2.4.4 criterion. With respect to the understandable principle, Al-Faries et al. [42] showed that the 3.2.5 criterion is the most common violated criterion. However, in the conducted experiment, this criterion is violated less than the 3.3.2 criterion. Regarding the robust accessibility principle, guideline 4.1 is intended to support compatibility with assistive technologies such as screen readers. The finding of the conducted experiment showed that the 4.1.2 criterion tend to be violated less than other criteria. According to Al-Khalifa [32,33], the major violations include the following: no text alternatives, no keyboard access, and no language identification. These findings are similar to those of our study. Moreover, the conducted study showed that other violations such as empty link and empty heading are very common in all the homepages.

7. Conclusions and Future Works

This study set out to propose novel frameworks in terms of tool comparison and webpage accessibility evaluation. WAVE and SiteImprove were selected as they are well-known tools and utilized to substantiate the concepts of the proposed frameworks. CER and WAA metrics were proposed as measurements for both frameworks. The CER metric was proposed to measure the capability of tools in detecting accessibility issues. The CER scores demonstrate the capability of SiteImprove compared to WAVE in detecting web accessibility issues. One of the main advantages of the tool comparison framework is the ability to compare more than two tools by implementing the same steps and utilizing the CER equation to compare tools' performance. We recommend the webmasters and developers use multiple efficient web-accessibility tools in order to detect a variety of accessibility barriers for disabled users. In this context, selecting the most efficient tools can be performed relying upon CER scores that can be computed for multiple accessibility tools.

The WAA metric is proposed as an indicator of the accessibility level for a given webpage. In this study, we use the WAA metric to compare six homepages of Saudi universities to determine which homepage is most accessible with respect to WCAG 2.0. One can employ this metric to compare two versions of webpages for the same site. The key benefit of this approach is that is helps the web developers decide whether the new version of the webpage is more accessible than the old version. Section 6.4 summarized a general finding based on analysing accessibility issues reported by both WAVE and SiteImprove tools. Based on that, the majority of homepages have accessbility issues related to empty links, empty heading, image links without alternative texts, and missing alt attribute for images.

One future direction is to compare the proposed WAA metric with other relevant metrics reported in Section 3.3. An interesting future extension of the study would be creating online databases that contain all possible web accessibility violations and the corresponding error messages generated by all tools for each potential violation. This could be updated by the developers of web accessibility checkers. Once such a database exists, we could train various classifiers such as support vector machines to use all possible errors and their target categories such as perceivable, operable, understandable, and robust. Thus, the classifier will be able to categorize all accessibility issues without any human intervention. Furthermore, this classifier could be integrated into the proposed frameworks presented in this study to make them fully automated in computing the CER and WAA metrics.

Moreover, one could apply artificial intelligence and machine learning methods to help the IT manager assign each web accessibility violation to the person responsible (webmaster, editor,

developer) for it to be fixed. This may include crawling distinct web accessibility tools to provide the optimal solution for any web accessibility error.

References

1. Caldwell, B.; Cooper, M.; Reid, L.G.V.; Erheiden, G. Web Content Accessibility Guidelines (WCAG) 2.0. 2008. WWW Consortium (W3C). Available online: https://www.w3.org/TR/WCAG20/ (accessed on 20 August 2019).

2. United States Lawsoverview of the Rehabilitation Act of 1973 (Sections 504 And 508). Available online: https://webaim.org/articles/laws/usa/rehab (accessed on 15 September 2019).

3. National Government Website Standard. Available online: https://www.yesser.gov.sa/EN/Methodologies/NATIONAL_ICT_DIGITALSTANDARDS_PRINCIPLES/Pages/National-Government-Website-Standard.aspx (accessed on 13 September 2019).

4. Brewer, J. Web accessibility highlights and trends. In Proceedings of the 2004 International Cross-disciplinary Workshop on Web Accessibility (W4A '04), New York, NY, USA, 17–22 May 2004; pp. 51–55. doi:10.1145/990657.990667. [CrossRef]

5. Vigo, M.; Brown, J.; Conway, V. Benchmarking web accessibility evaluation tools: measuring the harm of sole reliance on automated tests. In Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility, Rio de Janeiro, Brazil, 13–15 May 2013; ACM: New York, NY, USA, 2013; p. 1.

6. Lundqvist, S.; Ström, J. Web Accessibility in E-Learning: Identifying And Solving Accessibility Issues for Wcag 2.0 Conformance in an E-Learning Application. Student Thesis, 2018. Available online: http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-148913 (accessed on 25 September 2018).

7. Alshamari, M. Accessibility evaluation of arabic e-commerce web sites using automated tools. *J. Softw. Eng. Appl.* **2016**, *9*, 439–451. [CrossRef]

8. Kaur, A.; Dani, D. Banking websites in india: An accessibility evaluation. *CSI Trans. ICT* **2014**, *2*, 23–34. doi:10.1007/s40012-014-0040-x. [CrossRef]

9. Akgül, Y.; Vatansever, K. Web accessibility evaluation of government websites for people with disabilities in turkey. *J. Adv. Manag. Sci.* **2016**, *4*, 201–210. [CrossRef]

10. Adepoju, S.A.; Shehu, I.S.; Bake, P. Accessibility evaluation and performance analysis of e-government websites in nigeria. *J. Adv. Inf. Technol.* **2016**, *7*. [CrossRef]

11. AlMourad, B.; Kamoun, F. Accessibility evaluation of dubai e-government websites: Findings and implications. *J. -Gov. Stud. Best Pract.* **2013**, *2013*, 978647.

12. Brajnik, G. Comparing accessibility evaluation tools: A method for tool effectiveness. *Univers. Access Inf. Soc.* **2004**, *3*, 252–263. [CrossRef]

13. Martín, A.; Cechich, A.; Rossi, G. Comparing Approaches to Web Accessibility Assessment. In *Handbook of Research on Web Information Systems Quality*; IGI Global: Hershey, PA, USA, 2008; pp. 181–205.

14. Shadi Abou-Zahra, N.S.; Keen, L. Selecting Web Accessibility Evaluation Tools. Available online: https://www.w3.org/WAI/test-evaluate/tools/selecting/ (accessed on 5 May 2018).

15. Eggert, E.; Abou-Zahra, S. Web Accessibility Evaluation Tools List. Available online: https://www.w3.org/WAI/ER/tools/index (accessed on 5 July 2018).

16. Al-Khalifa, H.S. Wcag 2.0 semi-automatic accessibility evaluation system: Design and implementation. *Comput. Inf. Sci.* **2012**, *5*, 73. [CrossRef]

17. Abuaddous, H.Y.; Jali, M.Z.; Basir, N. Web accessibility challenges. *Int. J. Adv. Comput. Sci. Appl.* **2016**, *7*, 172–181. [CrossRef]

18. Abanumy, A.; Al-Badi, A.; Mayhew, P. e-government website accessibility: In-depth evaluation of Saudi Arabia and oman. *Electron. J. Gov.* **2005**, *3*, 99–106.

19. MasoodRana, M.; Fakrudeen, M.; Rana, U. Evaluating web accessibility of university web sites in the kingdom of Saudi Arabia. *Int. J. Technol. Knowl. Soc.* **2011**, *7*, 1–15.

20. Al-Khalifa, H.S.; Al-Kanhal, M.; Al-Nafisah, H.; Al-soukaih, N.; Al-hussain, E.; Al-onzi, M. A pilot study for evaluating arabic websites using automated wcag 2.0 evaluation tools. In Proceedings of the 2011 International Conference on Innovations in Information Technology (IIT 2011), Abu Dhabi, United Arab , 25–27 April 2011; pp. 293–296.

21. Alahmadi, T.; Drew, S. An evaluation of the accessibility of top-ranking university websites: Accessibility rates from 2005 to 2015. In Proceedings of the DEANZ Biennial Conference, Hamilton, New Zealand, 17–20 April 2016; pp. 224–233.

22. Brajnik, G.; Yesilada, Y.; Harper, S. The expertise effect on web accessibility evaluation methods. *Hum.–Comput. Interact.* **2011**, *26*, 246–283.

23. Hong, S.; Katerattanakul, P.; Lee, D. Evaluating government website accessibility: Software tool vs. human experts. *Manag. Res. News* **2007**, *31*, 27–40. [CrossRef]

24. Elkabani, I.; Hamandi, L.; Zantout, R.; Mansi, S. Toward better web accessibility. In Proceedings of the 2015 5th International Conference on Information & Communication Technology and Accessibility (ICTA), Marrakech, Morocco, 21–23 December 2015; pp. 1–6. doi:10.1109/ICTA.2015.7426915. [CrossRef]

25. Grantham, J.; Grantham, E.; Powers, D. Website accessibility: An australian view. In Proceedings of the Thirteenth Australasian User Interface Conference (AUIC '12), Melbourne, Australia, 31 January–3 February 2012; Volume 126, pp. 21–28. Available online: http://dl.acm.org/citation.cfm?id=2512125.2512129 (accessed on 9 March 2019).

26. Alayed, A.S. A Framework and Checklist for Localised Web Content Accessibility Guidelines for Arabic University Websites in Saudi Arabia. Ph.D. Thesis, University of Southampton, Southampton, UK, 2018.

27. Petrie, H.; Kheir, O. The relationship between accessibility and usability of websites. In Proceedings of the SIGCHI Conference on Human Factors In Computing Systems, San Jose, CA, USA, 28 April–3 May 2007; ACM: New York, NY, USA, 2007; pp. 397–406.

28. Petrie, H.; Bevan, N. The evaluation of accessibility, usability, and user experience. *Univers. Access Handb.* **2009**, *1*, 1–16.

29. Petrie, H.; Hamilton, F.; King, N.; Pavan, P. Remote usability evaluations with disabled people. In Proceedings of the SIGCHI Conference on Human Factors In Computing Systems, Montreal, QC, Canada, 22–27 April 2006; ACM: New York, NY, USA, 2006; pp. 1133–1141.

30. Kumar, K.L.; Owston, R. Evaluating e-learning accessibility by automated and student-centered methods. *Educ. Technol. Res. Dev.* **2016**, *64*, 263–283. doi:10.1007/s11423-015-9413-6. [CrossRef]

31. Latif, M.H.A.; Masrek, M.N. Accessibility evaluation on malaysian e-government websites. *J. E-Gov. Stud. Best Pract.* **2010**, *2010*, 935272.

32. Al-Khalifa, H.S. Exploring the accessibility of Saudi Arabia e-government websites: A preliminary results. In Proceedings of the 4th International Conference on Theory and Practice of Electronic Governance (ICEGOV '10), Cairo, Egypt, 1–4 December 2008; ACM: New York, NY, USA, 2010; pp. 274–278. doi:10.1145/1930321.1930378. [CrossRef]

33. Al-Khalifa, H.S. The accessibility of Saudi Arabia government web sites: An exploratory study. *Univers. Access Inf. Soc.* **2012**, *11*, 201–210. doi:10.1007/s10209-010-0215-7. [CrossRef]

34. Al-Khalifa, H.S.; Baazeem, I.; Alamer, R. Revisiting the accessibility of Saudi Arabia government websites. *Univers. Access Inf. Soc.* **2017**, *16*, 1027–1039. doi:10.1007/s10209-016-0495-7. [CrossRef]

35. Khan, M.A.; Buragga, K.A. Effectiveness of accessibility and usability of government websites in Saudi Arabia. *Can. J. Pure Appl. Sci.* **2010**, *4*, 1127–1131.

36. Alotaibi, S.J. Evaluating the blackboard system based on web accessibility and usability guidelines. *Int. J. Digit. Soc.* **2015**, *6*, 1094–1101. [CrossRef]

37. Sullivan, T.; Matson, R. Barriers to use: Usability and content accessibility on the web's most popular sites. In Proceedings of the 2000 Conference on Universal Usability (CUU '00)y, Arlington, VA, USA, 16–17 November 2000; ACM: New York, NY, USA, 2000; pp. 139–144. doi:10.1145/355460.355549. [CrossRef]

38. Parmanto, B.; Zeng, X. Metric for web accessibility evaluation. *J. Am. Soc. Inf. Sci. Technol.* **2005**, *56*, 1394–1404. doi:10.1002/asi.20233. [CrossRef]

39. Cluster W. Unified Web Evaluation Methodology (UWEM 1.0), 2006. In Proceedings of the 11th International Conference (ICCHP 2008), Linz, Austria, 9–11 July 2008; pp. 394–401.

40. Bühler, C.; Heck, H.; Perlick, O.; Nietzio, A.; Ulltveit-Moe, N. Interpreting results from large scale automatic evaluation of web accessibility. In *Computers Helping People with Special Needs*; Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 184–191.

41. Song, S.; Bu, J.; Shen, C.; Artmeier, A.; Yu, Z.; Zhou, Q. Reliability aware web accessibility experience metric. In Proceedings of the Internet of Accessible Things (W4A '18), Lyon, France, 23–25 April 2018; ACM: New York, NY, USA, 2018; pp. 24:1–24:4. doi:10.1145/3192714.3192836. [CrossRef]

42. Al-Faries, A.; Al-Khalifa, H.S.; Al-Razgan, M.S.; Al-Duwais, M. Evaluating the accessibility and usability of top saudi e-government services. In Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance (ICEGOV '13), Seoul, Korea, 22–25 October 2013; ACM: New York, NY, USA, 2013; pp. 60–63. doi:10.1145/2591888.2591898. [CrossRef]

43. Alturki, U.T.; Aldraiweesh, A. Evaluating the usability and accessibility of lms "blackboard" at king saud university. *Contemp. Issues Educ. Res.* **2016**, *9*, 33–44. [CrossRef]