QUANTITATIVE METRIC FOR RANKING WEB ACCESSIBILITY BARRIERS BASED ON THEIR SEVERITY

Hayfa Yousef Abuaddous, **Mohd Zalisham Jali & Nurlida Basir**
Faculty of Science &Technology,
Universiti Sains Islam Malaysia, Malaysia

haddose@yahoo.com; zalisham@usim.edu.my; nurlida@usim.edu.my

ABSTRACT

Web accessibility aims at providing disabled users with a barrier-free user experience so they can use and contribute to the Web more effectively. However, not all websites comply with WCAG 2.0 which results in Web accessibility barriers in websites. Thus, assistive technologies such as screen readers would not be able to interpret the presented contents on the monitor due to these barriers and this will contribute to making websites inaccessible to disabled users. This paper proposed an innovative metric that assigns measurable weight to each identified barrier based on its severity and impacts on the accessibility level, and then ranks the barriers accordingly. Following, Web developers can fix the highly ranked severe barriers instead of wasting time in studying and fixing less severe types of barriers that may rarely occur. An experiment was conducted to check the metric validity. We found the metric was valid and thereby we suggested the usage of the metric as a valid scientific measurement.

Keywords: Web accessibility barrier, WCAG 2.0, WABS metric, automated tools, A-Checker, disabled users.

INTRODUCTION

The web accessibility discipline struggles to make the Web a more comfortable environment to disabled people. According to the World Wide Web Consortium (W3C, 2005), web accessibility enables people with disabilities to use the Internet to perform a variety of tasks such as online purchasing and browsing. When accessibility is guaranteed, people with disabilities will have equal

access and equal opportunity to use and contribute to the Web more effectively. In spite of the importance of making websites accessible to users with special needs, the majority of websites being developed are still not fully accessible (Hashemian, 2011; Nahon, Benbasat, & Grange, 2012; Abuaddous, Jali, & Basir, 2013; Ahmi & Mohamad, 2016). Most websites endure accessibility barriers that make it hard for disabled people to use them. A barrier can be defined as any condition caused by the website that hinders a user's progress towards accessing a website resulting in a failure mode. Since this paper considered checking against the WCAG 2.0 guidelines, failing to meet a success criterion in WCAG 2.0 would result in a barrier. Web accessibility is about eliminating accessibility barriers so disabled users can use the Web. In 2008, the WAI developed Web Content Accessibility Guidelines 2 (WCAG 2.0). WCAG 2.0 is organized around four design principles that provide the foundation for web accessibility (perceivable, operable, understandable and robust) (WCAG 2.0, 2008). Each principle has guidelines and each guideline has testable success criteria (SCs) at levels A, AA, or AAA, with 'A' being the **"minimum standard"** and the most important one which a website must meet to be accessible. SCs are the basis of determining the conformance of a level in WCAG 2.0.

Web inaccessibility equates to barring disabled users from accessing and navigating the Internet. For example, users suffering from visual impairments may not be able to access most websites even with the use of screen readers. The problem stems from Web developers' noncompliant ways of coding that do not adhere to accessibility guidelines due to challenges facing Web developers during the website's design and development (Abuaddous, Jali, & Basir, 2016). This will produce inequities and barriers for assistive technologies. Thus, assistive technologies such as screen readers would not be able to interpret the presented contents on the monitor which will contribute to making websites inaccessible to disabled users.

Although the WAI of the W3C has published online guidelines, many IT professionals are unaware of them (Brown & Scott, 2015). Even if they had heard about the guidelines many are not motivated to follow them (Villena, Ramos, Fortes, & Goulartea, 2014). According to (Shin, Lim, Lee, & Kyung, 2013) only few Web designers follow accessibility guidelines. Long and detailed WAI documents such as WCAG 2.0 tend to be time consuming, hard to read and explore. There are many SCs to follow and accessibility barriers to avoid. This would take extensive hours to navigate and understand and require a certain level of technical knowledge of accessibility. Therefore, when developers or designers are required to implement accessibility, they do not always understand how to achieve the desired requirements unless

they undertake sufficient training (Alonso, Fuertes, González, & Martíez, 2010). As a result, real-world developers do not follow WCAG 2.0 and the accessibility level remains low. Thus, a new way to simplify the following guidelines is needed to increase the accessibility.

Therefore, this paper took a step forward in simplifying adherence to the guidelines for Web developers by proposing a metric that reveals the most prevalent and severe web accessibility barriers that affect websites in a certain sample, and then ranks them based on their severities. Web developers and designers can increase the accessibility level of their websites if they focus on checking against a concise list of SCs within WCAG 2.0 that are related to the most severe barriers which occur frequently and repair them accordingly, instead of wasting time and effort in studying large sets of barriers that may rarely occur (Hudson, 2010).

The proposed metric will be suitable to be used with any evaluation technique that reveals the type and the frequency of the barriers for each website such as automated tools, expert reviews and lab studies. However, automated testing was chosen to show the mechanism of the metric because of its ability to reduce evaluation time and support large scale evaluation studies.

WEB ACCESSIBILITY BARRIER SEVERITY METRIC

Several accessibility metrics to measure website accessibility have been discussed in the literature. Sullivan & Matson (2000) proposed a metric to classify websites based on their accessibility into four levels: highly accessible, mostly accessible, partly accessible and inaccessible. The websites evaluated automatically and manually then were classified based on failure rate (FR). FR was defined on the basis of a subset of WCAG 1.0 checkpoints where eight checkpoints were evaluated.

FR= Real_violations (checkpoints)/ Potential_violations

The advantage of such a metric lies in its simplicity: However, it does not consider other factors such as error impact, error nature and severity of detected violations.

Brajnik & Lomuscio (2007) proposed SAMBA which is a method for measuring website accessibility based on WCAG 1.0. This method uses the output of automated tools and couples it with human judgment, so that correct estimates of tool errors can be assessed, and estimations of severities

of barriers are used. However, applying this method absolutely needs experts for evaluation, otherwise it cannot be applied.

Zeng (2004) proposed the Web Accessibility Barrier (WAB) formula to evaluate website accessibility where a high WAB means a low accessibility level. This formula uses the total pages of a website, total accessibility errors as well as potential errors in a web page and error priority as input parameters. WAB_score= $\sum$ (real-errors/ (potential_errors x priority))

However, the returned ratings are not restricted to a limited range of values and not normalized. Thus, it can be useful only for ranking web pages according to their accessibility level. Moreover, this metric needs an expert to check the potential violations.

This paper proposed a Web Accessibility Barrier Severity (WABS) metric which was different from the previous metrics since it did not focus on measuring the accessibility level of websites. It was more concerned about revealing the persistent barriers that limit the accessibility based on their severity rather than total conformance to priority levels. Moreover, this metric was based on WCAG 2.0 which is the latest one from WAI and does not require human intervention to assign scores and weights. Besides, different approaches were suggested to determine severe barriers (Erickson, Trerise, Lee, VanLooy, & Bruyere, 2007; Socitm, 2008; Hudson, 2010). However, these approaches lack a clear mathematical model to identify and rank barriers based on their severities, since these approaches are either based on evaluator experience in determining severe barriers or on surveys sent to Web developers to indicate the most common barriers they faced when evaluating websites. Thus, this paper proposes a quantitative metric that assigns numerical weight to each identified accessibility barrier based on its severity and impact on accessibility level. The barriers can be ranked based on their weights. Thus, novice Web developers can concentrate on fixing a limited set of highly ranked barriers (most severe) instead of wasting their time in correcting less important types of barriers that score low weights (least severe).

REQUIREMENTS AND ASSUMPTIONS

In this section, the requirements and assumptions for WABS metric are described, based on the properties of a good web accessibility metric proposed by (Zeng, 2004; Vigo, Arrue, Brajnik, Lomuscio, & Abascal, 2007). Then, this section shows how WABS metric meets the measurements for scientific research.

<u>Requirement 1:</u> *Metric results should be normalized.*

Assumption1: The barriers must be measured in a quantitative score that provides a continuous range of values from perfectly major barriers to those completely not important.

Assumption 2: The metric values must have large discriminating power beyond that of simply major to minor.

In order to rank barriers according to their severity, a weight with a positive value associated with each barrier is chosen so that values of the final metric range from 0 to 1. The closer the result of the metric is to 0 the less severe the barrier is.

<u>Requirement 2:</u> *The metric should give one value for each barrier based on its influence on the priority class that it violates as well as to the whole dataset (the whole set of web page being assessed).*

Assumption 1: Beside the total frequency of errors for each barrier in the web page, the metric should also take into account the total number of times each barrier has been tested.

The metric should not be based on the absolute frequency of errors found for each barrier but on the relative number of errors found in relation to the number of tested cases, i.e. the ratio of errors and the number of tested cases. In other words, the number of websites that contain the same type of barrier should be taken into account.

Assumption 2: The metric should be scalable to conduct large-scale Web accessibility studies.

Assumption 3: The priority of unfulfilled SC (barrier) should be reflected in the final result.

Within WCAG 2.0 priority "A" SCs have more impact on the accessibility level of a web page than priority "AA" SCs and so on. No matter the value assigned for each priority, the value should reflect the difference between these priorities based on their importance. The unique restriction when selecting these weights is that:

1 >priorityA_weight>priorityAA_weight>priorityAAA_weight> 0

In this paper, priorities weight suggested by Vigo et. al. (2007) were adopted, where priority "A"=0.8, priority "AA"=0.16, priority "AAA"=0.04.

Requirement 3: *The measurement should be normative.*

The metric should be derived from standard guidelines of Web accessibility such as the WCAG. WCAG 2.0 is used as the foundation for WABS metric since it is the latest one from WAI.

Requirement 4: *Problems that need human judgment should not have influence on the final metric.*

Theoretically, the metric can be used to calculate scores based on SCs in WCAG 2.0 priorities. Nevertheless, this paper focuses only on barriers that can be checked automatically.
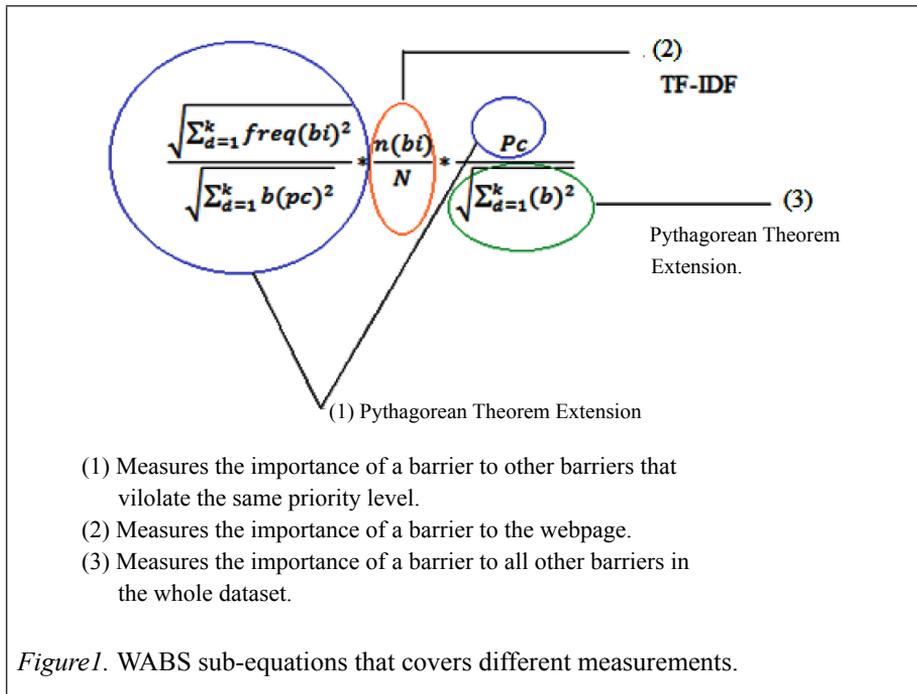
Metric Formulation

WABS metric is basically formed to meet the requirements of building a good metric as suggested by (Zeng, 2004; Vigo et al., 2007). Furthermore, it utilizes some parameters used by the metric proposed by Zeng (2004). Parameters such as barrier frequency and the priority level that has been violated as well as the concept of including the number of test cases, are used. The metric needs to take into consideration how a barrier weight can be affected by: other barriers that belong to the same priority level, other barriers that violate that same barrier's webpage, and by all barriers across the whole document. Thus, WABS needs different measurements for a scientific exploration, which are: (a) the importance of the barrier to the other barriers that violate the same priority level, (b) the importance of the barrier to the webpage, and (c) the importance of the barrier to all the other barriers in the whole dataset (all websites). The metric formulation is inspired by the vector space model which is part of the Information Retrieval field. The vector space includes three stages (Salton, Wong, & Yang, 1975). The first is document indexing; the second is term weighting which consists of the term frequency factor, the collection frequency factor and the length normalization factor; and the third is weights ranking. The same idea is applied to WABS since each document will be indexed by the barriers, then each barrier will be assigned a weight that will take into account the barrier importance to a priority level, to the webpage being tested and to the whole dataset, and finally the barriers will be ranked based on their weights. Two principles are used from the vector space model. The first one is the extension of the Pythagorean Theorem (Weisstein, 2014) which is used to measure barrier length in different contexts (this

principle covers the first and third measurements of the metric). The second principle is TF-IDF (Sparck Jones, 1972) which is adjusted to meet the second measurement of the metric (The importance of the barrier to the webpage).

The final metric

$$\frac{\sqrt{\sum_{d=1}^{k} freq(bi)^2}}{\sqrt{\sum_{d=1}^{k} b(pc)^2}} * \frac{n(bi)}{N} * \frac{Pc}{\sqrt{\sum_{d=1}^{k} (b)^2}}$$

This formula expresses the final metric (WABS). In order to demonstrate how the final metric is built, we broke it down into three main sub equations, namely formulas (1), (2) and (3). Equations cover the measurements which relate the importance of a barrier to its priority level, to its webpage and to the whole evaluation dataset. Figure 1 demonstrates the final metric (WABS) after aggregating the sub formulas. Besides, it displays how the sub formulas are connected to the extension of the Pythagorean Theorem and the TF-IDF principles. WABS is applied to each barrier separately to calculate its weight. The barriers severity can be ranked once each barrier weight is found.

(1) Measures the importance of a barrier to other barriers that vilolate the same priority level.
(2) Measures the importance of a barrier to the webpage.
(3) Measures the importance of a barrier to all other barriers in the whole dataset.

Figure1. WABS sub-equations that covers different measurements.

The importance of the barrier to the other barriers that violate the same priority level

$$\frac{\sqrt{\sum_{d=1}^{k} freq(bi)^2}}{\sqrt{\sum_{d=1}^{k} b(pc)^2}} * Pc \qquad (1)$$

where

> d= document (webpage) being tested,
> k= last document to be tested in the dataset,
> bi= barrier (violation) being checked,
> b= total number of barriers that appear in document d,
> $b(pc)$= total number of barriers that violate the same priority level, and
> Pc= priority level weight which the tested barrier violates.

This equation is inspired by the vector space model, where a document is represented in a 3-dimensional term vector space. A vector (v) can be expressed as a sum of elements such as,

$v = a_1 v_{i1} + a_2 v_{i2} + \ldots + a_n v_{in}$

where (a_n) are called scalars or weights and v_{in} are components or elements.

In order to calculate the document length in a vector space, the extension of the Pythagorean Theorem (Weisstein, 2014) is used as follows:

$d = (x1, x2, x3, ..., xn)$ is a vector in an n-dimensional vector space. Length of x is given by (extension of Pythagoras's theorem):

> $|d|^2 = x1^2 + x2^2 + x3^2 + ... + xn^2$
> $|d| = (x1^2 + x2^2 + x3^2 + ... + xn^2)^{1/2}$

The webpages in a collection can be viewed as a set of vectors in the vector space in which there is one axis for every barrier. Each webpage in the dataset has a different type and frequency of errors (barriers). As a result, webpage length is different and should be measured based on barriers' impact on the priority level.

The only problem with this assumption is that a webpage cannot be treated as a document of terms. The severity of barriers is not only affected by a webpage in the dataset since many factors are controlling the severity, such as

other barriers in the dataset and the priority level that has been violated. So, the vector length formula is adjusted to measure the severity of barriers that violate a priority level across the whole dataset.

Let x be the length of all barriers that violate the same priority level across the whole dataset as illustrated in Figure 2. Then,

$$|x|^2 = b1^2 + b2^2 + b3^2 + ... + bn^2$$
$$|x| = (b1^2 + b2^2 + b3^2 + ... + bn^2)^{1/2}$$

The same idea is applied for a certain barrier length across the whole dataset as well as the total length of barriers across the dataset.
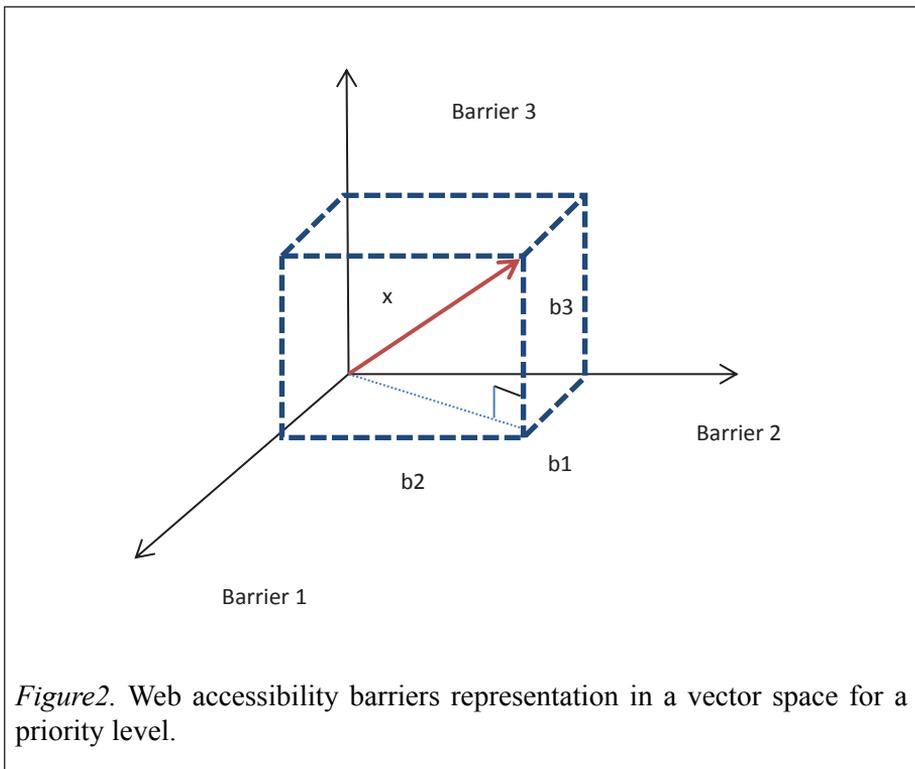

Figure2. Web accessibility barriers representation in a vector space for a priority level.

Since all the values cannot be negative, there is no need for the absolute value. This leads to Equation (1). For simplicity, Equation (1) is refined into Formulas (1.1) and (1.2), where Formula (1.1) represents the numerator in Equation (1), while (1.2) represents the denominator. These formulas are discussed later in this section.

Eq. (1) steps:

1.   Divide the result of Formula (1.1) by the result of Formula (1.2).
2.   Multiply the result of step 1 by the weight of the priority class which the barrier (bi) violates, where (Priority "A"=0.8, priority "AA"=0.16, priority "AAA"=0.04).

$$\sqrt{\sum_{d=1}^{k} freq(bi)^2} \tag{1.1}$$

Equation (1.1) defines the length of a certain barrier that violates a specific priority level across the whole document. To calculate Equation (1.1), follow the steps below:

1.   For each webpage, count how many times a certain barrier (bi) appears (i.e. frequency).
2.   Calculate the square of step 1.
3.   Find the summation of squares for step 2 across the whole dataset.
4.   Calculate the square root for step 3 final results.

$$\sqrt{\sum_{d=1}^{k} b(pc)^2} \tag{1.2}$$

Eq. (1.2) calculates the length of all the barriers across the whole document that violates the same barrier's priority class. Note that result of this formula will be fixed for each set of barriers that violate the same priority level. The steps are described below:

1.   For each webpage, find the total number of barriers that belong to the same priority class (check first against priority "A").
2.   Find the square of step 1.
3.   Repeat the previous steps for the whole dataset (all webpages).
4.   Sum up the results of step 3 (the summation of barriers squares).
5.   Find the square root for the final result (step 4).
6.   Repeat the previous steps for the remaining priority classes ("AA" and "AAA").

The importance of the barrier to the whole document

$$\frac{n(bi)}{N} \tag{2}$$

where,

> *n(bi)= the number of different documents (webpages) the barrier (bi) appears in, and N= the total number of documents.*

This formula shows the ratio of the documents that the barrier appears in to the total number of documents in the dataset.

This formula is inspired by a principle in Information Retrieval known as TF-IDF (Sparck Jones, 1972). TF-IDF assesses a document's word importance by counting its frequency.The term frequency (TF) is simply the number of times a given term appears in a specific document. The inverse document frequency (IDF) is a measure of the general importance of the term which means the terms which appear in many documents are not very useful for distinguishing a relevant document from irrelevant ones. However, in WABS metric, a barrier which appears in many webpages indicates how significant the barrier is to the whole set of webpages. Roughly speaking, WABS is more concerned about measuring how common a barrier is across an entire collection of webpages. Thus, the TF-IDF is adjusted to fit the metric goal.

Formula 2 steps are described below:

1. Calculate how many times a certain barrier (bi) appears in a different document.
2. Divide step 1 by the total number of documents (fixed across the documents).

The importance of the barrier to all the other barriers in the whole dataset (the whole tested webpage)

$$\sqrt{\sum_{d=1}^{k}(b)^2} \tag{3}$$

Eq. (3) is fixed for all the documents and calculated once. It calculates the total lengths of all barriers across the whole document. The steps are as follows:

1. For each document, find the summation of barriers that appear in it.
2. Find the square of step 1.
3. Repeat the first two steps for all documents.
4. Sum up the result of step 3.
5. Find the square root of step 4.

METRIC ATTRIBUTES EVALUATION

Freire, Russo, & Fortes (2008) in their paper discussed how to evaluate web metric based on its attributes. Yet, they did not have an interest in measuring barriers severity; hence we adopted some attributes that can be applied to the proposed metric to evaluate its measurements. Table 1 shows the attributes and their descriptions and how they were mapped and addressed in WABS metric.

Table 1

WABS Attributes Evaluation

Attribute	Description	WABS
Guidelines set	Set of guidelines used to the metric, such as WCAG, self-defined guidelines set or customized set of guidelines.	WABS metric utilizes WCAG 2.0 as basis for metric computation.
Coupling level with guidelines	How coupled the metric is with the guidelines set and whether it is easy or not to change the guidelines set.	WABS is strongly linked to WCAG 2.0 guidelines and priority levels to define barriers weights.
Type of evaluation	Type of evaluation methods that a metric is supposed to support, such as automatic testing, manual inspections or user testing.	WABS is designed to work with any evaluation technique that enumerates the type and the frequency of barriers for the tested webpage. Nonetheless, this work will only analyse the results obtained by automated tools to conduct large-scale assessment.
Considers site complexity	The metric considers the complexity and size of the site, usually in number of pages.	The main goal of WABS is not website evaluation rather than barriers severity evaluation.

(continued)

Attribute	Description	WABS
		As a consequence, WABS considers the size and the complexity of each webpage in the dataset, related to the influence of the number and types of barriers in that webpage.
Type of barriers weights coefficients	Use of predefined coefficients based on the priority of barriers in a given set of guidelines, user-derived coefficients or other approaches.	WABS considers weights based on WCAG 2.0 barriers.
Default coefficients	Default values for barrier weights coefficients.	WABS uses fixed values related to the priority of each SC in WCAG 2.0, (priority "A"=0.8, priority "AA"=0.16, priority "AAA"=0.04).
Automated tool	Automated tool for metric computation, if any. WABS is used in the context of automatic evaluation.	The metric is associated with A-checker tool.
Used in large scale	Use of a metric in large scale evaluations.	The experiment conducted in this paper includes a large-scale assessment to stimulate the metric.

METRIC VALIDITY

Validation of software metric confirms or rejects the correctness of a given implementation of that particular software metric regarding its specification (Rüdiger, 2009), i.e. the calculated metric values correspond to values which are expected by the metric specification. There are two types of validations for metrics: theoretical and empirical validations (Srinivasan & Devi, 2014). The theoretical validation confirms that the measurement does not violate any necessary properties of the metric measurements. The empirical validation confirms that measured values of attributes are consistent with the values predicted by models involving the attribute (Rüdiger, 2009).

The theoretical validation approves that WABS metric was not built arbitraril since it adheres to the requirements and assumptions discussed earlier. The

requirements to build a good metric suggested by the eminent researchers Zeng (2004) and Vigo et al. (2007) were carefully followed when formulating variables, constants and the final metric. Furthermore, Web accessibility metrics proposed by different researchers (Sullivan & Matson, 2000; Zeng, 2004; Arrue, Vigo, & Abascal, 2005; Velleman et al., 2006) have been extensively studied from the viewpoint of formulations and goals. WABS metric utilizes some parameters used by the metric proposed by Zeng (2004). The metric formulation in terms of barriers weighting and ranking is inspired by the vector space model. Thus, the metric measurements cover: (a) the importance of the barrier to the other barriers that violate the same priority level, (b) the importance of the barrier to the webpage, and (c) the importance of the barrier to all the other barriers in the whole dataset (all websites). Two principles were used from the vector space model, namely the extension of the Pythagorean Theorem (Weisstein, 2014), which was used to measure barrier length in different contexts (this principle covers the first and the third measurements). The second principle is TF-IDF (Sparck Jones, 1972) which was adjusted to meet the second measurement.

For empirical validation, Sheppard & Darrel (1993) stressed the point that the measurement method should be capable of discrimination and will not assign the same value to every object. Another assumption that requires validation is that the more the barrier appears in a dataset the higher the likelihood the barrier will be more severe. Thus an experiment was conducted and analyzed to check if the metric values had discriminating power and if the barrier frequency had a positive correlation with its weight.

EXPERIMENTAL STUDY

This section introduces the experiment procedure and findings. Following these, the results are analyzed and further discussed.

Experiment Procedure

The proposed metric (WABS) is designed to work well with automated tools. It is calculated automatically from evaluation reports yielded by the evaluation tool. Although automated testing does not check all the SC, they can reduce the time and effort required for the evaluation process. Our metric utilized the results of the A-Checker tool available at https://achecker.ca/, which is an open source accessibility evaluation tool that allows the user to select the guidelines he wants to check against, then submit a webpage via its URL or by uploading its HTML file. The SCs that A-Checker can evaluate are listed in Table 2 (Note that the barrier id corresponds to the SC id that has been violated).

Table 2

Success Criteria that can be Checked by A-Checker

Guideline number	Success criterion ID	Success criterion description	Priority level
1.1	1.1.1	Alternatives for non-text content.	A
1.3	1.3.1	Info and relationships are correctly formatted and in order.	A
1.4	1.4.1	Colour is not the only identifier.	A
1.4	1.4.4	Text can be resized without assistive technology up to 200% without loss of content or functionality except for captions and images.	AA
1.4	1.4.6	The visual presentation of text and images of text has a contrast ratio of at least 7:1.	AAA
2.1	2.1.1	All functionality of the content can be navigated by keyboard only.	A
2.2	2.2.1	Time limits have timing controls.	A
2.2	2.2.2	Pause, Stop, Hide: For moving controls > 5 sec.	A
2.4	2.4.1	Provide a "skip to content" link.	A
2.4	2.4.2	Web pages have titles that describe topic or purpose.	A
2.4	2.4.4	Link Purpose (In context): The purpose of each link can be determined from the link text alone.	A
2.4	2.4.6	Headings and labels describe topic or purpose.	AA
3.1	3.1.1	The default human language of each Web page can be programmatically determined.	A
3.3	3.3.2	Labels or instructions are provided when content requires user input.	A
4.1	4.1.1	When parsing: No major HTML errors.	A

Homepages of 500 different Malaysian homepages were selected. After selecting the dataset of websites and the automated tool, an experiment was conducted to examine the validity of the metric by automatically evaluating the selected dataset using A-Checker.

RESULTS AND FINDINGS

Table 3 shows the total frequency of each barrier generated by A-Checker across the whole dataset in descending order. The most frequent barrier across the dataset was scored by (1.1.1) with 4626 occurrences. On the other side, barrier (2.2.1) only appeared once.

Table 3

The Total Frequency of each Barrier across the Dataset

Barrier id	Total frequency	Priority level
1.1.1	4626 **(most frequent)**	A
1.4.6	2327	AAA
2.1.1	710	A
1.4.4	592	AA
1.3.1	311	A
2.4.4	222	A
3.3.2	156	A
3.1.1	117	A
2.4.6	104	AA
1.4.1	76	A
4.1.1	52	A
2.2.2	35	A
2.4.2	16	A
2.4.1	15	A
2.2.1	1 **(least frequent)**	A
Total	9360	

After collecting the needed data, WABS metric was applied to calculate the weight for each barrier. Table 4 presents the weights generated by WABS for each barrier ordered from the most severe barrier (highest weight) to the least. The most severe barrier found in the dataset was (1.1.1) while the least severe one was (2.2.1).

Table 4

The Weight for each Barrier in the Dataset using A-Checker

Barrier id	Barrier weight	Priority level
1.1.1	0.000107 **(most severe)**	A
1.4.4	1.22E-05	AA
1.3.1	6.14E-06	A
1.4.6	4.4E-06	AAA
2.1.1	2.04E-06	A
3.3.2	1.82E-06	A
2.4.4	1.05E-06	A

(continued)

Barrier id	Barrier weight	Priority level
3.1.1	8.67E-07	A
4.1.1	1.35E-07	A
2.4.6	5.99E-08	AA
1.4.1	4.73E-08	A
2.2.2	1.73E-08	A
2.4.2	4.37E-09	A
2.4.1	3.81E-09	A
2.2.1	7.87E-11 **(least severe)**	A

Testing the Relation between Barrier Frequency and its Weight

From Tables 3 and 4, it can be noticed that barrier (1.1.1) was the most frequent and severe, while (2.2.1) had the least frequency and severity in dataset (A). Thus, a further inferential analysis was carried out to investigate if there is a relation between barriers frequency and severity. Table 5 shows each barrier along with its total frequency and weight.

Table 5

Barriers with Corresponding Weight and Frequency

Barrier id	Weight	Total frequency
1.1.1	0.000107	4626
1.3.1	6.14E-06	311
1.4.1	4.73E-08	76
1.4.4	1.22E-05	592
1.4.6	4.40E-06	2327
2.1.1	2.04E-06	710
2.2.1	7.87E-11	1
2.2.2	1.73E-08	35
2.4.1	3.81E-09	15
2.4.2	4.37E-09	16
2.4.4	1.05E-06	222
2.4.6	5.99E-08	104
3.1.1	8.67E-07	117
3.3.2	1.82E-06	156
4.1.1	1.35E-07	52

A Spearman rank-order correlation was conducted in order to determine if there were any relationships between barriers frequency and weights. A

two-tailed test of significance indicated that there was a significant positive relationship between the frequencies and weights $r_s(15) = .96$, $p < .001$. Thus, the more frequently the barrier appears in the evaluation set the more likely it would be more severe.

DISCUSSION

Table 4 shows that the metric results were positive, finite and normalized with continuous range of values falling within the range (0-1) as intended by the specification. It is obvious that metric values have large discriminating power. This supports Sheppard & Darrel's (1993) contention. Table 4 shows how barriers weights that belong to the same priority level are different. For example, barriers (1.1.1), (3.3.2) and (2.2.1) scored different weights even though they belong to priority "A". Each barrier has a unique value even for barriers belonging to the same priority level. This allows the influence of each barrier that belongs to the same priority level to be observed.

Another interesting finding was that barriers that belong to the higher priority level are not necessarily more severe than the other types of barriers. For example, in Table 4, barriers (1.4.6) and (1.4.4) which belong to priority classes "AAA" an "AA" respectively, were more severe than the other barriers that belong to class "A".

The empirical validity of the metric highlighted in this paper was also tested by the correlation analysis between barrier frequency and barrier weight. The Spearman rank correlation showed that there is a positive correlation between the total number of accessibility barriers and the weight obtained for each barrier with $r_s(15) = .96$ and $p < .001$. This experiment confirms that barriers' weight does not merely depend on priority level. The total frequency of a barrier in the dataset has a large effect on its weight. This confirms the logical assumption that the more the barrier appears in a dataset the more likely the barrier will be more severe. Therefore, when the WABS metric was formulated it took into account the barrier frequency measure.

CONCLUSION AND FUTURE WORK

Web Accessibility Barrier Severity (WABS) is proposed to rank accessibility barriers based on their severity. The metric takes into account (a) the importance of the barrier to the other barriers that violate the same priority level, (b) the importance of the barrier to the webpage, and (c) the importance

of the barrier to all the other barriers in the whole dataset. An experiment was conducted to examine the metric validity and a dataset of 500 websites was selected from Malaysia. The experiment analysis shows that WABS did meet the properties which a valid metric should have. However, in order to check the metric reliability and sensitivity, several experiments should be conducted to examine the metric behavior under different contexts. Thus, a future study is needed to show the metric's reliability when different tools, time span and samples are used.

REFERENCES

Abuaddous, H., Jali, MZ, & Basir, N. (2013). Study of the accessibility diagnosis on the   public higher institutions websites in Malaysia. *Proceedings of the 4th International Conference on Computing and Informatics* (ICOCI 2013) (pp. 122-127). Sarawak, Malaysia: Universiti Utara Malaysia.

Abuaddous, H., Jali, MZ, & Basir, N. (2016). Web accessibility challenges. *International Journal of Advanced Computer Science and Applications* (IJACSA), *7*(10), 172-181.

Ahmi. A., & Mohamad, R. (2016). Evaluating accessibility of Malaysian public universities websites using Achecker and Wave. *Journal of Information and Communication Technology*, *15*(1), 193-214.

Alonso, F., Fuertes, L., González, L., & Martíez, L. (2010). On the testability of wcag 2.0 for beginners. *7th International Cross-Disciplinary Conference on Web Accessibility* (W4A'10) (pp. 1-9.). Raleigh, North Carolina: ACM Press, New York, NY, USA.

Arrue, M., Vigo, M., & Abascal, J. (2007). Quantitative metrics for web accessibility evaluation. *Proceedings of the ICWE 2005 Workshop on Web Metrics and Measurement*.

Brajnik, G., & Lomuscio, R. (2007). SAMBA: A semi-automatic method for measuring barriers of accessibility. *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility* (Assets '07) (pp. 43-50). ACM, New York, NY, USA.

Brown, J., & Scott, H. (2015). The challenges of web accessibility: The technical and social aspects of a truly universal web. *First Monday*, *20*(9).

Erickson, W., Trerise, S., Lee, C. G., VanLooy, S., & Bruyere, S.M. (2007). Web based student processes at community colleges: Removing barriers to access. Retrieved from: http://digitalcommons.ilr.cornell.edu/edicollect/1241.

Freire, A.P., Russo, C.M., & Fortes, R. P. M. (2008). A survey on the accessibility awareness of people involved in web development projects in Brazil. *Proceedings of 2008 International Cross-disciplinary Conference on Web Accessibility* (W4A'08) (pp. 87-96). Beijing, China: ACM Press, New York, NY, USA.

Hashemian, B. (2011). Analyzing web accessibility in Finnish higher education. *Proceedings of the 13th international ACM Sigaccess Conference on Computers and accessibility* (Assets '11) (101, pp. 8-16). Dundee, Scotland, UK: ACM Press, New York, NY, USA.

Hudson, R. (2010). Ten common accessibility problems. Retrieved from: http://usability.com.au/2010/07/ten-common-accessibility-problems/.

Nahon, K., Benbasat, I., & Grange, C. (2012). The missing link: Intention to produce online content accessible to people with disabilities by non-professionals. *Proceedings of 2012 45th Hawaii International Conference on System Sciences* (HICSS) (pp. 1747-1757). Maui, HI, Hawaii: IEEE Computer Society, Washington, DC, USA.

Rüdiger, L. (2009). *Validation of a standard- and metric-based software quality model* (Unpublished doctoral dissertation).Växjö University. Acta Wexionensia, Nr.

Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. *Commun, ACM, 18*(11), 613-620. doi: 10.1145/361219.361220

Sheppard, M., & Darrel, I. (1993). *Derivation and validation of software metrics.* Oxford, Clarendon Press.

Shin, H., Lim, J., Lee, J., & Kyung, K. (2013). Background display for visually impaired people in mobile touch devices. *Proceedings of 2013 IEEE International Conference on Consumer Electronics* (ICCE) (pp. 464-465). Las Vegas, NV, USA: IEEE Computer Society, Washington, DC, USA.

Socitm. (2008). A world denied: A supplement for better connected 2008 on website accessibility. Retrieved October 17, 2016 from: http://www.socitm.net/downloads/download/37/a_world_denied_a_supplement_for_better_connected_2008_on_website_accessibility

Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, *28*, 11–21.

Srinivasan, K.P., & Devi, T. (2014). Software metrics validation methodologies in software engineering. *International Journal of Software Engineering & Applications* (IJSEA), *5*(6), 87-102.

Sullivan, T., & Matson, R. (2000). Barriers to use: Usability and content accessibility on the Web's most popular sites. *Proceedings of ACM Conference on Universal Usability*, 139-144.

Vigo, M., Arrue, M., Brajnik, G., Lomuscio, R., & Abascal, J. (2007). Quantitative metrics for measuring web accessibility. *Proceedings of the 2007 International Cross-Disciplinary Conference on Web Accessibility* (W4A '07) (pp. 99-107). ACM, New York, NY, USA.

Velleman, E., Meerveld, C., Strobbe, C., Kosh, J., Velasco, C.A., Snaprud, M., & Nietzio, A. (2007). D- WAB4 unified web evaluation methodology (uwem 1.2 core). Technical report, *WAB Cluster*.

Villena, J., Ramos, B., Fortes, R.P.M., & Goulartea, R. (2014). Web videos – concerns about accessibility based on user centered design. *Elsevier Procedia Computer Science*, *27*, 481-490.

W3C WEB Accessibility Initiative. (2005). Introduction to web accessibility. *Web accessibility means that people with dissabilities can use the web*. Retrieved from: http://www.w3.org/WAI/intro/accessibility

Web Content Accessibility Guidelines 2.0. (2008). *A guide to understanding and implementing Web content Accessibility guidelines 2.0*. Retrieved from: http://www.w3.org/TR/WCAG20

Weisstein, E. (2014). Pythagorean Theorem. *MathWorld-A Wolfram Web Resource*. Retrieved from: http://mathworld.wolfram.com/Pythagorean Theorem.html.

Zeng, X. (2004). *Evaluation and enhancement of web content accessibility for persons with disabilities* (Unpublished doctoral dissertation). University of Pittsburgh.