

Threefold Web Accessibility Evaluation by the Use of an Integrated Tool

Andreas P. GIANNAKOULOPOULOS
Faculty of Communication and Media Studies, University of Athens
Athens, Greece

ABSTRACT

Following W3C's WAI and Section 508 suggestions and regulations there has been a continuously increasing interest in the field of websites' evaluation for conformance with accessibility guidelines. In this paper an attempt is made to provide an integrated tool, both in theoretic and implementation levels, which evaluates websites for accessibility in a multidimensional manner and outputs comparable results. The proposed tool evaluates websites in three different aspects of accessibility. Each kind of evaluation outputs a numeric grade which serves as input to a unifying formula. This formula, in turn, results to a final numeric grade for the website under evaluation. The modularity which underlies the threefold approach allows researchers to customize the technique according to the needs of each particular research, while retaining the core characteristics of the approach, namely uniformity and comparability. The operational details of the threefold web accessibility evaluation tool should be considered to be a working example of the platform's potential to improve accessibility research towards a common path by the integration of today's prevalent research inclinations which favour either objective or subjective methods exclusively.

Keywords: Web accessibility, web usability, website evaluation, website development and accessibility scale.

1. INTRODUCTION

Accessibility research is a relatively recent field in which scientific methods are applied in order to evaluate the level of accessibility of the websites under investigation [6]. From this fact stem two important characteristics of this research field: (a) despite of the fact that various paradigms have been developed, none of them has been up to now widely accepted as a prototype, which in turn results to the lack of an accessibility scale, beyond the three levels

of conformance proposed by the W3C [9], and (b) accessibility researches are conducted with a variety of tools, methods and techniques, utilizing different evaluation scales and obviously produce heterogeneous, incomparable outputs. Characteristically, it has been reported that "*in order to help organizations make their websites accessible, a number of methods and tools have been developed by researchers, practitioners and Information Technology (IT) companies. There are different views about the best evaluation method, yet there are relatively few studies that have been conducted comparing AEMs. A review of the small number of papers available suggests that comparisons are difficult because different methods measure different variables. For this reason the validity of the conclusions drawn in some studies may be questionable*" [2].

2. EVALUATION CRITERIA

Inspired by the will to obtain more accurate records of accessibility measure in comparative studies, this paper proposes an integrated tool for threefold accessibility evaluation. This tool follows to a large extent W3C's suggestions with respect to the appropriate methodology for the evaluation of websites' accessibility.

The main objective is to integrate in a common platform the two major types of web content accessibility guidelines provided in [9]: the subjective guidelines (i.e. the ones which are difficult or impossible to be evaluated by measurable and objective criteria, such as the use of the "*clearest and simplest language appropriate*" [9]), and the larger in number objective guidelines (i.e. the ones which may be automatically checked, such as the existence of alternative text for images) [3, 6]. Due to the technical nature of the guidelines, usually accessibility evaluation studies utilize automation tools exclusively [2, 10]. As a result, the main problematic area concerns the evaluation by humans, which may be employed either by field

experts, or, in special cases, like the one of disabled persons, by members of that particular group of users [7, 9]. The complementarity among the two research techniques (automated and human evaluation) is considered to be essential and, consequently have been both incorporated in the tool under discussion, though diversely weighted. It is still worth mentioning that the dissimilarities among the various research variables raise serious obstacles with respect to the quantification of the final results.

The core essence of the proposal lies within the methodological adoption of tools which favor measurable results produced by both objective and subjective techniques, the latter being controlled and delimited, but not, no matter what the case, eliminated.

3. THREEFOLD EVALUATION PROCESS

The proposed evaluation platform consists of three reckoning procedures and a unifying equation which assigns a final grade to each evaluated website. The three procedures are the following: (a) the code evaluation, in which the webpage code is evaluated against the objective criteria of the accessibility guidelines, (b) the validity checks with respect to the published grammars of markup languages and style sheets, and (c) the evaluation by one or more experts against a checklist of subjective criteria (grouped under similar categories) and the assignment of a numerical grade of conformance for each category of criteria. The results obtained from each procedure are weighted by predefined special weight factors and the final grade is computed by the use of the unification formula.

Introductory comments:

- Trial 1 (code evaluation) is the basic and most important evaluation process with respect to the objective criteria. Trial 1 corresponds to the semi-automated check which is often the dominant –and in some cases the only– evaluation mechanism in accessibility researches [10].
- The validity checks, which examine the syntax of the markup language and of the style sheets used in the websites under examination, fully conform to the relevant guidelines, where valid markup and style sheet syntax is considered an inescapable accessibility feature. Additionally, the validity checks reflect the overall spirit of content accessibility guidelines, according to which conformance to web standards and WCAG “*will also make Web content more available to all users,*

whatever user agent they are using (e.g., desktop browser, voice browser, mobile phone, automobile-based personal computer, etc.) or constraints they may be operating under (e.g., noisy surroundings, under- or over-illuminated rooms, in a hands-free environment, etc.)” [9]. Following the Web standards, on the side of the developers, enhances the internationalization of the Web, highlights its global scope and makes content accessible to all users.

- The unification formula is obviously the key element as far as the mechanistic construction of the tool is concerned, since it allows researchers to compose a final value of accessibility for each website under examination and, most important of all, produces numeric results, which, in turn, may be used for the construction of an accessibility scale. Consequently, the unification formula facilitates inter-website comparisons and meta-research.

4. TRIAL (A): CODE EVALUATION

The publicity and support that the W3C’s Web Accessibility Initiative (WAI) received after its establishment resulted in an increased number of software products and mechanisms for the automatic evaluation of websites with respect to WCAG [8]. In order to select the appropriate mechanism for the first trial a large number of products was tested before the final selection which adopted the accessibility evaluator developed by the France Télécom and partners under the project-name Ocawa (Operational Control and Analysis for Web Accessibility) [5].

The selection of Ocawa mechanism as the basic tool for code evaluation against a broaden set of rules (of the objective type in most cases) is substantiated by the following reasons:

- It is a relatively recent tool, developed after the year 2000, and embodies the latest advances in the filed of Web accessibility.
- It evaluates multiple pages per request which offers apparent advantages as far as the economic aspects of the research are concerned. Evaluating, in an efficient and cost effective manner, multiple pages also increases the reliability of the research since the probability of random effects is decreased (in analogy to the number of evaluated pages per website).
- The mechanism works as an expert-system, evaluating HTML streams against a “pool of rules” [5]. The rule set of the mechanism includes (a) the

rules of USA's Federal Law Section 508, (b) the W3C's Web Content Accessibility Guidelines, (c) the rules established by France Télécom, which address also ergonomic issues, (d) the guidelines of the French Agency for the Development of Electronic Administration and (e) the rules collection entitled Accessiweb which has been composed by the Braillet association [5].

During the first trial of the threefold accessibility evaluation, websites are submitted to the selected engine for multiple pages evaluation. For every page of each website the engine returns the number of remarks (i.e. errors) and the number of tags that appear in the code of the page. The value of the composite variable "error rate" is calculated from the division of the total number of remarks (sum of remarks) by the total number of tags (sum of tags).

$$ERR_{rate} = \frac{\sum_{i \in A} R_i}{\sum_{j \in A} T_j} \quad \text{Eq. (1)}$$

Where:

ERR_{rate} : the error rate, R_i : the number of remarks for page i (i belongs to the subset $A = \{n \in \mathbb{N}: 1 \leq n \leq P_{total}\}$ of the set of natural numbers \mathbb{N} , and covers the total number of pages - P_{total} - under evaluation) and T_j : the number of tags to be found in page j (i belongs to the subset $A = \{n \in \mathbb{N}: 1 \leq n \leq P_{total}\}$ of the set of natural numbers \mathbb{N}). Following the calculation of error rate, the set of R_i is denoted as R_{total} and, respectively, the set of T_j is denoted as T_{total} .

The error rate variable is considered to be a reliable measure of accessibility problems, due to the fact that for each page the number of tags -i.e. the basic structural elements of the Web- is taken into consideration. The calculated values of the variable reflect number of accessibility problems per tag, which is a far more dependable measure than mere absolute number of errors which are affected by extraneous factors, such as the size of each webpage.

5. TRIAL (B): VALIDITY CHECKS

The second trial includes the validity checks with respect to the formally established grammars for markup and style sheets. The final output of the second trial is the value of the variable named "multiplier", the range of which is the set $\{1, 2, 4\}$.

The final value is calculated by multiplying the values of the Boolean variables "grade of markup validity" and "grade of style sheet validity", the range of values for each one of them being the set $\{1, 2\}$. Validity value equal to 1 denotes approved usage of the respective grammar and validity value equal to 2 denotes erroneous syntax with respect to the corresponding standards.

$$M = G_{HTML} \cdot G_{CSS} \quad \text{Eq. (2)}$$

Where:

M : the multiplier with range of values the set $\{1, 2, 4\}$, G_{HTML} : grade of markup validity with range of values the set $\{1, 2\}$ and G_{CSS} : grade of style sheet validity with range of values the set $\{1, 2\}$.

In details, according to the threefold accessibility evaluation method, every website under investigation is checked for markup and style sheet validity. The semi-automatic mechanisms proposed for those tests are the ones provided as services by the W3C, namely the Markup Validation Service and the CSS Validation Service, respectively. Validity checks correspond to particular accessibility guidelines and, additionally, are clearly suggested by W3C as an integral part of a complete accessibility evaluation scheme [9, 6].

With regards to validity checks the following comments have to be taken into consideration:

- The range of values for the variables utilized in the second trial have been defined in such a way that the final output (i.e. the multiplier) will always be a strong indicator of the accessibility level employed in each evaluated website. Given that during the unification process the error rate (i.e. the output of the first trial) is multiplied by the value of the multiplier the overall penalty grade of the website under discussion is doubled for each and every failure in validity checks. The powerful influence of the multiplier to the final outcome mirrors the importance attributed by the threefold process to conformance with Web standards, as part of highly accessible Web constructions.

- It must be also noted that validity checking mechanisms, and especially the markup validation service of the W3C proposed in the current framework, are extremely austere as far as conformance to the grammar is concerned. Minor errors, syntax mistakes or even typos, which are usually overlooked by the common browsers, may result in validity failures, which, as described right above, intensely affect the final penalty grade. Of

course, the severity of the validity mechanism neither distorts, nor biases the final research results which retain their comparability within the sample population of websites, since all websites of the same sample are submitted to the same tool for the validity checks of the second trial.

6. TRIAL (C): HUMAN REVIEW

The third trial covers the proposed in the relevant bibliography manual evaluation by experts [3, 6]. The human review within the threefold evaluation framework turns into a measurable trial by the assignment of grades to websites with respect to subjective criteria for content accessibility. The checklist of criteria for the third trial stems from the subjective guidelines and can be found in the checklist proposed by the W3C, along with relevant usability criteria [4], all of which share in common the fact that it is impossible up to now to be automatically checked by some sort of software. All of the criteria in the checklist are classified under five broad categories for analysis and examination. Each website under evaluation receives a grade in each of the five categories, reflecting the extend to which the particular website fulfils the criteria falling under the specific category. The grade scale for each category is the set $\{n \in \mathbb{N}: 0 \leq n \leq 5\}$, where a grade equal to 0 corresponds to the ideal status in which all criteria of the category are adequately fulfilled, whereas a grade equal to 5 corresponds to the worst case in which the website is totally insufficient with respect to accessibility and usability criteria of the particular category. The in between values degrade gracefully and analogically from the excellent status to a severely problematic one.

$$E_{avg} = \frac{\sum_{k \in B} E_k}{\max B} \quad \text{Eq. (3)}$$

Where:

E_{avg} : The expert average grade of all five evaluation categories, marked by the expert -or the average of the grades of experts if more than one, so that the input to Eq.(3) will always be a single value per category, E_k : the grade given to the website by the expert for category k (k belongs to the subset $B = \{n \in \mathbb{N}: 1 \leq n \leq 5\}$ of the set of natural numbers \mathbb{N} , and covers all the categories of the analysis) and $\max B$: the maximum value element in set B, so that

the quotient of the division in Eq. (3) will be the average of grades in all categories.

The five categories of subjective criteria against which websites are graded by experts in the third trial according to the threefold accessibility evaluation method are the following:

Navigation: the grade in this category collectively reflects the conformance to subjective criteria related to the usability, consistency and clarity of the navigation mechanisms (e.g. the existence of a clear and consistent navigation bar across the whole website).

Content: the grade in this category collectively reflects the conformance to subjective criteria related to the content of the website and the usage of natural language (e.g. the descriptiveness and the explicitness of the text used in titles and hyperlinks).

Structure: the grade in this category collectively reflects the conformance to subjective criteria related to the structural organization of information within the website (e.g. the construction of meaningful nodes and leaves in tree-like structure for content, correlated in a clear and distinct manner).

Requirements: the grade in this category collectively reflects the conformance to subjective criteria related to software requirements which must be fulfilled by the prospective user in order to successfully use the website (e.g. the design and coding for specific browsers as opposed to the design and coding for any user-agent and platform).

Support: the grade in this category collectively reflects the conformance to subjective criteria related to help and support mechanisms and facilities (e.g. the existence of wizards for the accomplishment of specific tasks, the existence of help pages or Frequently Asked Questions and the like).

6. UNIFICATION FORMULA

Recalling that the objective of the threefold platform is to produce a numerical value reflecting the accessibility level of each website, in order to promote comparability and gradation among websites with respect to content accessibility, a unifying equation has been developed which integrates the outputs of the various trial and produces a single numeric grade for each website (in the form of penalty degrees). The formula of the equation reads as follows:

$$F = (ERR_{rate} \cdot M \cdot W_1 + E_{avg} \cdot W_2) \cdot 100$$

Eq. (4)

Where:

F (final score): The final score (penalty grade) outputted by the integrated evaluation tool.

ERR_{rate} (error rate): The value of the error rate variable, as received from the first trial.

M (multiplier): The value of the multiplier variable, as received from the second trial.

W₁ (weight 1): Weighting factor for the conjunct outcome of the first and second trial. The predefined value for this factor is equal to 0.7.

E_{avg} (expert average grade): The average grade of all five categories for each website, assigned by the expert analysts as part of the third trial.

W₂ (weight 2): Weighting factor for the outcome of the third trial. The predefined value for this factor is equal to 0.3.

With respect to the unification formula the following observations clarify certain issues:

- The constant at the end of the formula (100) bears no influence the research results and the ordering of the websites, since the unified outcome of all trials (expressed by the sum within the parentheses) is multiplied by the same constant. The numeric constant has been added in order to make final grades more legible.
- The sum within the parentheses, in which all three trials are included, has been set up in such a manner that separates from one side all trials concerning the objective criteria (i.e. automatically conducted) and from the other side the trial concerning the subjective criteria (i.e. the evaluation conducted manually by experts). Consequently, the first addend product includes the results of the code evaluation and the validity checks (values of the variables “error rate” and “multiplier”, assigned by automated techniques), while the second addend product includes the output of the human evaluation (value of the variable “expert average grade”, assigned manually by the experts).
- The weighting factors W1 και W2 reflect the relative specific weight attributed by the threefold accessibility evaluation platform to the two major components of the final grade, namely the objective and the subjective evaluation techniques. By the use of the weighting factors the influence of the objective trials to the final grade is increased, whereas the influence of the subjective trial is equally decreased.

Utilizing the unification formula researchers obtain a final grade for each website under evaluation. This final grade expresses degrees of penalty, thus the greater the grade, the more accessibility obstacles a website poses to its potential users. Final grades may be used for the construction of an accessibility scale in which websites that have undertaken may be positioned in reverse order according to their accessibility levels. Such an accessibility scale would present degrees of penalty for problematic websites according to their performance in the threefold accessibility evaluation.

7. CONCLUSIONS

As argued above, the evaluation platform utilizes existing automation technologies under a unification formula in order to bypass the methodological fragmentation which is observed in the relevant research field.

The growing corpus of research material suggests that up to now approaches to web accessibility evaluation techniques have diverged from each other, especially with respect to the dualism implied by the two main categories under which evaluation criteria may be classified: the criteria which may be automatically checked and the criteria which may only be checked by humans [2, 3].

W3C has provided the research community with general accessibility evaluation guidelines which – though not employable right from the shelf due to their general and guiding character– cover most of the aspects of the evaluation task and therefore may suitably be used for the development of integrated tools and techniques for web accessibility evaluation. The threefold approach, consisting of (a) the code evaluation, (b) the validity checks with respect to markup and style sheets and (c) a ratio scale checklist with respect to subjective criteria, bypasses the afore mentioned limitations and offers the researcher comparable results.

The major merits and improvements of such an approach –as opposed to the usage of automated only or purely subjective evaluation methods– may be summarized as follows:

- I) It integrates under one formula the different aspects of the so far proposed methods of evaluation.
- II) It provides uniform data for each website which facilitate inter-website comparisons and meta-research.
- III) It serves as the basis for the development of an abstract accessibility scale featuring arithmetically

defined ranges of accessibility levels, thus enhancing the already established W3C's levels of conformance.

Practical implementation of the proposed technique and tool has taken place within the framework of a comparative study regarding Greek and international senior-oriented websites with rather encouraging conclusions as regards the ability of the tool to provide univocal, comparable evaluation results. Though the threefold web accessibility evaluation platform –and especially the respective tool– is developed to a fully operation level, it should not be regarded as a fixed, irrevocable method to be used as proposed or not at all. Therefore, more than presenting a specific tool, this paper attempts to provide an operational paradigm as a means to support the development of more integrated, less subjective and uniform techniques for web accessibility evaluation. And this goal may be more easily accomplished if the research tools provide clearly comparable results, such as numeric data representing unambiguous levels of conformance to accessibility guidelines.

8. REFERENCES

- [1] Tania Lang, “Comparing website accessibility evaluation methods and learnings from usability evaluation methods”, in **Peak Usability**, [<http://www.peakusability.com.au/>], Accessed: May 2006.
- [2] John-Paul Mueller, **Accessibility for everybody: understanding the Section 508 accessibility requirements**, Berkeley: Apress, 2003.
- [3] Jakob Nielsen & Marie Tahir, **Homepage Usability: 50 Websites Deconstructed**, Indianapolis: New Riders Publishing, 2002.
- [4] OCAWA, **Accessibility Validator**, [<http://www.ocawa.com/>], Accessed: May 2006.
- [5] John Slatin & Sharron Rush, **Maximum accessibility**, Boston: Addison–Wesley, 2003.
- [6] World Wide Web Consortium, **Core Techniques for Web Content Accessibility Guidelines 1.0**, [<http://www.w3.org/TR/WCAG10-CORE-TECHS/>], Accessed: May 2006.
- [7] World Wide Web Consortium, **Web Accessibility Initiative**, [<http://www.w3.org/WAI/>], Accessed: May 2006.
- [8] World Wide Web Consortium, **Web Content Accessibility Guidelines 1.0**, [<http://www.w3.org/TR/WCAG10/>], Accessed: May 2006.

- [9] Panayiotis Zaphiris & Giorgos Zacharia, “Website Content Accessibility of the Cyprus Domain” in Yannis Manolopoulos, Skevos Evripidou & Antonis Kakas (eds) **Lecture Notes in Computer Science: Advances in Informatics**, Heidelberg: Springer Verlag, 2003, pp. 248-261.