

A COMBINED AGILE METHODOLOGY FOR THE EVALUATION OF WEB ACCESSIBILITY

Firas Masri and Sergio Lujan Mora

Department of Software and Computing Systems, University of Alicante, Spain

ABSTRACT

To assure and certify the fulfillment of web accessibility guidelines (*WCAG 1.0 & 2.0*) and to guarantee accessibility to all disabled users, various accessibility evaluation methods have been proposed, basically classified in two types: qualitative methods (*analytical and empirical*) and quantitative methods (*metric-based methods*). As no method by itself is enough to guarantee full accessibility, many studies advice to combine these methods between each other in order to guarantee better results. Some recent studies present combined evaluation methods between qualitative methods only, thus leaving behind the great power of metrics that guarantee objective results and task diversity. To achieve this goal, the current paper proposes a systematic combined agile accessibility evaluation method based both on qualitative and quantitative evaluation methods. This proposal presents an evaluation methodology that combines between the essential analytical evaluation methods and the empirical user test methods. Finally, WAB metric (*Web Accessibility Barrier*) is included to summarize objectively the final results and amplify the use of this method to cover all types of evaluations tasks, like validating, certifying and comparing processes.

KEYWORDS

Web accessibility, Empirical evaluation methodology, User test.

1. INTRODUCTION

Nowadays the web is present in all fields of life, from access to information and service web pages to management of legal documents by means of electronic public administration (*e-government*).

This makes that users will be a heterogeneous public, with different abilities and disabilities (visual, hearing, cognitive & motorize impairments). These characteristics represent a huge challenge if we hope to provide universal access to all possible users, specially if the intention is to fulfill web accessibility guidelines WCAG 1.0¹ and WCAG 2.0² (*Web Content Accessibility Guidelines*) of the WAI (*Web Accessibility Initiative*) or the other national and international guidelines and laws (Requisiti tecnici, 2005; Section 508, 2007; UE Commission, 2002) which regulate and protect the right of disabled users to access information.

To assure web accessibility, several studies have suggested numerous evaluation methods (Brajnik, 2006; Bühler, 2006; Vigo, 2007) as a means to verify, measure and certify the fulfillment of the accessibility guidelines and therefore to supply full accessibility to disabled people. Currently, there are two types of evaluation methods: the qualitative methods (analytical and empirical) and the quantitative methods.

The qualitative methods have been the most used until now, specially the analytical ones, which are characterized by their low cost and ease of use. Automatic evaluation tools such as ATRC,³ WAVE 4.0⁴, EvalAccess 2.0⁵, TAW⁶ and Cynthia Says⁷ have been the pioneers and the most well-known, due to its usability, ease of use and its quick results, although they are not the final and complete solution. The other

¹ Available at: <http://www.w3.org/TR/WCAG10/>.

² Available at: <http://www.w3.org/TR/WCAG20/>.

³ Web Accessibility Checker. Available at: <http://achecker.ca/checker/index.php>.

⁴ WAVE 4.0 - Web accessibility evaluation tool. Available at: <http://wave.webaim.org/>.

⁵ EvalAccess 2.0 - Web service tool for evaluating web accessibility. Available at: <http://supt07.si.edu.es/evalaccess2/index.html>.

⁶ TAW - Web accessibility test. Available at: <http://www.tawdis.net/>.

⁷ Cynthia Says evaluation tool. available at: <http://www.contentquality.com>

analytical evaluation methods, which are based on the manual heuristic inspection of code, do not guarantee full accessibility (Brajnik, 2008); it depends largely on evaluators' experience and the adopted guidelines. On the other hand empirical methods are generally more expensive, but more accurate, because they clearly show the most catastrophic accessibility faults. User test is the most reliable and complete one (Masri & Lujan, 2010a).

The quantitative methods help to understand, control and improve the final product (Fenton & Pfleeger, 1998), thus its main goal is to assure the quality results and monitor the accessibility level by establishing values and summarizing results. These methods and due to their nature aren't sufficient enough to assess accessibility and evaluators can't depend only on them.

As consequence of the above-mentioned situation related to evaluation methods and because no method on its own would guarantee the detection of all accessibility barriers, some studies like (Lopez, 2010; Villegas, 2010), considering W3C and researchers recommendations, starts to apply in their evaluations process a combined evaluation method based completely on qualitative methods (analytical and empirical methods). These works ignored totally the inclusion of the important quantitative methods to their methodology.

Due to this situation, and addressing the growing need of providing combined evaluation methods, arises the idea to present both kind of methods in a clear documented evaluating methodology, with a clear user – centered orientation, specially that there is no homogeneous and unified practice in this field (Masri & Lujan, 2010b), and those that already exist are not complete enough as to guarantee the coverage of all kinds of evaluations tasks from on side and does not provide a clear and objective results from other side.

The methodology presented in this paper combines the essential analytical methods with the empirical user test method and concludes with the WAB metric (*Web Accessibility Barrier*) (Parmanto and Zeng, 2005) that summarize objectively the results; thus permitting the coverage of all evaluation tasks.

2. ACCESSIBILITY EVALUATION METHODS

Qualitative methods include two essential types of evaluation methods: the analytical and the empirical ones. The analytical methods are: standard expert's revisions (conformity revisions), the automatic tools and barrier walkthrough method. On the other hand, the empirical methods are: user tests, subjective revisions, screen techniques (Masri & Lujan, 2010a).

As for the qualitative evaluation methods finality, they are used to undergo formative evaluations (identification of the list of problems) during the development phase; and summative evaluations (validation and comparison) in the final phase of the product and after the final users have used it. These methods estimate the accessibility of an interface so as to validate it. The results will always show the descriptions of the failure modes, defects or even solutions and recommendations for the developers.

On the other hand, quantitative methods are: Failure Rate metric, WEBQEM (*Web Quality Evaluation Metric*), WAB metric (*Web Accessibility Barrier*), UWEM (Unified Web Evaluation Methodology), A3 (Aggregation Metric), WAQM (*Web Accessibility Quality Metric*) and T1 metric. These methods helps evaluators to monitor and improve accessibility levels, also permits to summarize results objectively, their results can be used to compare quality among web pages, or to track quality improvement in the quality assurance process (Vigo et al, 2007; Freire, 2008; Sirithumgul, 2009).

2.1. Comparison between Web Accessibility Evaluation Methods

Evaluation methods whether qualitative (analytical & empirical) or quantitative have advantages and disadvantages when evaluating web accessibility.

Analytical methods are characterized by their great capacity of identifying a wide range of diverse problems for diverse audiences, apart from their ability of marking the exact violations of the adopted guidelines. However, they are criticized for requiring skilful evaluators and for not distinguishing between the important from unimportant web accessibility problems (Brajnik, 2008). Analytical methods have also proved methodologically weakness when qualifying the gravity of identified problems (Petrie & Kheir, 2007). A comparison between automatic tools also shows quite contradictory results (Thatcher et al, 2006; Diaz & Cachero, 2009).

Empirical methods are more exact when qualifying a web site; they discover the more catastrophic accessibility mistakes in a real time, especially if they are applied in the correct context taking into account the specific characteristics of the web accessibility explained in (Masri and Lujan, 2010a). Many studies (Dey, 2004; DRC, 2004) have shown that empirical methods and specially user tests are the unique ones that detect and qualify the severity of the real mistakes faced by users. They have shown also that 45% of accessibility difficulties emerge with the application of user tests to disabled people after being evaluated with software and other methods. However, they are criticized for their elevate cost due to necessity for equipped labs and real users.

On the other hand, the quantitative metrics are being used in large-scale evaluations processes; only the metric for WEBQEM did not. WAB, UWEM, A3, WAQM and T1 metrics have shown a great correlation between their results (Freire, 2008; Sirithumgul, 2009), for that reason, it is not possible to state which metric could be more effective in general cases. Each metric may be more suitable for different projects, according to their needs. In this sense, in order to help the definition of good metrics, (Daskalantonakis, 1992) it is needed to identify important characteristics of useful software metrics. According to Daskalantonakis, software metrics must be: (1) simple to understand and precisely defined; (2) objective; (3) cost effective; and (4) informative (ensure that changes to metric values have meaningful interpretations).

From all the above mentioned characteristics, we consider that a combined accessibility evaluation method must contain: (1) user test method (due to its reliability and effectiveness); (2) quantitative metric (to control and monitor accessibility results and cover all tasks types). Particularly, in our proposal, we consider that WAB metric presented by (Parmanto and Zeng, 2005) is the best metric that fits our approach due to its simplicity, objectivity and task coverage, beside, it has a fixed defined barrier weight that correspond to WCAG levels (A, AA & AAA). Also, it is considered inclusive because it includes all user groups.

3. THE COMBINED METHODOLOGY FOR EVALUATING WEB ACCESSIBILITY

When we talk about accessibility evaluation, first of all, we must think directly what our principal objective is and what we know about it. When assessing accessibility, we can face three different situations; the first is: evaluating a site during development in order to see if it fits all accessibility guidelines; the second is to validate or compare its compliance with accessibility guidelines after being adapted by developers and posterior to its launch; however, in the third situation, we don't know anything about the site we want to evaluate and we are obliged to start from the initial steps to check the real state of the site, which means from the analytical evaluation methods.

The first situation can be resolve by applying the common analytical automatic evaluation methods, while in the second case, it's more complicated and the best practice is to apply an empirical evaluation test, while in the third situation, many evaluators lose time to determine how to assess it. So taking into consideration the three possible cases, we propose two steps for the three goals commonly or separately.

- The analytical evaluation step
- The empirical evaluation step

3.1 The Analytical Evaluation Step

This step consists of two stages:

- **The pre analysis stage:** illustrate a rapid and clear idea about the site that is going to verify.
- **The automatic evaluation stage:** Helps the evaluator to have a complete idea about the site understudy and helps to determine whether the site is qualified to continue to the next step directly or not.

3.1.1 Pre Analysis Stage

This stage is composed of the following main activities:

- Localize and verify the identity of the site under evaluation to make sure that's the correct one
- Compare the actual version of the site with oldest version using www.archive.org, thus avoiding unexpected mistakes and having an idea about the sites evolution.

- Navigate the site in order to have a more complete idea about the used design.
- Apply software tools like Web Developer Tool to revise HTML cleanliness and detect image volume used in the design.
- Apply text and sound browsers like WebbIE and others to verify if information is accessible and equivalent to that provided by a graphical user interface browser.

3.1.2 Automatic Evaluation Stage

This stage is composed of the following main activities:

- Validation of mark-up language using *HTML Validation Service of W3C*.
- Validation of CSS style sheet using *CSS validation service*.
- Automated Accessibility evaluation using TAW or other available automatic evaluator software.

3.2 The Empirical Evaluation Step

Current legislations like (UE Commission, 2002; Section 508, 2007) establish priority/level AA as mandatory for all public and electronically products; so it was decided to consider two levels of test:

1. Test of level 1.
2. Test of level 2.

Test level 1 has to do with the conformity degree A and AA, where the biggest users groups will access information without troubles. On the other hand, test level 2 is related to the conformity degree AAA that is a higher level of fulfillment in which all users groups will access without any kind of barriers. It is suggested doing the test in just one day in order to reduce costs and increase the test control and systematization.

3.2.1 Users and Staff

The following users and staff are involved in an empirical evaluation:

- Users with different disabilities (Vision, hearing, motor & cognitive impairments), 6 are enough according to (Nielsen, 2000), 1 or 2 per each disability. Users must have some experience with internet and assistive tools.
- Facilitator, one person is enough to explain users the sequence of steps.
- Assistant or observer. An observer per disability is enough to guarantee the understanding and observation of the process.

3.2.2 Room and Equipment

It can be carried out in a small office or laboratory supplied with two computers for each selected disability. The computer must have the specific devices and software for the disability in question, for example: voice browsers, screen readers, a screen magnifier program, Jaws program, Synthesizer + Tiflowin software, etc. For more information about devices and software see (Pernice & Nielsen, 2001).

3.2.3 Test Planning

Test plan: Accessibility guidelines are prepared, classified and revised according to the test level (A & AA for test level 1 and AAA for test level 2). The web site must be studied to localize functionalities and the supposed potential barriers. For better results, it is advisable to assign tasks focusing on the fulfillment of the most essential functionalities with an implicit orientation towards each disability.

Material Preparation: The preparation of all tasks to be developed is undergone based on the test planning. The obtained material must be attractive and not exhausting for the user, with questions that put the site accessibility at stake. For example:

1. Browsing through the main menus, localizing the most attractive links (suggesting some of them and covering most of them).
2. Following a link predetermined by the evaluator (image-link).
3. In case there is a multimedia, asking for a summary of the multimedia subject.
4. Locating some essentials functionality and ordering user to access through the keyboard.
5. Filling up a form or sending a message through the page.

6. In case the site has sale service or donations options, asking the user to select one product and proceed to the purchase.

7. In case the site allows the formal step of legal documents, request the user to proceed to fill in the application form and document's formalities.

3.2.4 Required Tests

Pilot Test: It is performed to make sure that everything will turn out just as it has been planned and that the media work properly. The user does not have to be necessarily a disabled person.

Final test: The test is performed by the selected users in a familiar environment, free of tension. Users are asked to use the protocol, think aloud.

3.2.5 Qualification Methodology

A proposal of this property should be done through an approach of observation on one hand and a metric that establishes a value and summarizes the results on the other hand. The analysis system must be as it follows:

- **User's interview:** The interview results are compared with the corresponding observer's notes.
- **Recording analysis:** The recording is analyzed in order to find errors omitted by the observers.
- **Data analysis:** Errors found are analyzed and classified, taking into account the priorities and its corresponding checklists.
- **Conclusion:** A final report on the data analysis results is drawn up
- **Evaluations Metric:** The evaluation metric criteria come to be the test barometer (measure) that allows providing a concrete result. The formula results as it follows (Parmanto and Zeng, 2005):

$$WAB = \frac{\sum_{j=1}^T \sum_{i=1}^n \left(\frac{b_{ij}}{B_{ij}} \right) (W_i)}{T} \quad \text{Where,}$$

b_{ij} Represents the real errors and barriers in a priority

B_{ij} Represents the potential barriers of each page in such priority

W_i Represents the gravity of the real error inverse to each priority (1-2-3)

T Represents the number of pages of the whole site

This metric will be applied to the final report of the concluded data analysis results, it is considered best result those that is close or equal to 0.

4. CONCLUSION

The proposed combined method is clearly exposed step by step as a clear guide for web accessibility evaluations, the methodology ends with a clear and objective qualification method based in WAB score weighted metric formula that resume and validate the obtained results. This methodology characterized by its systematic and objective results with a user centered evaluation focus; it is systematic because all of its steps are developed formally and under observation, and considered objective because it resumes the results by mean of an evaluation metric and not subjectively with human opinion. The next step of our research will be the application of this method to evaluate different Spanish public and commercial websites, thus improving improper derived mistakes and including more user experience to this combined evaluation solution.

REFERENCES

- Brajnik, G., 2008. Beyond conformance: the role of accessibility evaluation Methods. *In 2nd International Workshop on Web Usability and Accessibility IWWUA08*, Auckland, New Zealand.
- Brajnik, G., 2006. Web accessibility testing with barriers walkthrough. Visited: Feb 2011. Available at: www.dimi.uniud.it/giorgio/projects/bw.

- Bühler, C., Heck, H., Perlick, O., Nietzio, A., and Ulltveit-Moe, N., 2006. Interpreting Results from Large Scale Automatic Evaluation of Web Accessibility. In *Computers Helping People with Special Needs*, Springer, and Berlin/Heidelberg.
- Daskalantonakis M. K., 1992. A Practical View of Software Measurement and Implementation Experiences within Motorola. *IEEE Transactions on Software Engineering*, 18(11):998 - 1010.
- Dey, A., 2004. Accessibility evaluation practices—survey, results. Visited: Feb 2011. Available at: <http://deyalexander.com/publications/accessibility-evaluation-practices.html>.
- Díaz, B., Cachero, C. 2009. Fiabilidad de herramientas automáticas para el aseguramiento de la accesibilidad en Web. In *IADIS Ibero-Americana WWW/Internet 2009*, Madrid, Spain.
- Disability Rights Commission (DRC). 2004. *Formal investigation report: web accessibility*. ISBN 0 11 703287 5.
- Fenton N. E., Pfleeger S.L. Software Metrics: A Rigorous and Practical Approach. PWS Publishing Co., Boston, MA, USA, 1998.
- Freire, A. P., Fortes, R., Turine, M., Paiva, D., 2008. An Evaluation of Web Accessibility Metrics based on their Attributes. *SIGDOC'08, September 22-24, 2008*, Lisbon, Portugal.
- López, L. AWA: *Marco metodológico específico en el dominio de la accesibilidad web para el desarrollo de aplicaciones web*. Tesis Doctoral, Universidad Carlos III, Leganés, Madrid. 2010.
- Masri, F., Lujan, S., 2010a. Análisis de los métodos de Evaluación de la Accesibilidad web. *Proceedings of the: Universidad 2010 conference*. Havana, Cuba.
- Masri, F., Lujan, S., 2010b. Test de usuario: Un Método Empírico Imprescindible para la Evaluación de la Accesibilidad Web. *Proceedings of the: 5ª conferencia ibérica de sistemas y tecnologías de información 2010*. Santiago de Compostela, España.
- Nielsen, J., 2000. *Why You Only Need to Test with 5 Users*. Visited March 2011, available at: <http://www.useit.com/alertbox/20000319.html>.
- Parmanto, B., & Zeng, X. M. 2005. Metric for web accessibility evaluation. *Journal of the American Society for Information Science and Technology*, 56(13), 1394-1404.
- Pernice, K., Nielsen, J., 2001. Beyond Alt Text: Making the web easy to use for users with disabilities. Visited: Feb 2011, Available at: <http://www.nngroup.com/reports/accessibility/testing/>.
- Petrie, H., Kheir, O., 2007: The relationship between accessibility and usability of websites. In: *Proc. CHI 2007*, pp. 397–406. ACM, CA
- Requisiti tecnici e i diversi livelli per l'accessibilit`a agli strumenti informatici, 2005. (G.U.n.1838/8/2005), Visited Feb 2011, www.pubbliaccesso.it/normative/DM080705.htm.
- Romen, D., Svanaes, D., 2008. Evaluating Web Site Accessibility: Validating the WAI Guidelines through Usability Testing with Disabled Users, *proceedings: NordiCHI 2008, October 20 to 22, 2008*.
- SEC.508, 2007. ELECTRONIC AND INFORMATION TECHNOLOGY, Visited Feb 2011, Available at: www.section508.gov/index.cfm?FuseAction=Content&ID=14.
- Sirithumgul, P., Suchato, A., Punyabukkana, P. 2009. Quantitative Evaluation for Web Accessibility with Respect to Disabled Groups.
- (SSWAE), 2011. Spanish Senate web accessibility evaluation. Available at: <http://accesibilidadweb.dlsi.ua.es/?menu=ej-analisis-senado-parte-1> y <http://accesibilidadweb.dlsi.ua.es/?menu=ej-analisis-senado-parte-2>.
- Thatcher, J., Burks, M., Heilmann, C., Henry, S., Kirkpatrick, A., Lauke, P., Lawson, B., Regan, B., Rutter, R., Urban, M., Waddell, C., 2006: Web Accessibility: Web Standards and Regulatory Compliance; Friends of ED.
- UE Commission., 2002. Visited: Feb 2011. Available at: http://ec.europa.eu/ipg/standards/accessibility/index_en.htm.
- Vigo, M., Arrue, M., Brajnik, G., Lomuscio, R., and Abascal, J. 2007. Quantitative metrics for measuring web accessibility. In *W4A '07: Proceedings of the 2007 international cross-disciplinary conference on web accessibility (W4A)*. Banff, Canada.
- Villegas, E., Pifarré, M., Fonseca, D., 2010. Diseño metodológico de experiencia de usuario aplicado al campo de la accesibilidad. *Actas de la 5ª conferencia ibérica de sistemas y tecnologías de información 2010*. Santiago de Compostela, España.