Web Not For All: A Large Scale Study of Web Accessibility

Rui Lopes
LaSIGE/University of Lisbon
Campo Grande, Edíficio C6
1749-016 Lisboa, Portugal
rlopes@di.fc.ul.pt

Daniel Gomes
FCCN
Av. do Brasil, 101
1700-066 Lisboa, Portugal
daniel.gomes@fccn.pt

Luís Carriço
LaSIGE/University of Lisbon
Campo Grande, Edíficio C6
1749-016 Lisboa, Portugal
lmc@di.fc.ul.pt

ABSTRACT

The Web accessibility discipline strives for the study and improvement of front-end Web design towards people with disabilities. Best practices such as WCAG dictate how Web pages should be created accordingly. On top of WCAG, several evaluation procedures enable the measurement of the quality level of a Web page. We leverage these procedures in an automated evaluation of a nearly 30 million Web page collection provided by the Portuguese Web Archive. Our study shows that there is high variability regarding the accessibility level of Web pages, and that few pages reach high accessibility levels. The obtained results show that there is a correlation between accessibility and complexity (i.e., number of HTML elements) of a Web page. We have also verified the effect of the interpretation of evaluation *warnings* towards the perception of accessibility.

Categories and Subject Descriptors

H.5.4 [**Information Interfaces and Presentation**]: Hypertext/Hypermedia—*User issues*; K.4.2 [**Computers and Society**]: Social Issues—*Assistive technologies for persons with disabilities*

General Terms

Measurement, Human Factors.

Keywords

Web Science, Web Accessibility, Web Characterisation, Quality Assessment, Automated Evaluation.

1. INTRODUCTION

Since its inception, the Web has been growing both in size and complexity. It is argued that this happens due to its decentralised properties: anyone can contribute to the Web without a central authority dictating what is allowed to be published or not. Consequently, people with insufficient technological skills to produce quality content (e.g., *usable*) became massive Web publishers. This makes the Web being perceived as a *living organism*, constantly evolving in different directions regarding its size and quality [1]. With the ever growing number of people interacting with the Web, minority groups such as people with disabilities become more representative and, consequently, the adequacy of Web sites to them becomes relevant. Due to the openness of Web technologies, there is no implied quality control ensuring that every Web page is accessible. In this paper we provide insights on the shape of accessibility on the Web, through the automatic evaluation of a large Web document collection. In the light of this study, we hypothesise how Web accessibility design and education should evolve.

W4A2010 - Communications, April 26-27, 2010, Raleigh, USA. Co-Located with the 19th International World Wide Web Conference.

2. BACKGROUND

2.1 The Portuguese Web Archive Initiative

The Portuguese Web Archive (PWA) is a project of a non-profit organisation named Foundation for National Scientific Computing[1] (FCCN), which has the mission of providing infrastructures to the national academic and research communities (e.g., managing the National Research and Education Network). The PWA periodically crawls contents mainly from the *.pt* domain (Portugal) and stores them into a repository. It also supplies a computational platform to researchers that wish to analyse the archived data. Besides supporting research, the PWA has a direct interest in contributing to monitor Web content quality.

2.2 Web Accessibility Evaluation

Several recommendations have been proposed on how Web technologies should be used without posing barriers to people with disabilities, such as WCAG, the Web Content Accessibility Guidelines [3]. WCAG defines a set of guidelines that should be followed by developers, designers, etc. when creating Web pages, in order to ensure a good level of accessibility for all users. Each guideline is composed by a set of verifiable checkpoints that creators must follow accordingly on Web technologies. WCAG can also be used as a base for evaluation, through different metrics. [4] discusses the strengths and pitfalls of existing metrics. These metrics depend on the single results obtained from the verification of checkpoints conformance, manually performed by experts or automatically evaluated through software. While expert evaluation provide an in-depth answer of the accessibility quality of a Web page, it poses problems on the scalability of the process, and brings potential comparison bias [6].

[1] http://www.fccn.pt

On the other hand, software evaluation, by being fully automated in its nature, has the benefit of scalability and objectivity. But since several checkpoints cannot be machine-verified, the evaluation is less detailed when compared with expert analysis. Nonetheless, the rate of non-verifiable evaluation failures is proportional to automated ones [9]. Having an automated way of evaluating the accessibility of Web pages opens the way to perform large scale analysis of Web accessibility. To our knowledge, no large-scale accessibility evaluations of the Web (and its evolution) have been performed before. We assume that this is due to the dependency of computational resources for large-scale analysis. To mitigate this problem, work is conducted on evaluating smaller scale collections of Web documents [9], typically based on sampling methods. However, there is always a significant sampling bias induced by these methods [2].

3. METHODOLOGY

We began our research with the question: *What is the shape of accessibility on the Web?*, as accessibility quality of Web pages is often less than desired. While this has been measured and studied several times before at a small scale, e.g. at the Web site level, there is still a lack of understanding of Web accessibility's macroscopic properties.

We defined a two step approach for data acquisition in this experiment. First we obtained one Web document collection, as detailed in [5], and then we performed an accessibility evaluation on each of the Web documents in the collection. We implemented 39 checkpoints from WCAG 1.0 (from priorities 1 and 2) based on the suggested implementation of UWEM [8]. Each checkpoint evaluation result was classified as: (1) *PASS*: it is applicable to an HTML document *and* its compliance is verified; (2) *FAIL*: it is applicable to an HTML document *and* its compliance unachieved; and (3) *WARN*: it is applicable to an HTML document *but* it is impossible to verify its compliance.

For this study, we defined three evaluation metrics based on the *failure rate* metric by Sullivan & Matson [7] to verify different aspects of the Web accessibility quality of a given Web page. Each criterion yields a percentage, with the semantics from *not accessible* to *fully accessible*. We have defined our automated evaluation process with the distinction of whether a checkpoint evaluation of a given HTML element results in *PASS*, *FAIL*, or *WARN*, as follows.

Conservative rate: *WARN* results are interpreted as *failures*. The semantics of this rate conveys the worst-case scenario on accessibility evaluation:

$$rate_{conservative} = \frac{passed}{applicable} \quad (1)$$

Optimistic rate: *WARN* results are interpreted as *passed*. This rate is related to a best-case scenario where developers and experts dismiss warnings (often incorrectly, as explained in [6]) as accessibility issues that were taken into account:

$$rate_{optimistic} = \frac{passed + warned}{applicable} \quad (2)$$

Strict rate: *WARN* results are *dismissed* (thus accounting only the actual *FAIL* results):

$$rate_{strict} = \frac{passed}{applicable - warned} \quad (3)$$

4. RESULTS

The evaluation of the Web document collection has yielded results for 28,135,102 Web pages, out of 48,718,404 contents (e.g., images, PDFs, etc.) that have been crawled in total (nearly 58%), spanning through 21GB of data. This evaluation resulted on a total 40,831,728,499 HTML elements that were analysed, with an average of approximately 1451 HTML elements per Web page. Of these, 1,589,702,401 HTML elements successfully met all applicable Web accessibility criteria, an average of 56 HTML elements per Web page (around 3.89%). On failures, 2,918,802,078 HTML elements failed to comply, corresponding to an average of more than 103 errors per Web page (approximately 7.15%). Finally, 36,323,224,020 HTML elements have triggered *warnings*, accounting for an average 1291 per page (nearly 89%).

4.1 Distribution of Rates

We aggregated the accessibility quality rate by permillage values, and performed a count of how many Web pages belong to each aggregation. Figure 1 presents a linear-log plot for the distribution of *conservative rate* metric versus page count. Since all *warnings* are interpreted as errors, and no Web page was missing the HTML elements detectable in the checkpoints that yield warnings, no Web page was able to reach the maximum value of accessibility quality. The depicted exponential decay starts around 5% of compliance, where the number of pages with good quality is minimal.

Figure 2 presents a linear-log plot for the distribution of *optimistic rate* metric versus page count. Since this metric takes into account all *warnings* as positively complied, all checkpoints that cannot decide on the their own criteria have a significant positive effect on its page count distribution. Here, we observed that there is a rapid progression of the number of pages for each aggregated rate, with a lower bound of accessibility quality around 50% and a 90% mean.

When analysing from the perspective of the 100% detectable problems (i.e., *errors*), we found that the there is a near constant distribution of Web pages according to their accessibility quality, as depicted in Figure 3. The only exceptions on this are at the edges of the distribution, especially when approaching fully compliance with the detectable errors. Here, the decay on the page count is significant, despite the fact that it is less steep comparing to *conservative rate*.

4.2 Rates and Page Complexity

Our second incursion on this study relates to verifying if there is a correlation between the rate and complexity of each Web page. We have defined the criterion of page complexity as the number of HTML elements present in a Web page, encompassing both the breadth and depth of the Web pages' HTML node tree.

Regarding *conservative rate*, with the exponential decay of node count (i.e., HTML elements), the accessibility rate approaches 10% quality, as presented on Figure 4. When taking into account the *optimistic rate* metric, there is no obvious correlation between node count and accessibility quality, as depicted on Figure 5. Nevertheless, there is a homogeneity on the distribution of *optimistic rate* regarding node count. Lastly, Figure 6 depicts the distribution for the *strict rate* metric. Like in the *conservative rate* metric, we have discovered the same kind of exponential decay between node count and the metric. However, in this case, the rate approaches 100%, since *warnings* were dismissed.

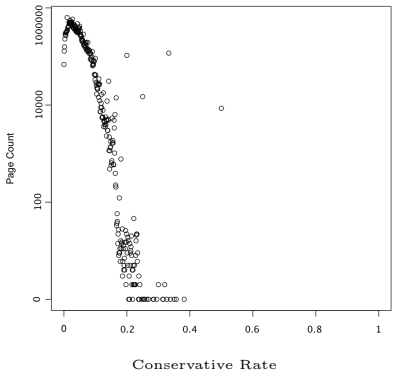

Figure 1: Accessibility distribution for *Conservative rate*

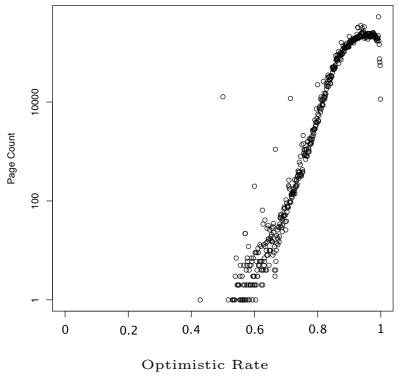

Figure 2: Accessibility distribution for *Optimistic rate*

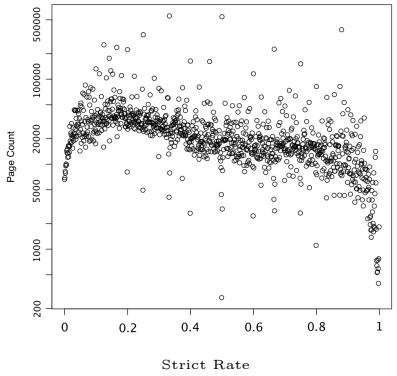

Figure 3: Accessibility distribution for *Strict rate*

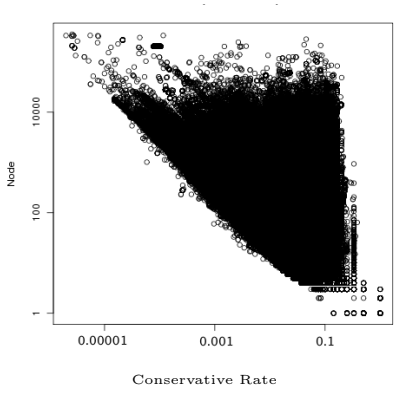

Figure 4: Accessibility *conservative rate* versus page complexity

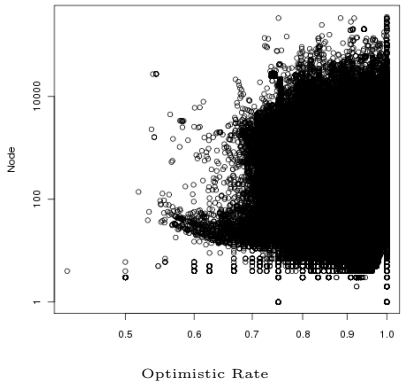

Figure 5: Accessibility *optimistic rate* versus page complexity

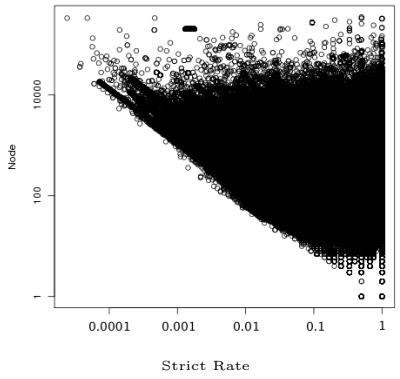

Figure 6: Accessibility *strict rate* versus page complexity

5. DISCUSSION

One of the interesting aspects of this experiment is the distribution of rates, according to the three metrics (*simple*, *positive*, and *ignore*). When looking at *errors* distribution (i.e., *strict rate* metric), its linearity implies that critical accessibility problems are likely to be encountered with the same probability by end users who depend on proper accessibility. However, when taking into account *warnings*, the picture of accessibility on the Web is not clear. When *warnings* are perceived as positive, accessibility quality quickly reaches high levels. But, as discussed by *Vigo et al.* [9], the rate of warnings goes hand in hand with error rate. This result has the direct consequence that indeed such small scale studies have been verified at the large scale.

Another important result from this experiment concerns the relationship between the number of HTML elements in a Web page (i.e., node count) and its accessibility quality. While applying the *optimistic rate* metric is insufficient to reach a significant conclusion, this changes in what respects to both *conservative rate* and *strict rate* metrics. In both cases, we have discovered that a high node count on a Web is directly related to its accessibility quality. There was no single Web page in the evaluated document collection that had both a small node count and a poor accessibility quality rate. We hypothesise that this happens due to the complexity of Web pages: simplicity leaves out several HTML structural compositions that hinder accessibility, and also that a smaller Web pages are more manageable.

5.1 Impact on Designing Accessible Web Pages

The results of our experiment can also be discussed towards more practical matters, i.e., how people who create Web pages (e.g., designers, developers, etc.) can mitigate the recurring accessibility problems encountered on the Web.

Paying attention to detail in the structure, rhetoric and discourse of a Web page conveying information is critical for its accessibility success. Some of the *warnings* raised by the evaluation process concern the lack of usage of HTML structural elements that help building the discourse of a Web page. Therefore, we believe that there is a strong need for a better education and dissemination of best practices for properly using the semantics of HTML elements.

Another issue concerns the aforementioned problem of the relationship between Web accessibility quality and the complexity of Web pages. Our position on this issue, in what respects to designers and developers, is that Web accessibility is more manageable in *smaller chunks*. Our advice to Web page creators is to follow a *simplicity* approach on defining the structure of Web pages, which lowers the burden of verifying accessibility compliance during development.

5.2 Impact on the Perception of Accessibility

Our results show the profound difference between the op-

posite perspectives of accessibility given by the *conservative* and *optimistic* rates. Overall, the *conservative* results are in pair with the *strict* analyses performed. This discovery confirms the expectations and model followed by the WAQM accessibility metric [9], in that errors and warnings tend to occur proportionally.

On the other hand, when comparing with the *optimistic* rate, it shows what developers and designers might interpret the accessibility quality of the Web sites they create, i.e., having an *optimistic* view. Since most developers and designers are not accessibility experts, and since non-experts tend to incorrectly evaluate accessibility [10], we hypothesise that the *optimistic* rate might shed light on the real perception of accessibility by non-experts at the large. This discrepancy further shows that guidelines and encompassing evaluation procedures are just starting points for proper accessibility adequacy. Consequently, we believe that improvements must be made on communicating guidelines and presenting accessibility evaluation results to motivate developers and designers on investigating the nature of accessibility evaluation *warnings*.

5.3 Limitations of the Experiment

The depth of evaluating Web accessibility with algorithms is shallower than that of expert evaluation or usability evaluation with people with disabilities. Therefore, the results we have presented and discussed are based on the existence of a *detectability* upper limit of accessibility quality.

Furthermore, there might be a difference between the Web page being served – its HTML and associated resources – and its rendered layout on a Web browser. The flexibility of CSS allows Web designers to create rich layouts that can result on a different rhetoric being conveyed on a Web page. Along the same lines, the ever increasing use of AJAX also poses limitations to automated evaluation.

The evaluated document collections are also limited by Web page crawling capabilities (named *spider traps*), including difficulty of reaching the *deep Web* through HTML forms and AJAX, infinite generation of Web pages through server-side scripts, *robots.txt* exclusion protocol, etc.

6. CONCLUSIONS AND ONGOING WORK

This paper presented a large-scale study of accessibility on the Web conducted over a Web document collection provided by the Portuguese Web Archive. We have discovered the effects of Web page quality in what respects to accessibility, and how it hinders the expected universality aspects of the Web. One of the aspects studied leveraged the confirmation that simpler, smaller Web pages tend to have a better accessibility quality. We hypothesise that is due to providing less margin of error for Web designers and developers. Our results also show that accessibility communication must be further improved. This was shown through the disparity between *conservative* and *optimistic* perspectives over Web accessibility evaluation results.

This work is the beginning of a series of studies about the accessibility quality of the Web and, therefore, ongoing work is being conducted in studying different facets of evaluation of accessibility at large scale, including (1) comparing document collections from different years to study the evolution of the Web in what respects to accessibility compliance, and (2) studying vertical cross-cuts of document collections, such as site aggregation, government, etc.

7. ACKNOWLEDGEMENTS

This work was funded by Fundação para a Ciência e Tecnologia (FCT) through scholarship SFRH/BD/29150/2006, the Multiannual Funding Programme and POSC/EU. The authors would like to thank to the Portuguese Web Archive members at FCCN for their tireless efforts.

8. REFERENCES

[1] T. Berners-Lee, W. Hall, J. A. Hendler, K. O'Hara, N. Shadbolt, and D. J. Weitzner. A Framework for Web Science. *Found. Trends Web Sci.*, 1(1):1–130, 2006.

[2] G. Brajnik, A. Mulas, and C. Pitton. Effects of sampling methods on web accessibility evaluations. In *Assets '07: Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, pages 59–66, New York, NY, USA, 2007. ACM.

[3] B. Caldwell, M. Cooper, W. Chisholm, L. G. Reid, and G. Vanderheiden. Web Content Accessibility Guidelines 2.0. W3C Recommendation, World Wide Web Consortium (W3C), December 2008. `http://www.w3.org/TR/WCAG20/`.

[4] A. P. Freire, R. P. M. Fortes, M. A. S. Turine, and D. M. B. Paiva. An evaluation of web accessibility metrics based on their attributes. In *SIGDOC '08: Proceedings of the 26th annual ACM international conference on Design of communication*, pages 73–80, New York, NY, USA, 2008. ACM.

[5] J. Miranda and D. Gomes. An Updated Portrait of the Portuguese Web. In *14th Portuguese Conference on Artificial Intelligence (EPIA 2009)*, Aveiro, Portugal, October 2009.

[6] D. Sloan, A. Heath, F. Hamilton, B. Kelly, H. Petrie, and L. Phipps. Contextual web accessibility - maximizing the benefit of accessibility guidelines. In *W4A '06: Proceedings of the 2006 international cross-disciplinary workshop on Web accessibility (W4A)*, pages 121–131, New York, NY, USA, 2006. ACM.

[7] T. Sullivan and R. Matson. Barriers to use: usability and content accessibility on the web's most popular sites. In *CUU '00: Proceedings on the 2000 conference on Universal Usability*, pages 139–144, New York, NY, USA, 2000. ACM.

[8] E. Velleman, C. Meerveld, C. Strobbe, J. Koch, C. A. Velasco, M. Snaprud, and A. Nietzio. Unified Web Evaluation Methodology (UWEM 1.2), 2007.

[9] M. Vigo, M. Arrue, G. Brajnik, R. Lomuscio, and J. Abascal. Quantitative metrics for measuring web accessibility. In *W4A '07: Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A)*, pages 99–107, New York, NY, USA, 2007. ACM.

[10] Y. Yesilada, G. Brajnik, and S. Harper. How much does expertise matter?: a barrier walkthrough study with experts and non-experts. In *Assets '09: Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*, pages 203–210, New York, NY, USA, 2009. ACM.