
Characterizing Cloud SaaS Accessibility of E-Governance websites

B.Gohin^{#1}, Dr. Viji Vinod^{#2}...

#1 Department of Computer Applications, Dr.MGR Educational and Research Institute
University, Chennai – 95, Tamil Nadu, India

#2 Dr. Viji Vinod, Research Supervisor, Professor & Head of the Department, Department of
Computer Applications, Dr.MGR Educational and Research Institute University, Chennai – 95,
Tamil Nadu, India

ABSTRACT

Characterizing the cloud computing software service accessibility is a nascent research area. It is known that developers and designers are largely unaware of accessibility practices, despite the efforts to educate and advise them with standards and guidelines such as WCAG and Section 508. Furthermore, when content production emerges as a common practice for end users, such as the case of e-governance, additional efforts should be taken, so that all content is accessible to people with disabilities. In the light of these problems, this research presents experimental studies over accessibility quality of e-governance sites. Objective of this study is centered on an initial foray on automated Web accessibility evaluation of a significant number of e-governance web pages (over 2000), in order to understand the main problems of automating these evaluation practices, including potential scalability issues.

Key words: Cloud Computing, E-Governance, SaaS, Accessibility, Disabilities, W3C, WCA;

Corresponding Author: Gohin B

INTRODUCTION

Due to the decentralized nature of the cloud software service's architecture, hyperlinks can be used as navigable and verifiable forms of making explicit the credibility of a Web page's content. Readers can follow these hyperlinks and judge by themselves whether the information they are reading is correct, especially since the linked information might have been created by authoritative and reputed sources [1]. The establishment of this type of links is based on the assumption that reputed sources are trusted sources of confirmation from the point-of-view of readers. Independent verifiability of human-created content is at the core of its credibility [2]. Therefore it is of the uttermost importance to ensure that readers can access and understand information, otherwise it will be too cumbersome to verify it. While the content itself is at the core of credibility, the interface between content and the reader must be adequate. In other words, the readers must be able to access reputed sources without any kinds of barriers. For example, in a scenario where a Web page links to a supposedly authoritative source with a Web

page that uses images to convey information, blind readers will not be able to grasp that information and, therefore, the credibility associated to this linking scenario is inherently flawed for this group of readers [1][2].

The ability to obtain accurate government services online quickly and conveniently. Little is known, however, about whether this online information is accessible to people with disabilities that have to rely on special devices or technologies to process online information due to their sensory, mobile, or Cognitive disabilities [3].

Large gap in Cloud Computing usage can be attributed in large part to problems with accessibility. A study by the Nielsen & Norman Group found that the usability of the Web is about three times better for users without disabilities than it is for users with disabilities [4].

For people with disabilities, the Web very often is the only source of information and knowledge to which they may have access without having to depend unduly on others. Greater access to government services over the Cloud Computing will open a door to people with disabilities by offering them the exciting possibilities for independence and community participation by overcoming environmental barriers [4]. An accessible e-governance website will serve to assist in raising the community participation, independent living, and quality of life of people with disabilities [5].

Next, this research details an experiment performed over the accessibility of e-governance sites and the external links embedded in them. This experiment was formulated to verify the extent of hyper linking and template mechanisms, as well as the accessibility levels of e-governance sites.

EVALUATION METHODOLOGY

To find the answers for these problems, an accessibility evaluation framework was created. This framework was based on the UWEM methodology [6]. Which provides directions towards the automation of the evaluation process(which becomes relevant due to scalability issues). The framework implements a subset of checkpoints from WCAG (the Web Content Accessibility Guidelines) that can assess particular accessibility issues of Web pages' HTML structures [7] each checkpoint is responsible for analyzing a specific detail on these structures (e.g., image tags without alternate text attributes). Afterwards, the outcome of each checkpoint contributes to a final, (quantitative) score representative of the accessibility level of the evaluated Web page. This experiment follows closely the setup presented below:

- A set of 150 e-governance sites was randomly selected through a special purpose URL provided by Google search engine.
- Each e-governance sites individual Web page was analyzed in order to extract all hyperlinks. Thus, we excluded all pages from e-governance site's sub domains, such as hyperlinks to the web page's equivalent in other languages, discussion pages, etc.
- A pre-processing task was applied to each Web page (more specifically JTidy2), allowing the verification of standard-conformance (i.e., whether it respects HTML

specifications), and pushing the application of healing processes on ill-formed Web pages (akin to what Web browsers do, as well), for structural analysis purposes;

• Finally, each e-governance site was evaluated against a set of accessibility checkpoints $C = \{c_1, \dots, c_n\}$. When combining the results yielded from this process into a quantifiable result A , we opted to apply a simple evaluation metric adapted from Sullivan & Matson [8] presented in Equation 1

$$A = \frac{1}{n} \sum_{i=1}^n c_i, c_i \in C \quad (1)$$

Here, each checkpoint c_i returns 1 if its corresponding assessment was successful when applied to the Web page that is being analyzed, or 0 if not. Consequently, the overall evaluation result A averages all checkpoint values in the $[0; 1]$ range (i.e., a percentage), stating the accessibility quality level for the analyzed Web page.

RESULTS

By choosing 150 randomly selected e-governance sites from Google, a total set of 2000 Web pages were crawled. Out of these, 100 were randomly selected, and 2691 hyperlinks were followed. From the followed hyperlinks, 2211 were targeted to other Web pages within e-governance sites, whereas 480 hyperlinks targeted to Web pages outside the scope of e-governance site. Table 1 shows the overall results for accessibility assessment:

Table 1: Overall accessibility assessment

	Number	Total	%
Webpages correct	272	2725	9.98
Checkpoints correct	5	14	35.71

From these results, we verify that only nearly 10% of the e-governance website's pages fully complied with the checkpoints. We expect this situation to become worse when applying more complex and thorough evaluation procedures. Out of the 14 evaluation checkpoints analyzed, only 5 were fully complied by all Web pages, based on UWEM tests: no blink elements, no marquee elements, no page refreshes through the meta element, the assurance that all field set elements have a legend child element, and the assurance that all optgroup elements have a label attribute. When applying the accessibility metric defined above, we have an average quality of accessibility for each page of 84.6%. The minimum accessibility quality obtained from the evaluated pages was 50% (i.e., passing 7 checkpoints), whereas the maximum was fully compliance.

When splitting the analysis between internal and external Web pages, the results were different from the average values presented. Tables 2 and 3 further detail these findings:

Table 2: Internal Web pages accessibility assessment

	Number	Total	%
Webpages correct	250	2311	10.81
Checkpoints correct	5	14	35.71

Table 3: External Web pages accessibility assessment

	Number	Total	%
Webpages correct	22	414	5.31
Checkpoints correct	5	14	35.71

Analyzing both tables, it becomes relevant to know the fact that while the number of checkpoints passed remains the same, the number of Web pages which are fully compliant with all the 14 checkpoints evaluated is quite different between both clusters, with the ratio of 2:1 (internal: external). This is due to the fact that the overall HTML structure of e-governance sites complies with more checkpoints than the external Web pages linked from it. Moreover, the fact that e-governance sites provides a simplified markup language, might provide additional support to these values. Wikipedia's standardized environment facilitates the usage of accessibility aids that require well-structured Web pages. Based on these results, Table 4 presents a comparison between both web page clusters based on the accessibility metric defined above. Once again, the ratio of compliance between e-governance Web pages and externally referenced Web pages differs almost by 8%, and the minimum expected quality for any Web page within e-governance sites is more than 7% higher than externally references Web pages.

Table 4: Internal vs. external Web pages accessibility assessment

%	Internal	External
Average	89.79	81.83
Maximum	100.00	100.00
Minimum	57.14	50.00
Standard deviation	8.68	9.66

Lastly, another analysis was performed based on clustering Web pages between the initially 100 Home Web pages and all the Web pages linked from these. Tables 5 and 6 present their respective summaries:

Table 5: Home Web pages accessibility assessment

	Number	Total	%
Webpages correct	13	100	13.00
Checkpoints correct	8	14	42.86

Table 6: Child Web pages accessibility assessment

	Number	Total	%
Webpages correct	259	2625	9.95
Checkpoints correct	5	144	35.71

Both tables yield corollaries from the previous findings: if a random hyperlink is followed from a home Web page, there is always the possibility of finding a Web page which cannot even be parsed (as child Web pages include externally linked Web pages). However, when comparing the accessibility evaluation between both clusters, as presented in Table 7, more results can be studied.

These numbers help on reflecting about the exploratory nature of interacting with large content, highly linked websites. If a user wants to follow a hyperlink from a e-governance web site, there is no guarantee the linked Web page has a higher or lower accessibility quality.

However, it is mostly expected that, due to the wilderness nature of the Web vs. the highly structured and templated nature of e-governance site, that there is a concrete probability that the minimum quality of accessibility will decrease significantly.

Table 7: Home vs. child Web pages accessibility assessment

%	Home	Child
Average	84.14	84.65
Maximum	100.00	100.00
Minimum	64.29	50.00
Standard deviation	9.27	8.76

Figures 1 and 2 these findings are further detailed. In both cases, the X-axis represents the accessibility quality level (in percentage, based on the accessibility metric previously defined) for Web pages, whereas the Y-axis represents the number of Web pages that meet those quality levels.

The shape of the graph for home Web pages depicts the aforementioned stability of the accessibility quality. The “spike” centers around 80%, with a sharp decrease towards lower quality Web page count, and a more stable decrease on higher quality Web page count.

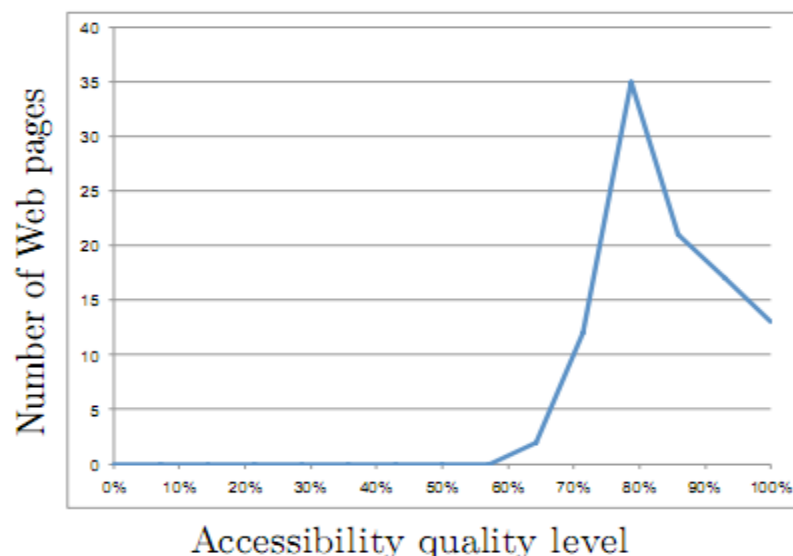


Figure 1: Accessibility quality distribution on home Web pages

Regarding the shape of the graph for external Web pages, a higher variation of quality is perceived. First, the initial spike shows that there is a high number of Web pages that do not meet the minimum requirements. This typically happens when pages are either totally inaccessible (e.g., HTTP 404 errors) or have severe parse errors. It is worth mentioning that the number of Web pages in this situation is 1.5 times more than the average Web page count for the large amount of external Web pages analyzed. From around 50%, the quality level distribution steadily increases until stabilizing between 70% and 90%. Finally, the quality level decreases abruptly once again for the highest percentile (between 90% and 100%), which contrasts to the results of home pages of e-governance sites.

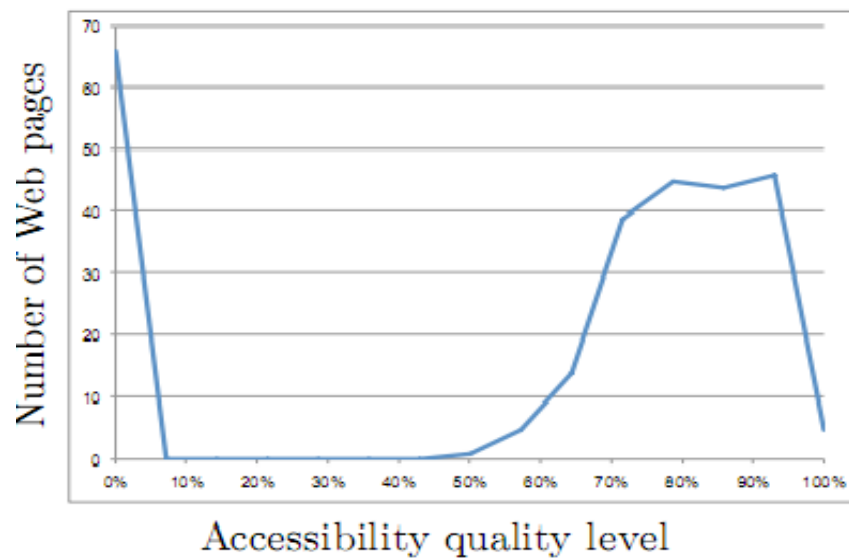


Figure 2: Accessibility quality distribution on external Web pages

Another aspect that has been analyzed in this experiment concerns the number of errors and warnings yielded by the HTML parser (i.e., HTML wellformedness). No e-governance sites processing has resulted on an error, which satisfies the high availability and good baseline quality of the cloud based software service platform and its instantiation into e-governance sites.

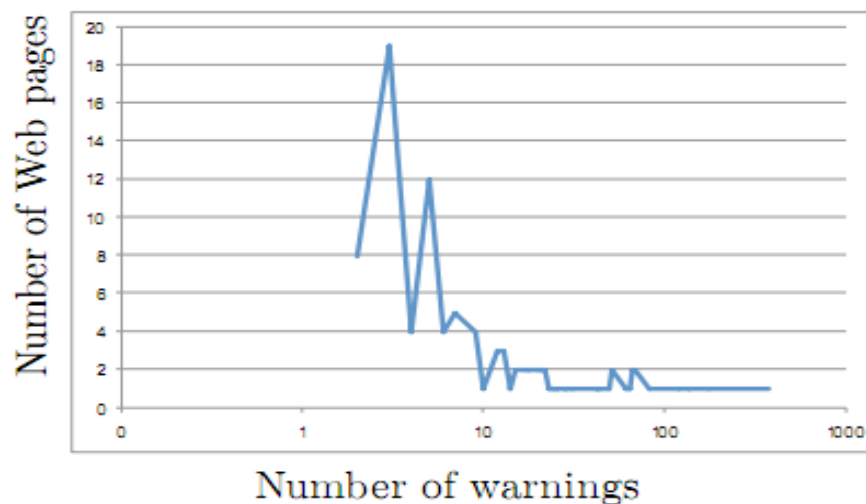


Figure 3: Warnings distribution on home Web pages

On the other hand, the high variability and uncontrolled nature of the Web has influenced the referencing mechanism used by e-governance sites contributors. 66 external Web pages have resulted on errors, which accounts for nearly 25% of all external Web pages analyzed. This aspect is manifested as the initial spike of the accessibility quality distribution presented in Figure 2. The analysis of the home e-governance sites has yielded 2323 warnings total, which accounts for an average of 23 warnings per page. In contrast, external Web pages have summed

up to 19523 warnings total, on an average of almost 74 warnings per Web page. This 31.5% increase on warnings is another indicator of the accessibility discrepancies between e-governance sites and the Web pages used for citation and referencing purposes.

A deeper analysis of this part of the experiment is presented in the graphs depicted in Figures 3 and 4. Each of these graphs presents on the X-axis the number of warnings on a Web page (represented in a logarithmic scale), and, on the Y-axis, the number of Web pages that have those warnings.

Regarding the distribution of warnings for external e-governance sites the figure changes dramatically. The wilderness of the Web pays a significant role on the different distribution of warnings, which might be smoothed only by the influence of contributors on choosing the reference Web pages on e-governance sites. The first interesting fact on this analysis is that there is a significant number of Web pages that have little to no warnings. However, this figure tends to shape differently afterwards. The biggest part of external Web pages orbit around 100 warnings, an order of magnitude above to those of e-governance sites.

The main outcome of the conducted experiment concerns the high variability of the accessibility quality of external Web pages, i.e., ranging from no accessibility (or even unavailability) to an average distribution between 50% and 100%.

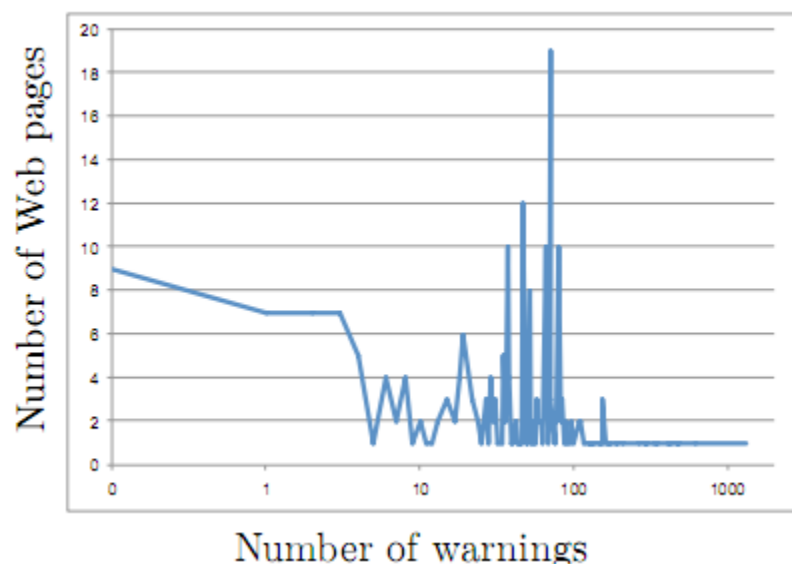


Figure 4: Warnings distribution on external Web pages

CONCLUSION

The results of these studies revealed a set of characterizations of Web accessibility, both on how it impacts users and how Web page creation hinders it. 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

e-governance sites tend to have a better accessibility quality. We hypothesize that is due to providing less margin of error for web page designers and developers. Our results also show that accessibility communication must be further improved. This was also shown through the poor compliance levels of WCAG Checkpoint 1.1 (alternative texts for media elements), as well as on the disparity between conservative and optimistic perspectives over cloud SaaS accessibility evaluation results. This study also uncovered how the fundamental hyper linking properties of the cloud software service differ when taking account accessibility.

REFERENCES

1. Kaeru., 2012, People with Disabilities Report [Online], <http://cis-india-org/about/accessibility> (accessed 20 November 2012).
2. Williams, B., Copestake, P., Eversley, J., & Stafford, B. (2007) "Experience and Expectations of Disabled People: A Research Report for the Office of Disability Issues", Office for Disability. [online] <http://www.officefordisability.gov.uk/research/pdfs/eedp-full-report.pdf>.
3. Rovira, C., Marcos, M., Codina, L. (2007) "Repositorios de publicaciones digitales de libre acceso en Europa: análisis y valoración de la accesibilidad, posicionamiento web y calidad del código", *El profesional de la información*, Jan/Feb 2007, Vol16, No. 1, pp. 24-38.
4. Nielsen, J., (2001), Beyond Accessibility: Treating Users with Disabilities as People , Alert box. Friedman, M. and Bryen, D., (2009) "Web Accessibility Design Recommendations for People with Cognitive Disabilities", *Technology & Disability*, Vol 19, No. 4, pp 205-212.
5. Foley, P., Alfonso, X., Fisher, J., Bradbrook, G. (2005) "eGovernment: Reaching socially excluded groups?", Report prepared by IECRC and Citizens Online for the IDEa. [online] <http://www.idea.gov.uk/idk/core/page.do?pageId=1074872>
6. Velleman, E., Meerveld, C., Strobbe, C., Koch, J., Velasco, C.A., Snaprud, M. & Nietzio, A. (2007). Unified Web Evaluation Methodology (UWEM 1.2). 46, 51, 61, 72, 90, 107
7. Google, <http://www.google.com>. Accessed on December 2012.
8. Sullivan, T. & Matson, R. (2000). 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, 139–144, ACM, New York, NY, USA. 51, 62, 73