An Evaluation on Accessibility of Websites Based on Popularity

Alberto Rodriguez
Department of Computing Science
School of Natural and Computing Sciences
University of Aberdeen
Student ID - 52092176
u14ar20@abdn.ac.uk

Abstract

Web accessibility is a big consideration for designers, both legally but also due to the perceived parallels in effective accessibility design and search engine optimisation. This report finds that in fact accessibility of two samples featuring 100 websites of high and low popularity respectively did not show a statistical difference. Web accessibility may not play as large of a part in popularity of websites and instead other factors play a much larger role in garnering large user attention.

1 Introduction

Starting from the first web content accessibility guidelines established in 1999, accessibility all over the internet has been a greater focus for developers and innovations and exploration into various different technologies and techniques have expanded the number of people who can access web content [1]. The rise of developing with accessibility in mind has come due to a variety of factors such as the accessibility practice guidelines outlined in the web accessibility access guidelines (WCAG) 2.2 as well as legal repercussions to inaccessible services [2] in an effort to make the web accessible to all 1.3 billion people .

As a result, there have been a variety of tools and approaches to aid in developing accessible content. There is consensus that effective accessibility helps in search engine optimisation (SEO) which helps in bringing people to a website [3]. While it is widely believed that good accessibility helps in SEO there are many other factors such as keyword selection that can also play a factor [3], [4]. This research sought to explore the relationship between website popularity and accessibility as well as the differences in accessibility across government and non-government domains.

1.1 Research Questions

Given the rapid change and focus on accessibility of web services over time by governing bodies worldwide [5], an evaluation of webpages have been conducted to measure the accessibility with relation to popularity and types of websites in an attempt to evaluate if a website's popularity is directly correlated with its accessibility. The hypothesis was that more popular websites have greater accessibility than their less popular counterparts due their better SEO and better usability [6]. In terms of different types of websites the hypothesis was that government and websites for public services are more accessible than their private company counterparts due to governments having stricter guidelines in terms of WCAG standards in many places around the world [7]–[11].
In short the two main questions the research aimed to answer were:

- Are more popular websites more accessible?
- Are government websites more accessible than those in the private sector?

2 Background and Related Work

2.1 Web Accessibility Guidelines

Accessibility on the web has always been a widely studied issue ranging as far back as the late 90s. In the early stages of the world wide web, accessibility guidelines varied or were not considered much [12] until the first W3C recommended guidelines WCAG 1.0 were published. In the years that followed these guidelines were extended and amended. The current W3C minimum required guidelines are WCAG 2.1 [13]. Within the WCAG guidelines there are a number of levels, denoted A, AA or AAA. Each level denotes an increase of requirements in accessibility and different guidelines than other levels. This means that A would be least strict and AAA being most strict [14]. If a web service does not meet the minimum accessibility guidelines they are open to serious legal consequences, which happened to American company Target in 2006 when a class action lawsuit was filed by the National Federation of the Blind [15].

2.2 The Need for Automatic Evaluation

Due to the importance of having proper accessibility there has been a push from developers and legislators in evaluation metrics to measure how accessible a website or web service is [16]. During the infancy of accessibility evaluation there were many attempts in developing tools and methods to spot accessibility issues, such as Bobby, Tawdis or WebXACT however it was found by Luque Centeno, Delgado Kloos, Arias Fisteus, *et al.* [17] that there were different interpretations of the the same WCAG guidelines leading to confusion. Among the many tools there are various metrics to quantify accessibility, such as the web accessibility barrier score metric devised by Parmanto and Zeng [18]. During recent years there are newer tools such as WAVE[1] and Accessibility Checker[2] as well as tools that seek to leverage artificial intelligence (AI) to spot issues such as AccessiBE[3]. However, with the current state of AI capabilities there are many issues and challenges to these tools which still require human evaluation [19], [20]. While research concludes that the need for manual and real user testing has not been fully abolished [21] there are many benefits to developers in using such tools as an aid in development and development, for example the fact that SEO and accessibility both benefit from the same implementation decisions leading to a potential increase in popularity of a website [3], [22].

3 Methods

To answer the two research questions it is important to quantify the popularity of websites. In the context of this research, Google's PageRank metric was be used for this [23]. The PageRank algorithm has been in use since it's inception in 1998 and is considered one of the main factors which determines Google's web page rankings and should be a factor that designers looking to perform effective SEO are aware of [24]. The research was be done using automatic evaluation to crawl a large number of websites for accessibility issues.

To answer the second research question there must be a way to identify between private and government domains. This can be done by checking the top level domain of the website, for example government websites will have the .gov domain whereas other websites may have .com. However in practice this is not always the case, some government domains may have other top level domains or a .gov domain is registered to a non-government entity [25]. In the case of this research any government website without a .gov or equivalent (.gob for some Spanish speaking countries or .gouv for France) government domain are not considered a government website. While there is one open source project for United State websites not using .gov [26], it is almost impossible to find or compile such a list for every government in the world. Therefore any non .gov domain is assumed to be a private website which, thanks to the big data-set, did not pollute the results. It is acknowledged that there is a possibility for fake .gov domains to exist but due to the difficulty of procuring such domains and the scope of the experiment, manual verification of domain authenticity was not done in this experiment [27]. See appendix A for the full list of top level government domains allowed in the experiment.

3.1 Data

The experiment used a list of 10 million most popular websites downloaded from Domcop[4] which uses websites taken from the Open PageRank initiative[5]. The calculated PageRank is computed using data provided by Common Crawl[6] and Common Search[7]. The data includes website URLs, PageRank values and their rank in the list (the higher the rank, the more popular they are). The amount of websites exponentially decreases as PageRank increases as shown in table 1. Note that the lowest PageRank included in the data-set is 2.88.

4 Metrics and Formulae

In terms of comparing accessibility of websites the A3 aggregation function was be used to compare them. A study by Martins and Duarte [28] evaluate

[1] https://wave.webaim.org/
[2] https://www.accessibilitychecker.org/
[3] https://accessibe.com/
[4] https://www.domcop.com/top-10-million-websites
[5] https://www.domcop.com/openpagerank/what-is-openpagerank
[6] https://commoncrawl.org/
[7] https://github.com/commonsearch

Amount of Domains for PageRank Range		
PageRank Range	**All Domains**	**Government Domains**
[2, 3)	1,716,494	66,359
[3, 4)	6,718,393	267,511
[4, 5)	1,534,954	71,720
[5, 6)	28,432	1,798
[6, 7)	1,528	130
[7, 8)	158	4,024
[8, 9)	25	7
[9, 10)	16	4
[2, 10)	1,000,000	407,553

Table 1: Distribution of domains in each PageRank bucket

it to be an effective metric for automatic evaluation. The formula is as follows:

$$A3 = 1 - \prod_{b} (1 - F_b)^{\frac{B_{pb}}{N_{pb}} + \frac{B_{pb}}{B_p}} \quad (1)$$

B_{pb} : is the number of actual points of failure of a checkpoint b in page p, b is the barrier (checkpoint violation)
N_{pb} : the number of potential points of failure of a checkpoint b in page p
F_b : identifies the severity of a certain barrier b
NOTE – a checkpoint is a test operation which verifies if the page conforms to a specific guideline.

The A3 aggregation function is especially effective in evaluating accessibility for different disability groups due to the varying severity value F_b [29], [30], for example a blind user will be impacted greater by missing alt text in an image than a user with motor control difficulties. In the case of this experiment the severity of all barriers were taken as equal so as to represent every disability fairly. The particularly beneficial part in this method for this experiment is the dynamic numeric value which changes based on the violations with respective to the amount of content on the page [31].

5 Experimental Approach

In the experiment there were two null hypotheses:

1. There is no difference between accessibility of high popularity and low popularity websites

2. There is no difference between accessibility of government and non-government domains

Firstly, for this experiment the domain data must be preprocessed so that only working websites (see appendix B for website verification process) are used. The data was split into:

- a high PageRank split - 100 URLs with a PageRank more than or equal to 6

- a low PageRank split - 100 URLs in the bottom 25% with respect to PageRank

- a government domain split - 100 URLs which have a government top level domain

- a non-government domain split - 100 URLs which do not have a government top level domain

Using these splits, an automatic website accessibility evaluation tool was be used to find any points of failure on all URLs in each split. This experiment was done using the WAVE accessibility evaluation tool[8]. Relevant checkpoint violations are saved and all possible violations are found to calculate the A3 metric for each URL. Specifics on implementation related to the experiment performed are discussed in appendix C.

6 Results

The results section is be split into two sections. One section outlines the results in comparison of high and low PageRank websites and the other compares government and non-government domains. Observations are made using independent T-tests[9] between samples with a confidence level of 95%.

6.1 High/Low PageRank

Figure 1 shows the distribution of A3 values for low and high PageRank websites. Table 2 shows the mean A3 values along with standard deviation.

Figure 1: PageRank A3 value distribution

PageRank Sample	Mean A3	A3 Standard Deviation
High, PageRank ≥ 6	0.256	0.261
Low, 25^{th} percentile	0.225	0.184
Government	0.340	0.325
Non-Government	0.202	0.136

Table 2: Mean and standard deviation for A3 values of samples

[8] https://wave.webaim.org/
[9] Independent T-test tool from https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html was used

Sample	A3 p-Value	PageRank p-Value
High/Low PageRank	0.364	N/A
Government/Non-Government	2.03×10^{-4}	9.53×10^{-5}

Table 3: p-Values for government/non-government and high/low PageRank samples

6.2 Government/Non-Government

Figure 2 shows the distribution of A3 values for government and non-government websites. Table 2 shows the mean and standard deviation for A3 value. Likewise in Table 4 for PageRank values. p-values for both PageRank and A3 value are located in Table 3.

Figure 2: Government/non-government A3 value distribution

Domain Sample	Mean PageRank	PageRank Standard Deviation
Government	4.638	0.288
Non-Government	4.502	0.157

Table 4: Mean and standard deviation for PageRank of government/non-government samples

7 Discussion & Conclusion

Regarding the first null hypothesis, due to p-value of 0.364 in Table 3, there is not enough evidence to suggest a difference, hence the hypothesis cannot be rejected. While it seems that there does not seem to be a correlation between popularity and accessibility there are a few factors which may have impacted the results in this experiment.

Due to the automatic evaluation performed in the experiment there may be mistakes in the accessibility report such as missing checkpoint violations or false positives [32]. While the tool does perform consistently over every website, it was impossible to manually check every domain due to the large sample size and thus this potential issue could not be avoided.

Another potential issue may be in the way that samples are preprocessed. While there was great effort put into only allowing working websites, there may be some domains which have content which have changed after they were added to the original sample data. Take for example a website that has later been deleted and the domain returns to some registrar company. The company then serves some generic webpage in place of the previous website. The automated tool may have instead scanned this generic page instead of the original page. To avoid this, manual filtering of the data sample should have been done or a more recent domain list should have been used but due to the automated nature of the experiment could not be accounted for.There is strong evidence to reject the second null hypothesis as $2.03 \times 10^{-4} < 0.05$.

Due to data gathered in Table 2 there is very strong evidence that government websites are clearly more accessible than their non-government counterparts which is supported by the generally stricter guidelines by government websites [8]–[11]. Possible differences in PageRanks are considered however due to earlier lack of evidence in disparity of A3 means for the high/low PageRank samples, this should not have an effect on the results taken for the second sample.

Future work in accessibility evaluations should seek to gather a greater sample of domains and manually vet domains both when picking which domains to add to the sample and when creating accessibility reports on pages. There is room for further research into accessibility across government domains for different countries and for more effective large-scale automated evaluation approaches and tools.

References

[1] W3C, *Introduction to web accessibility*, Web Accessibility Initiative (WAI), 2019. [Online]. Available: https://www.w3.org/WAI/fundamentals/accessibility-intro/.

[2] M. Wilson, *3 possible consequences of inaccessible websites*, Legal Marketing & Technology Blog, Feb. 2022. [Online]. Available: https://onward.justia.com/3-possible-consequences-of-inaccessible-websites/ (visited on 10/31/2023).

[3] L. Moreno and P. Martinez, "Overlapping factors in search engine optimization and web accessibility," *Online Information Review*, vol. 37, pp. 564–580, Aug. 2013. DOI: 10.1108/oir-04-2012-0063.

[4] T. Mavridis and A. L. Symeonidis, "Identifying valid search engine ranking factors in a web 2.0 and web 3.0 context for building efficient seo mechanisms," *Engineering Applications of Artificial Intelligence*, vol. 41, pp. 75–91, 2015, ISSN: 0952-1976. DOI: https://doi.org/10.1016/j.engappai.2015.02.002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197615000329.

[5] A. Akinyemi, *International web accessibility laws and policies*, www.whoisaccessible.com, Feb. 2020. [Online]. Available: https://www.whoisaccessible.com/guidelines/international-web-accessibility-laws-and-policies/ (visited on 11/21/2023).

[6] H. Petrie and O. Kheir, "The relationship between accessibility and usability of websites," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '07, San Jose, California, USA: Association for Computing Machinery, 2007, pp. 397–406, ISBN: 9781595935939. DOI: 10.1145/1240624.1240688. [Online]. Available: https://doi.org/10.1145/1240624.1240688.

[7] *Web content accessibility guidelines (wcag) 2.2*, www.w3.org. [Online]. Available: https://www.w3.org/TR/WCAG22/ (visited on 10/31/2023).

[8] M. Goodwin, D. Susar, A. Nietzio, M. Snaprud, and C. S. Jensen, "Global web accessibility analysis of national government portals and ministry web sites," *Journal of Information Technology & Politics*, vol. 8, pp. 41–67, Feb. 2011. DOI: 10.1080/19331681.2010.508011. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/19331681.2010.508011 (visited on 11/24/2023).

[9] *Make your public sector website or app accessible*, GOV.UK, May 2018. [Online]. Available: https://www.gov.uk/guidance/accessibility-requirements-for-public-sector-websites-and-apps (visited on 10/31/2023).

[10] M. Rogers, *Government accessibility standards and wcag 2*, Powermapper.com, 2017. [Online]. Available: https://www.powermapper.com/blog/government-accessibility-standards/ (visited on 11/24/2023).

[11] *Web accessibility laws & policies*, Web Accessibility Initiative (WAI), Mar. 2019. [Online]. Available: https://www.w3.org/WAI/policies/ (visited on 11/24/2023).

[12] J. Hoffmann, *Putting web accessibility first*, The History of the Web, Jun. 2017. [Online]. Available: https://thehistoryoftheweb.com/putting-web-accessibility-first/ (visited on 10/31/2023).

[13] W3C, *Web content accessibility guidelines (wcag) 2.1*, W3.org, Jun. 2018. [Online]. Available: https://www.w3.org/TR/WCAG21/ (visited on 11/21/2023).

[14] *Understanding conformance*, www.w3.org. [Online]. Available: https://www.w3.org/WAI/WCAG21/Understanding/conformance#levels (visited on 10/31/2023).

[15] M. Meyers, *Target settles with blind patrons over site accessibility*, CNET, Aug. 2008. [Online]. Available: https://www.cnet.com/tech/services-and-software/target-settles-with-blind-patrons-over-site-accessibility/ (visited on 10/31/2023).

[16] P. Collins, *What is the law on accessibility?* info.webusability.co.uk, Jun. 2022. [Online]. Available: https://info.webusability.

co.uk/blog/what-is-the-law-on-accessibility.

[17] V. Luque Centeno, C. Delgado Kloos, J. Arias Fisteus, and L. Álvarez Álvarez, "Web accessibility evaluation tools: A survey and some improvements," *Electronic Notes in Theoretical Computer Science*, vol. 157, no. 2, pp. 87–100, 2006, Proceedings of the International Workshop on Automated Specification and Verification of Web Sites (WWV 2005), ISSN: 1571-0661. DOI: https://doi.org/10.1016/j.entcs.2005.12.048. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1571066106002477.

[18] B. Parmanto and X. Zeng, "Metric for web accessibility evaluation," *Journal of the American Society for Information Science and Technology*, vol. 56, no. 13, pp. 1394–1404, 2005. DOI: https://doi.org/10.1002/asi.20233. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.20233. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20233.

[19] J. ARAa and S. Cecilia, "Artificial intelligence in web accessibility: Potentials and possible challenges," *Proceedings of IAC*, p. 173, 2022.

[20] S. Abou-Zahra, J. Brewer, and M. Cooper, "Artificial intelligence (ai) for web accessibility: Is conformance evaluation a way forward?" In *Proceedings of the 15th International Web for All Conference*, ser. W4A '18, Lyon, France: Association for Computing Machinery, 2018, ISBN: 9781450356510. DOI: 10.1145/3192714.3192834. [Online]. Available: https://doi.org/10.1145/3192714.3192834.

[21] J. Abascal, M. Arrue, and X. Valencia, "Tools for web accessibility evaluation," in *Web Accessibility: A Foundation for Research*, Y. Yesilada and S. Harper, Eds. London: Springer London, 2019, pp. 479–503, ISBN: 978-1-4471-7440-0. DOI: 10.1007/978-1-4471-7440-0_26. [Online]. Available: https://doi.org/10.1007/978-1-4471-7440-0_26.

[22] R. Ferraz, "Accessibility and search engine optimization on scalable vector graphics," in *2017 IEEE 4th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, 2017, pp. 94–98. DOI: 10.1109/ISCMI.2017.8279605.

[23] T. Haveliwala *et al.*, "Efficient computation of pagerank," Citeseer, Tech. Rep., 1999.

[24] A. Haynes, *What is google pagerank? (+ why should you care in 2023)*, Loganix, Apr. 2022. [Online]. Available: https://loganix.com/google-pagerank/ (visited on 10/31/2023).

[25] *It's way too easy to get a .gov domain name – krebs on security*, krebsonsecurity.com, Dec. 2019. [Online]. Available: https://krebsonsecurity.com/2019/11/its-way-too-easy-to-get-a-gov-domain-name/ (visited on 11/21/2023).

[26] *Government domains outside .gov and .mil*, Search.gov, Mar. 2022. [Online]. Available: https://search.gov/about/policy/govt-urls.html (visited on 11/21/2023).

[27] M. Humphries, *It's now (a bit) harder to register a .gov domain*, PCMAG, Mar. 2020. [Online]. Available: https://www.pcmag.com/news/its-now-a-bit-harder-to-register-a-gov-domain (visited on 11/21/2023).

[28] B. Martins and C. Duarte, "Large-scale study of web accessibility metrics," *Universal Access in the Information Society*, Dec. 2022, ISSN: 1615-5297. DOI: 10.1007/s10209-022-00956-x. [Online]. Available: https://doi.org/10.1007/s10209-022-00956-x.

[29] C. Bühler, H. Heck, O. Perlick, A. Nietzio, and N. Ulltveit-Moe, "Interpreting results from large scale automatic evaluation of web accessibility," in *Computers Helping People with Special Needs*, K. Miesenberger, J. Klaus, W. L. Zagler, and A. I. Karshmer, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 184–191, ISBN: 978-3-540-36021-6.

[30] A. P. Freire, T. J. Bittar, and R. P. M. Fortes, "An approach based on metrics for monitoring web accessibility in brazilian municipalities web sites," in *Proceedings of the 2008 ACM Symposium on Applied Computing*, ser. SAC '08, Fortaleza, Ceara, Brazil: Association for Computing Machinery, 2008, pp. 2421–2425, ISBN: 9781595937537. DOI: 10.1145/1363686.1364259. [Online].

Available: https://doi.org/10.1145/1363686.1364259.

[31] P. Sirithumgul, A. Suchato, and P. Punyabukkana, "Quantitative evaluation for web accessibility with respect to disabled groups," Apr. 2009, pp. 136–141. DOI: 10.1145/1535654.1535687.

[32] M. Vigo, J. Brown, and V. Conway, "Benchmarking web accessibility evaluation tools: Measuring the harm of sole reliance on automated tests," in *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, ser. W4A '13, Rio de Janeiro, Brazil: Association for Computing Machinery, 2013, ISBN: 9781450318440. DOI: 10.1145/2461121.2461124. [Online]. Available: https://doi.org/10.1145/2461121.2461124.

A Top Level Government Domains

The list of top level government domains which were allowed in the experiment. https://github.com/Litolo/Research_Methods_Final/blob/main/allowed_gov_domains.txt

B Checking Validity of Websites

Below checks if a website is working, i.e. webserver is running, we get a HTTP message within 10 seconds of the first GET request and we do not receive an error status code. Note that some websites may still be 'working' but should not be included in the sample. For example a web server may send a 301 (Moved permanently) response in which case we should visit that website. In the experiment in this paper it assumed that websites are either automatically re-routing or all the URLs serve correct content, which is not always the case unfortunately. See section 7 if you'd like to read more about drawbacks/limitations of the experiment.

```
import requests

def get_url_status(url):  # checks status for each url in list urls
    url = "https://"+url
    try:
        r = requests.get(url, timeout=10)
        return True if r.status_code < 400 else False
    except Exception:
        return False
```

C Implementation Specifics

For checkpoint violations read the error violations from the your selected tool. In the case of the experiment in this paper WAVE was used and due to the inability to pay for access to the API, a web scraping tool was used to extract the data from the website and transformed into json and csv files. Once each violation is saved, run another tool to scan the website (selenium was used in this experiment) and find all possible violations on the page. For example in the case of missing alt text errors, the page should be scanned for any <img> tags as possible violations. See the full code source for exact implementation in this experiment at https://github.com/Litolo/Research_Methods_Final