JOURNAL
COMPUTER SCIENCES INSTITUTE

JCSI 26 (2023) 23–28
Received: 19 September 2022
Accepted: 21 November 2022

Comparative analysis of social media accessibility

Analiza porównawcza dostępności mediów społecznościowych

Bartosz Henryk Bocheński*

Department of Computer Science, Lublin University of Technology, Nadbystrzycka 36B, 20-618 Lublin, Poland

Abstract

The paper assesses the accessibility compliance level of social media websites and their usability in terms of universal design. The study was performed on "reddit" and a social media website created by the author of the article for the express purpose of meeting the WCAG requirements as closely as possible, named "Twittn't". The analysis has been performed with the use of a survey study, an automated WAVE tool and an eye-tracker. The research groups consisted of fifteen and twenty people for survey and eye-tracker study respectively. The test results have shown strong correlation between usability of the social media website interface and number of its WCAG errors.

Keywords: social media; eye tracker; WCAG 2.0; universal design

Streszczenie

W artykule zbadano poziom dostępności mediów społecznościowych i ich funkcjonalność pod względem projektowania uniwersalnego. Badania zostały przeprowadzone na stronie „reddit" oraz medium społecznościowym stworzonym na potrzeby artykułu, spełniającym zalecenia WCAG w najwyższym możliwym stopniu, nazwanym „Twittn't". Analiza stron została przeprowadzona przy pomocą ankiety, automatycznego narzędzia WAVE oraz okulografu. Grupy badawcze zawierały 15 osób w przypadku ankiety oraz 20 w przypadku badania okulograficzengo. Wyniki badań wykazały silną korelację między użytecznością interfejsu mediów społecznościowych i liczbą błędów WCAG.

Słowa kluczowe: media społecznościowe; okulograf; WCAG 2.0; projektowanie uniwersalne

*Corresponding author

Email address: bartosz.bochenski@pollubl.edu.pl (B H Bocheński)

1. Introduction

The concept of web accessibility relates to the need to help people suffering from disabilities with accessing the Web, regardless if the disability is temporary, permanent or related to person's age. However, Web accessibility is also supposed to create a better experience for people with slow internet and those without any disabilities. Thus, an accessible website is easy to navigate, understand and interact with regardless of disability or age [1].

The Web Content Accessibility Guidelines (WCAG) Version 2.0 [2] is the internationally accepted Web accessibility compliance standard. For the purpose of this article's research, the United States and Poland both enforce the WCAG standard, with former doing so from the year 1998 with WCAG 1.0 and updating to 2.0 in year 2008, but only applying it to government owned or financed websites, and latter, as a member of the European Union, applying the standard in version 2.0 to all public sector websites since 2016 [3]. The EU (European Union) law was updated to use WCAG 2.1 in year 2018.

The main uses of social media [4] include but are not limited to: company promotions, communicating with friends or family, sharing art or random thoughts online or simply arguing with strangers over a typo. Due to their overwhelming popularity and wide variety of users, the social media websites should be designed to meet everyone's needs, including people with disabilities. However, many social media websites either follow accessibility guidelines coincidentally or completely ignore them [5], causing frustration due to unintuitive or simply poorly designed interface. Furthermore, social media are not forced to adhere to accessibility standards. In the US (United States) social media aren't usually financed or related to the federal government and thus free from the US accessibility legislation, for it doesn't apply to private websites. In the EU only public sector websites, which are websites owned by state, regional and local authorities or bodies governed by public law are required to follow accessibility guidelines. Social media do not belong to any of these categories and are thus exempt from adhering to accessibility standards.

The following study aims to evaluate how quality and clarity of social media's interface relates to its accessibility. Thus, to analyse this correlation, a UD (Universal Design) standard has been chosen (WCAG 2.0) and a hypothesis has been put forward: applying the standards of universal design greatly improve quality of a social media's interface and its usability for all users.

2. Literature review

With web accessibility being a long-standing issue for design of websites a whole there are a myriad of publications assessing accessibility standards and their application.

The article by W Arasid et al [6] analyses websites of 13 universities with the aim of improving their accessibility based on the WCAG 2.0 guidelines. The evaluation was performed with the use of TAW (Test de Accesibilidad Web), which is an automated evaluation tool. Results were presented in a form of graph showing error rate of each WCAG category. The study found that almost all of the studied websites have repeated the same WCAG 2.0 errors.

Reliance on automated tools has its disadvantages. An article by Markel Vigo et al [7] shows the consequences of relying only on automated test when identifying accessibility barriers. While automated test allows for quick evaluation, forgoing user test or expert evaluation often causes negative consequences. A total of 6 state-of-the-art tools were tested in their coverage, completeness and correctness of WCAG 2.0 conformance. The coverage averaged at most 50%, completeness reaching up to 38% and those with higher completeness averaged lower correctness, due to finding as many violations as possible, thus causing false positives. It was summarised that while using only automated test means that half of success criteria will not be tested and only 40% of the analysed criteria will be caught at the risk of generating false positives.

While the guidelines for disabled people are available, most practitioners do not conform to them, believing that accessibility guidelines provide no benefit to majority of users (unimpaired people) or cause negative impact for them. The study by Schmutz, S. et al [8] analyses the impact of implementing Web accessibility guidelines for users without impairments. Higher level of Web accessibility led to better performance compared to low or very low. There was no discernible difference between low or very low. Contrary to common concerns high conformance with Web accessibility guidelines provides benefits to unimpaired users.

3. Study subject

The study analyses user interface of two social media websites. While the websites aren't close to each other in interface design nor do they fulfil the exact same role, they are similar enough to warrant a direct comparison.

The first website under the study is hosted on https://www.reddit.com and is thus named reddit (stylized in all lower case) [9]. It is a social media website founded in 2005 by University of Virginia and is written in Python and JavaScript. Reddit has been a popular social media platform worldwide for the last several years, reaching about 430 million active users monthly current year (2022). However, it is also well known for its user interface being infamously unintuitive and not conforming to the rules of universal design (UD).

The second studied website is hosted on https://kullublin.xyz, and while it has no official name, it has been titled "Twittn't" as a reference to "Twitter", one of the more accessible social media platforms. It has been created for the needs of the study for the express purpose of fulfilling the WCAG 2.0 accessibility requirements and general rules of universal design. Twittn't was created using Wordpress software and the Buddypress plugin.

4. Study methods

The two social media websites were tested by methods provided below during the period of February to April 2022. Both websites were tested in their light mode to maintain consistency.

4.1. WAVE automatic tool

The WAVE tool [10] was used to quickly assess the WCAG compliance level of both websites. This tool presents results of its analysis of elements of a website into six categories:

- Errors
- Contrast errors
- Alerts
- Features
- Structural elements
- ARIA labels

For the purposes of this study the only noteworthy categories are errors and contrast errors, unless the other categories show an outstanding result, e.g., a website with a grand total of 0 ARIA labels (e.g., 4chan.org).

The tool was used in a form of an extension for the Mozilla Firefox browser. The evaluation was performed on live versions of both websites during the period of April 2022. For consistency's sake, both sites had enough content for scrolling to be necessary to load more and the test was done without scrolling. This way, the website content will affect the test equally on both websites.

4.2. Survey study

The survey study was performed with the assistance of fifteen people with no prior experience with either website. They were provided a list of ten task to complete in sequence on both websites:

1. Enter the provided website
2. Sign up and login
3. Create a new post
4. Upvote a post
5. Comment under a post of choice
6. Check your profile
7. Change your profile picture
8. Go to main page
9. Switch the website to dark mode
10. Log out

After which they were asked to fill out an interface assessment survey prepared by researchers at the Lublin University of Technology [1] (survey is under address https://forms.gle/7jFe1EMjPLATHxpS9). The survey consists of 31 questions divided into five categories which are as follows:

- Navigation and structure
- Communication, feedback and user assistance

- Application interface
- Subpage text
- Data input.

The results of the survey are presented on a scale from 1 to 5. With 1 meaning a function of an interface is unintuitive and difficult to perform and 5 meaning that it works without any complications and is intuitive.

4.3. Eye-tracker

The eye-tracker [11] study was performed with the assistance of twenty users with no prior experience with either of the websites. The study group contained both near sighted people and those without any vision impairments. They were asked to locate ten different interface elements present on both websites which are:

1. Element used to add an image to a post
2. Element used to enter the comment section of a post
3. The name of the group in which the post on screen has been posted
4. Dark mode button
5. Profile image change button
6. Element used to start creating a new post
7. Button used to reply to a comment
8. Button used to create a new group
9. The username of the post's author
10. Button used to leave a group

The study was performed on a computer equipped with a Gazepoint GP-3 HD eye-tracker. The result categories analysed by the study are as follows:

- task completion time
- Time To First Fixation (TTFF) [12] on an Area of Interest (AOI)
- success rate of a task.

5. Results

5.1. WAVE automatic tool

The following figures (Figure 1, 2) show the WAVE tool result summaries for reddit (Figure 1) and Twittn't (Figure 2).

The following table (Table 1) describes the number of WCAG 2.0 error occurrences per category on both websites. Table 1 contains only the WCAG categories in which at least one of the websites had at least one error.

Table 1 Comparison of number of WCAG 2.0 errors for WAVE test

WCAG 2.0 Category	Occurrences on reddit	Occurrences on Twittn't
1.1 Text alternatives	42	8
1.2.1 Audio/Visual only	11	0
1.2.2 Captions pre-recorded	10	0
1.2.3 Audio description or media alternative pre-recorded	10	0
1.3.1 Info and Relationships	2	1
1.4.2 Audio control	11	0
1.4.3 Contrast minimum	147	1
2.1.1 Keyboard	1	1
2.4.1 Bypass blocks	1	1
2.4.4 Link purpose (in context)	59	7
2.4.6 Headings and labels	2	1
3.3.2 On input	1	0
4.1.2 Name, role, value	1	0
Total number	**298**	**20**

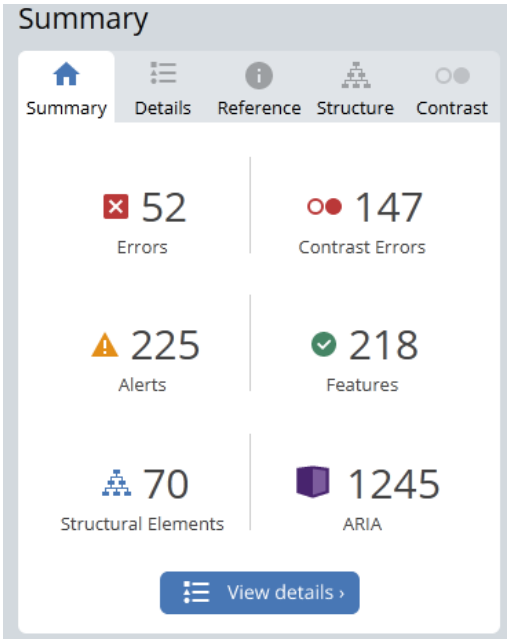

Figure 1: Summary of WAVE tool results for reddit.

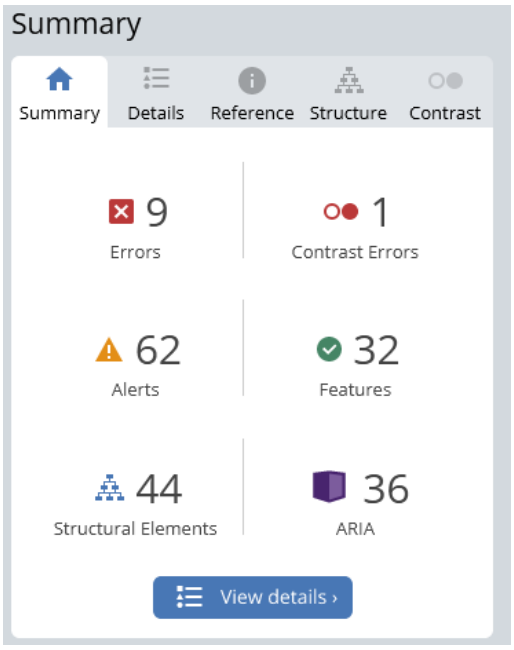

Figure 2: Summary of WAVE tool results for Twittn't.

The totals error counts between the summaries (Figure 1, 2) do not match the total amounts in the

table (Table 1) due to summaries only showing the number of elements with errors and not the total error count. For example, an empty button is a singular element but the error it causes falls under WCAG 2.0 category 1.1 and 2.4.4 thus causing 2 errors.

5.2. Survey study

The interface assessment survey has resulted in a total average of 2 points for reddit and 4.21 for Twittn't. The averages for each question of the survey are presented in Figure 3.

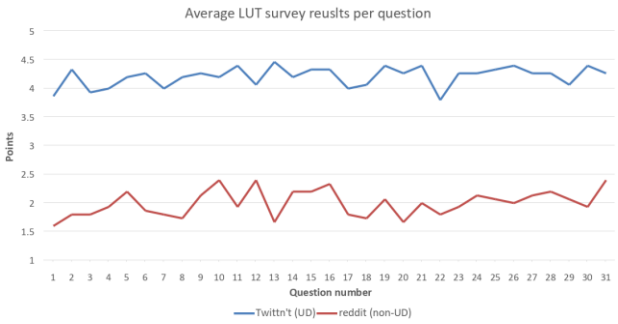

Figure 3: Average survey results on a scale from 1 to 5 for both websites.

The average results of each category are presented in Table 2.

Table 2 Average survey results for each category of the website assessment survey

Category	reddit	Twittn't
Navigation and structure	1.87	4.1
Communication, feedback and user assistance	2.07	4.19
Application interface	1.96	4.26
Sub page text	1.99	4.12
Data input	2.12	4.27

5.3. Eye-tracker study

The results of eye-tracker study have been compiled into tables task success rates and the averages for both TTFF and task completion time. Table 3 shows study results for reddit and the Table 4 displays the results for Twittn't. Shorter task completion times and TTFF and higher success rates indicate better quality of the interface.

The distinction between near sighted people and the unimpaired wasn't made due to vision impairment not causing any significant difference in overall results.

To visually represent the locations of users' fixations a set of heat maps [13] has been generated. The heat maps shown in Figure 4 are for reddit's stimulus number 2, the task of which was to locate the element or elements responsible for redirecting the user to post's comment section.The second heat map (Figure 5) shows the fifth stimulus for Twittn't, the task of which was to locate element or elements used to change user's profile photo.

Table 3: Eye-tracker data table for reddit containing average task completion times, average TTFF and task success rates

Stimulus	Task completion time	TTFF	Task success rate
1	12677.37	8184.44	35%
2	8670.37	4061.44, 6745.42	95%
3	11750.51	3673.82	85%
4	9395.05	7857.48	85%
5	6457.87	2911.54	80%
6	7812.689	2289.511	90%
7	6506.373	3792.282	55%
8	14198.73	9861.914	75%
9	6198.946	3646.215	75%
10	14549.77	5199.44	95%

Table 4: Eye-tracker data table for Twittn't containing average task completion times, average TTFF and task success rates

Stimulus	Task completion time	TTFF	Task success rate
1	5348.40	3417.50	70%
2	7046.20	3153.80	100%
3	4020.80	5295	65%
4	4278	3221.20	75%
5	5916	486.40, 4957.10	95%
6	2989	2420.90	100%
7	2781.70	2039.20	85%
8	6765.60	3426	65%
9	2212.40	1753.60	95%
10	2560.50	1701.90	95%

Figure 4: Heat map of stimulus 2 for reddit.

Figure 5: Heat map of stimulus 5 for Twittn't.

6. Discussion

For the automated WAVE test the total number of errors on reddit is quite high in comparison to Twittn't, 1490% of the Twittn't total error count (Table 1). The difference couldn't have come from user generated content due to less than 10 posts being loaded on each of the websites at the time of the tests. Social media websites tend to struggle with text alternatives and link purpose categories due to most of the content being generated by the user on the front-end side of website. However, the number of occurrences of these errors on reddit exceeds double the number of loaded posts, meaning that the errors, especially empty links, must originate from parts of the website interface. Surprisingly enough, part ofthe image-based posts did include proper alternative text. The only real outlier on the side of Twittn't is text alternatives and link purposes. The errors in these categories originate from user profile pictures not having their alternative text handled properly when they appear as a part of a post. This issue is slightly mitigated by profile picture never appearing without the username beside it. Those errors, while notable, will not affect the majority of users.

The contrast issues on the other hand will affect all users. Those errors on reddit originated from text colour being light grey while the background is either slightly lighter grey or white. The user error or malice doesn't contribute to the issue due to reddit not allowing the change of text colour on posts and WAVE tool not counting images for contrast category. For comparison, the reddit's text to background contrast ratio is 4.21:1 while the Twittn't is 8.59:1. In theory this means that reddit fails WCAG's AA requirement for normal text and barely passes it for large text, while the Twittn't passes the contrast test. In practice, this means that users might have issues reading the labels of interface elements or information such as post author's name or post's title on reddit and visually impaired users might be incapable of reading them.

The survey study has shown that the survey respondents, after completion of the tasks, were highly biased against reddit. The average difference between reddit and Twittn't was above 2 points on a 5-point scale (Figure 3).

Important detail to mention before the analysis of results is that the areas of interest set in the application used to collect eye-tracking data were highly lenient, i.e., the area of interest was about 20% larger than the element itself for all stimuli.

The task completion rates for the eye-tracker study only account for any fixation happening on a stimulus. This decision was made to reduce the risk of misinterpreting the reason for time difference between last fixation and the task completion.

The time difference between TTFF and task completion times is quite long for reddit, with task completion times being up to 240% longer than TTFF. Such difference indicates that the users had issues discerning the purpose of the element they were looking at. The time differences vary from 4 to 8 seconds (Table 3). The discrepancy was made clear with stimuli 1, 8 and 10. On those stimuli the users averaged over 30 total fixations per stimuli, while averaging 0.35, 4 and 3 fixations on the areas of interest (AOI) for stimulus 1, 8 and 10. An edge case happened in stimulus 4 where one respondent took slightly over 16 seconds and 100 (5 times the average) fixations to find the correct AOI. Furthermore, the subsequent fixations on AOI usually appeared early after the first fixation. Last fixations on the AOI usually happened over a second before the task completion time.

The eye-tracker results for Twittn't (Table 4) were far more stable, with times between TTFF and task completion time ranging between 500 and 3000 ms. The stimuli with longer delays between those times usually had an average of 2.2 subsequent fixations close to the task completion time. Additionally, the last fixation on the AOI usually occurred less than 500 ms before task completion time.

The notable outlier on the Twittn't eye-tracker data is stimulus 3 (Table 4), where the average TTFF ended up being higher than task completion time by over a second. The cause of this is a couple of respondents that decided they found the correct answer less than a second after the start of the stimulus and either missing the element or the eye-tracker not recording the fixation due to it being too short.

Due to high difference in eye-tracker study results between analysed websites and substantial difference in task completion times, especially for reddit, a Levene's test [14] was performed on the eye-tracker results. The test results for all of the stimuli were above 0.05, thus all of the eye-tracker data met the assumption of homogeneity of variance, with the highest result being stimulus 6 with p value of 0.915 and lowest being stimulus 9 with p value of 0.117.

On a side note, many of important interface elements of reddit are placed on a long dropdown menu on the top right side of a screen. The element that opens the menu has severe contrast issues. The notable elements hidden by the menu are: user profile options, community creation menu and dark mode. The unintuitive nature of the menu could have affected the results of the survey and eye-tracker study.

7. Conclusions

In summary, results of the study have confirmed the hypothesis put forward at the beginning of the article.

The automatic WAVE test revealed the first concerning detail about reddit: while total number of errors is quite high (298), there are websites that function properly with higher error counts. The real problem is that the half of them are contrast errors that make reddit difficult to navigate at best and painful for the eyes at worst. Twittn't on the other hand has low number of total errors (20) but only a singular contrast error, making it significantly easier to navigate and to read the website's content.

The survey has confirmed the predicted issues with interface due to contrast. The survey results per category (Table 2), in reddit's, case is the lowest for navigation and structure (1.87) and application interface (1.96), while on Twittn't the application interface is the second highest rated category at 4.26.

The eye-tracker study has shown that the users had significantly easier time locating element on Twittn't than on reddit. The task completion times being shorter on Twittn't with higher success rates among users shows a clear difference in interface clarity. Times to first fixation support that: on reddit the TTFF tends to be upwards of 8 seconds before task completion time. Either the user doubted that the element they found fulfilled the ascribed purpose or deemed the element to not be the correct one and continued their search. On the other hand, the TTFFs on Twittn't were closer to task completion times. A singular exception being a stimulus with two AOIs (Table 4, stimulus 5), meaning that unlike on reddit the users were not second guessing themselves once they located the correct element.

In conclusion simply following the rules of universal design, such as WCAG 2.0, greatly improves the quality and clarity of social media website interface.

References

[1] M. Miłosz, Ergonomia systemów informatycznych, Biblioteka Cyfrowa Politechniki Lubelskiej, 2014.

[2] W3C Recommendation, Web Content Accessibility Guidelines, 11 December, http://www.w3.org/TR/WCAG20/, [20.07.2022].

[3] Web Accessibility Initiative, Web Accessibility Laws & Policies, https://www.w3.org/WAI/policies/, [20.07.2022].

[4] A.M. Kaplan, M. Haenlein, Users of the world, unite! The challenges and opportunities of Social Media, Business Horizons, 53(1) (2010) 59-68, https://doi.org/10.1016/j.bushor.2009.09.003.

[5] M.G. Dinis, C. Eusébio, and Z. Breda, Assessing social media accessibility: the case of the Rock in Rio Lisboa music festival, International Journal of Event and Festival Management, 11(1) (2020) 26-46, https://doi.org/10.1108/IJEFM-02-2019-0012.

[6] W. Arasid, A.G. Abdullah, D. Wahyudin, C.U. Abdullah, I. Widiaty, D. Zakaria, N. Amelia, A Juhana. An Analysis of Website Accessibility in Higher Education in Indonesia Based on WCAG 2.0 Guidelines, IOP Conference Series: Materials Science and Engineering, 306(1) (2018) 012130, https://dx.doi.org/10.1088/1757-899X/306/1/012130.

[7] M. Vigo, J. Brown, 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 (W4A '13). Association for Computing Machinery, New York, NY, USA, 1 (2013) 1–10, https://doi.org/10.1145/2461121.2461124.

[8] S. Schmutz, A. Sonderegger, & J. Sauer. Implementing Recommendations From Web Accessibility Guidelines: Would They Also Provide Benefits to Nondisabled Users, Human Factors, 58(4) (2016) 611–629, https://doi.org/10.1177/0018720816640962.

[9] K.E. Anderson, Ask me anything: what is Reddit?, Library Hi Tech News, 32(5) (2015) 8-11, https://doi.org/10.1108/LHTN-03-2015-0018.

[10] L.R. Kasday. A tool to evaluate universal Web accessibility. In Proceedings on the 2000 conference on Universal Usability (CUU '00), Association for Computing Machinery, New York, NY, USA (2000) 161–162, https://doi.org/10.1145/355460.355559.

[11] K. Holmqvist, M. Nyström, and F. Mulvey. Eye tracker data quality: what it is and how to measure it. In Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12), Association for Computing Machinery, New York, NY, USA (2012) 45–52, https://doi.org/10.1145/2168556.2168563.

[12] L.N. van der Laan, I.T.C. Hooge, D.T.D. de Ridder, M.A. Viergever, P.A.M. Smeets, Do you like what you see? The role of first fixation and total fixation duration in consumer choice, Food Quality and Preference, 39 (2015) 46-55, https://doi.org/10.1016/j.foodqual.2014.06.015.

[13] O. Špakov and D. Miniotas 2007. Visualization of Eye Gaze Data using Heat Maps, Elektronika ir Elektrotechnika, 74(2) (2007) 55-58.

[14] G.V. Glass. Testing Homogeneity of Variances, American Educational Research Journal, 3(3) (1966) 187–190, https://doi.org/10.3102/00028312003003187.