Accessibility Testing of European Health-Related Websites

Cecilia Sik-Lanyi¹ · Éva Orbán-Mihálykó²

Received: 12 January 2019 / Accepted: 29 June 2019 / Published online: 5 July 2019
© The Author(s) 2019

Abstract

The current development of the Internet and its growing use make it necessary to satisfy the needs of users with disabilities. The primary objective of this study is to examine healthcare-related websites in nine European countries in order to evaluate the status of their accessibility. Such a detailed statistical comparison has not yet been made in Europe, especially as the present study offers a dual measurement system combining both the application of automated testing software and statistical analysis of user feedback. The study compares 48 websites from Eastern Europe with 51 sites from Western and Northern Europe. The research phase was performed in three steps: firstly by using AChecker, secondly by Nibbler and subsequently followed by user feedback questionnaires evaluated by a group of experts. The overall goal of this study is to determine the most common accessibility problems and to draw site owners' attention to shortcomings so that they can improve the quality of service of their healthcare-related sites in the future. The investigated European websites are grouped into Eastern and Western–Northern countries. We compared our results from different perspectives and ascertained that no significant differences can be established between the two groups predicated on their respective economic situations. Equally, no correlations were observed while comparing the sizes of web pages in Kbytes, the number of barriers and their Nibbler accessibility scores. Furthermore, there appears to be no correlation between the results of the software tests and the percentage of the elderly population in the respective country.

Keywords Software engineering · WCAG 2.0 · Testing tools · Usability testing · Web design · Computer science

1 Introduction

The current development of the Internet and its growing use make it necessary to meet the needs of those disabled people who experience accessibility problems, be it related to visual, motor, mobility, auditory, cognitive or intellectual skills. Most software engineering companies have not been developing for users with special needs because they do not see a potential market in these users [1]. However, figures have proved that at least 10% of the world's population has some kind of impairment (<http://www.disabled-world.com/disability/statistics/>). This number is estimated to reach 14%

in the USA, and 65% of the population older than 65 years is likely to develop a handicap of some kind. Disabilities correlate with age. In developed societies, more and more people turning older than 75 are likely to have some kind of impairment. This group will comprise 14.4% of the population by 2040, compared to 7.5% in 2003. It is almost a twofold increase [2].

Another factor is that by 2020 25% of the EU's population will be over 65 years. Money spent on pension, health and long-term care is expected to increase by 4–8% of the GDP in the forthcoming decades. These expenditures will triple by 2050. It is not negligible either that the combined wealth of older Europeans is estimated more than €3000 billion [3]. If companies do not respond to the demand of producing barrier-free websites, they are likely to lose a considerable number of potential customers/users. New solutions are needed for those elderly people who might not be able to leave their homes and for healthcare monitoring [3, 4]. Barrier-free Internet and software are an essential part of this process. Figure 1 shows elderly populations in the countries of the European Union (EU-28) in the period of 2004–2014.

✉ Cecilia Sik-Lanyi
lanyi@almos.uni-pannon.hu

Éva Orbán-Mihálykó
orbane@almos.uni-pannon.hu

¹ Department of Electrical Engineering and Information Systems, University of Pannonia, Egyetem u. 10, Veszprem 8200, Hungary

² Department of Mathematics, University of Pannonia, Egyetem u. 10, Veszprem 8200, Hungary

Fig. 1 Elderly population in EU-28: total percentage of population, 2004–2014 [5]

Table 1 GDP, the total population and elderly population numbers in the investigated countries

Country	GDP 2017 US\$ billion [6]	Total population (millions) in August 2018 [7]	Region	The elderly population % of total population in 2013 or 2014 [5]	The elderly population % of total population in 2017 [8]
Germany	3677.44	82.323	W-E	21.45 (2014)	22.06
Switzerland	678.89	8.555	W-E	17.6 (2013)	18.15
Sweden	538.04	9.994	N-E	19.91 (2013)	20.26
Austria	416.60	8.754	W-E	18.21 (2013)	19.26
Finland	251.88	5.545	N-E	19.94 (2014)	21.1
Poland	524.51	38.093	V4	15 (2014)	16.86
Czech Republic	215.73	10.626	V4	17.6 (2014)	18.98
Hungary	139.14	9.683	V4	17.72 (2014)	19.05
Slovakia	95.77	5.45	V4	No data available	15.43

Table 1 shows the GDP, total population, region and percentage of the population in the countries included in this study from Central-Eastern Europe (the so-called V4 countries) and from West Europe (W-E) and North Europe (N-E). The size of the elderly population is increasing everywhere in Europe.

Table 2 shows the GNI data and the region by each country in this study.

Providing a barrier-free Internet, software and VR applications is a challenging task. Already existing principles and standards of universal design/design for all cannot guarantee accessibility, per se [11, 12], and the fact that providers overlook these regulations further delays the introduction of

a barrier-free Internet [13]. In 1999, the World Wide Web Consortium published the Web Content Accessibility Guidelines WCAG 1.0 [14]. Web Content Accessibility Guidelines, WCAG 2.0, was released in 2008 [15]. To complete WCAG 2.0, the researchers recorded the number of known problems identified for A, AA and AAA level guidelines. While WCAG 1.0 contains 14 guidelines with 62 checkpoints at three priority levels, the WCAG 2.0 has only four principles with 12 guidelines. Both have three levels of conformance: A, AA and AAA. Both provide technical advice.

The main principles and structure of WCAG 2.0:

- Principles—Top four principles.

Table 2 GNI per capita in the investigated countries in 2017

Country	GNI per capita, Atlas method (current US\$) [9]	GNI per capita, PPP (current international \$) [10]	Region
Germany	43,490.00	51,680.00	W-E
Switzerland	80,560.00	50,980.00	W-E
Sweden	52,590.00	65,610.00	N-E
Austria	45,450.00	52,500.00	W-E
Finland	44,580.00	45,400.00	N-E
Poland	12,730.00	27,970.00	V4
Czech Republic	18,160.00	34,450.00	V4
Hungary	12,870.00	26,960.00	V4
Slovakia	16,610.00	30,880.00	V4

- Guidelines—12 guidelines provide the basic goals.
- Success criteria—For each guideline testable success criteria are provided. Three levels of conformance are defined: A (lowest), AA and AAA (highest).

Several other guidelines based on WCAG 2.0 exist. For example, the standard in Canada is the Standard on Web Accessibility [16], which follows the WCAG 2.0 guidelines. The Web Content Accessibility Guidelines for Japan, called JIS (Japanese Industrial Standards), was released in 2004 [17]. Recognizing the growing demand for an accessible Internet, more and more countries pass legislation concerning accessible digital information. Countries frequently support and adopt WCAG 2.0 by referring to the guidelines in their legislation [18–21]. The USA, furthermore, specifies that all Federal electronic media must be made accessible to people with disabilities. WCAG 2.0 AA standard has been approved as an appropriate standard for accessibility by the US Department of Justice in multiple settlement agreements [22, 23]. The European Parliament approved a draft in February 2014 which states that all websites in the public sector have to be developed to be accessible for everyone [24]. This European standard, called EN 301 549 version 1.1.2, describes relevant accessibility requirements. The problem of accessibility is twofold at this juncture. Firstly, not all member states have harmonized their accessibility laws with that of the European Union, even though it was expected to take place by September 2018. Secondly, not all public sector websites had been able to comply with guidelines by September 2018. The next goal is that mobile apps in the public sector should comply by June 2021 [24, 25]. Although the newest version of WCAG, the WCAG 2.1 [26, 27], was released on 5 June 2018, there is no suitable, efficient automatic test tool based on these guidelines. There has been extensive research dealing with Web accessibility, mainly in the domain of testing the accessibility of governmental websites [28–36] only to conclude

that websites are barely accessible and further research is required. Goodwin et al. [33] conducted the first global analysis of the Web accessibility of government websites of 192 United Nation Member States almost 9 years ago. Among other things, they proved the following hypotheses:

- The wealthier a country is (GNI per capita), the fewer barriers will be present on its websites.
- A website with a WAI logo has better website accessibility scores.

Zheng [34] investigated 108 consumer health information websites in the USA. Although this research was performed in 2003 and based on WCAG 1.0 [14], Zheng's research methodology used a taxonomy that classified the websites into six categories—E-commerce, corporate, government, portal, community and education. The result was that none of the websites was completely accessible to people with disabilities. Governmental, educational, health-related websites, nevertheless, exhibited better performance on web accessibility than other websites in other categories. They also established correlation between the accessibility and the popularity of a website.

Investigating web accessibility has become a frequently researched topic and a number of publications have appeared that address—among others—the accessibility of such institutions as universities. A group of Brazilian researchers [35], for instance, set out to investigate Brazilian federal universities to see whether they complied with e-government standards. Unfortunately, their results confirmed that most federal university websites do not follow important accessibility standards. Some of the reasons behind the findings include IT teams that are unable to cope with staff shortage, high demand, tight deadlines and the lack of specific development methods. Our own research [37] looks at the websites of Hungarian universities to discover potential accessibility issues of students who have some colour deficiency problems. Our results make it clear that, unfortunately, not every website is clearly visible; therefore, students with colour deficiencies cannot acquire information the same way as healthy ones. Our recommendation for web designers here is to create websites that do not only feature colour visual cues, that is, the implementation of various patterns and huge visual contrasts would be more efficient to promote visibility in texts, buttons and links. In general, website designers do not tend to develop different versions of the same site for each disability type; in fact, it is unlikely that further features are added for the comfort of different disability groups [38–41].

The priorities of the European Commission include the simplification of modern digital contract rules, the promotion of access to digital content and the enhancement of online sales [42, 43]. The EC wishes to support digital market strategies in the member countries and introduce new

e-commercial regulations in order to make buying and selling goods inside the EU easier [44]. In Hungary, people made online purchases more than 22 million times in 2015 [45]. The retail turnover rate in Hungary grew by 18% in the first half of 2016 compared to the data from the previous year, and it reached 131 billion forints [46].

In this paper, we present the results of testing and evaluating almost 100 healthcare-related websites from nine European countries. Based on the results, we will review our earlier recommendations outlined in an earlier paper [41] to see whether they still hold. The following sections will set out the context of the research, its implementation and results and, finally, look at the proposed hypotheses. Section 3 details the research methodology and hypotheses. Section 4 discusses the results, and Sect. 5 presents the discussion and main conclusions of the research.

2 Materials and Methods

The present study highlights the most acute accessibility problems of the examined healthcare-related websites. The investigation includes healthcare-related sites from nine European countries; 48 from Eastern Europe (the so-called V4 countries); and 51 sites from Western and Northern Europe (N–W-EU). Table 3 shows the number of tested websites in the investigated European countries.

We examined the V4 countries together because, on the one hand, they form one region and, on the other hand, economically, historically and from an educational point of view they are very similar. Salaries in the V4 countries amount to one-third or one-fifth of that in the western part of Europe, while the prices of basic foods and daily necessities do not differ between significantly.

As the accessibility properties of the web pages can be considered random qualities, the comparisons require statistical methods. This article uses both quantitative and

qualitative research methods. The measurement of web accessibility is addressed in this section.

The websites that fall within the scope of interest of the present investigation were chosen based on Google search results, concentrating on top finds. For automatic testing, AChecker [47] and Nibbler [48] were used. AChecker is a free tool that checks single HTML pages for conformance with accessibility standards to ensure that the content can be accessed without limitations [47]. AChecker diagnoses known problems, likely problems and potential problems with their levels of conformance based on WCAG 2.0. Nibbler is a free tool for testing websites; it generates a report scoring the website on a scale of 10 for key areas including accessibility, SEO (search engine optimization), social media and technology [48]. Figure 2a and b presents screenshots featuring AChecker (Fig. 2a: before testing, Fig. 2b: test results), and Fig. 3 introduces the Nibbler automated tool. To improve the depth of the present research, we extended our automated tests by interviewing users in the framework of a questionnaire. We involved experts with considerable experience gained in the area of digital accessibility to compile a questionnaire (see Table 4) and elicit useful feedback from users with regard to their accessibility issues. The experts involved in the present investigation include MSc IT engineer university students specialized at software ergonomics, human–computer interaction, user interface design and web accessibility.

Despite the fact that both AChecker and Nibbler automatic are very effective, there were a number of cases in which they rather badly affected our research process. Some of the issues that influenced measurement and/or evaluation include test situations in which AChecker or Nibbler being unable to measure and detect certain human needs. Websites with a traditional blind version (yellow letters on a black background), for instance, are not refreshed frequently. Webmasters tend to forget to make new information available; consequently, those websites are likely to contain outdated information. Certain buttons lose their functions over time, which may make users confused. Automatic test tools, to move onto another area, do not measure colour vision problems well enough. At times, users of low or impaired can hardly read CAPTCHA figures (an audio version remedies this problem). Considering further user groups with accessibility issues, blind users or those who have motion problems with their hands use a special keyboard or just the TAB key to navigate. If one button or submenu is not available by the TAB key, the website becomes unsuitable for further use. These are the main aspects that we wanted our experts to investigate in a questionnaire based on our earlier experiences in the field.

As mentioned, the testing phase started with AChecker, followed by Nibbler and ended with the questionnaires (Table 4) including the use of the SEE web application [49]

Table 3 Number of the tested websites

Country	Number of the tested website	Region
Germany (GER)	9	N–W-EU
Switzerland (SWI)	16	N–W-EU
Sweden (SWE)	8	N–W-EU
Austria (AU)	8	N–W-EU
Finland (FIN)	10	N–W-EU
Poland (PO)	8	V4 countries
Czech Republic (CZR)	8	V4 countries
Hungary (HU)	19	V4 countries
Slovakia (SLO)	13	V4 countries

(a) Screenshot of the AChecker automated tool (at the beginning of the test)

(b) Screenshot of the AChecker automated tool (after the test, showing the results)

Fig. 2 **a** Screenshot of the AChecker automated tool (at the beginning of the test), **b** screenshot of the AChecker automated tool (after the test, showing the results)

(question 3 in Table 4). The SEE web application proved to be the best and easiest test software from the point of view of handling because the user only needs to set a colour deficiency type with a slider and does not have to upload a picture to a web page as in the case of other testers. The test phase is concluded with the use of Variantor special glasses [50].

All these tests were performed in the time period of June–July 2018. The most frequently visited and most popular sites were chosen by Google statistics for the test upon searching for “healthcare”, “assistive technology”, “health related”, etc. keywords. The focus of interest was web-

sites that contained health-related information or blogs about health topics or web shops for medicaments or assistive technologies. Moreover, we consulted native speakers in the respective countries and asked them to send us the URLs of the most popular websites in their areas of interest.

3 Hypotheses and Research Questions

This chapter presents our research questions (Qs) and hypotheses (H) based on related questions. They are grouped

Fig. 3 Screenshot of the Nibbler automated tool

based on closely related topics. The scenario is based on the order of first asking a question (Q); then, we formulate the statistical hypothesis together with its alternative (H).

3.1 Testing by Nibbler

Q1 Can we ascertain significant differences between the groups of V4 countries and N–W-EU countries performing the evaluation by the data of Nibbler test investigating the different aspects separately?

H1 Are the expectations of the scores of the V4 countries and N–W-EU countries regarding the aspects separately equal versus the alternative hypothesis that they are different?

Q2 How large are the correlation coefficients and determination coefficients between the points of the different aspects? Is there any relation between the different aspects? Can the scores of the different aspects be considered statistically independent or higher technology results higher accessibility?

H2 Is the correlation coefficient between the different aspects equal to zero or it is not? What about the one-sided alternative hypothesis?

3.2 Testing by AChecker

Q3 Can we see significant differences in average error numbers in the N–W-EU and V4 countries grouping the questions by their types (using the AChecker automated tool)? What is the situation if we focus on the questions separately and what about the most affected principles?

H3 There are no differences between the V4 and N–W-EU countries concerning the average error numbers, versus there is difference between the error numbers.

The expectations of the error numbers are the same statistically in the case of the most frequently affected principles comparing V4 and N–W-EU countries versus the non-equality.

Q4 Are the conformance levels (A, AA, AAA) the same in the N–W-EU and V4 countries based on the error numbers (by using the AChecker automated tool)?

H4 The average error numbers are equal comparing V4 and N–W-EU countries in case of each conformance level versus they are different.

Q5 Which are the most frequently violated principles based on the error numbers of AChecker results and how much are the extreme values?

H5 Most error numbers originate when web developers violate the first principle.

3.3 Comparison

Q6 Is there any relation between the scores of the separate websites given by Nibbler tests and AChecker tests or are they independent? What about the average scores of websites belonging to the same countries? How much are the correlation coefficients?

H6 The correlation between the scores by Nibbler test and the error numbers provided by AChecker is zero, versus it differs from zero, investigating websites separately and grouping them by countries.

Table 4 Questionnaire of the experts' test

Questions	Remarks
1. Does the website have a blind version? (yes/no)	Developing a blind version site is not a good solution
2. Each link can be determined based on the link text alone	For example, "click here" or "here" links are not understandable
2.1 Can linked texts be followed?	If the earlier answer is "yes", please write the text of the link here
3. Is information lost by a user with colour deficiency?	Check it with the Google SEE web application
3.1 Protanomaly	If you answered "yes" to the previous question, please indicate here which colour information was lost!
3.2 Deuteranomaly	
3.3 Tritanomaly	
3.4 Achromatopsy	
4. Is there CAPTCHA on the website? (yes/no)	For this question, look at the whole web page, not just the main page!
4.1 If CAPTCHA is available, can it be heard? (yes/no)	–
5. Is every part accessible by using the keyboard? (yes/no)	Use TAB to scroll through all menu items; can you enter all submenu points; are all functions available on the main page?
5.1 Inaccessible features	If your answer was "no" to the previous question, type here what was unavailable on the keyboard!
6. Is there an English-language website? (yes/no)	–
6.1 Is the information same on the opening page in English and in the national language? (yes/no)	–
7. Is the website responsive? (yes/no)	–
8. Is WAI logo indicated on the site? (yes/no)	–

3.4 Human aspects

Q7 Should users who have any colour deficiency be able to use the investigated websites the same way as people with no vision impairment (by using as test tool the SEE web application and the Variantor special glasses)?

H7 Users who have any colour deficiency are able to use the investigated websites the same way as people with no vision impairment (by using as test tool the SEE web application and the Variantor special glasses).

Q8 Is there any difference between the N–W-EU and V4 countries based on the experts' test?

H8 There is no significant difference between the N–W-EU and V4 countries based on the evaluation of the experts' questionnaire.

3.5 External and Internal Characteristics and Web Accessibility

Previous research upholds [33] that the wealthier a country is (GNI per capita), the fewer barriers will hinder the use of websites. Moreover, website with a WAI logo has better accessibility scores. Furthermore, there is a correlation between the size of web pages and the number of barriers detected. In addition, we challenged these findings only to prove the opposite; that is, neither of the above statements hold true for health-related web pages.

3.5.1 Correlation between the sizes of the web page and barriers

Q9 Is there any correlation between the sizes in Kbytes of the website and the number of known problems by AChecker and the Nibbler accessibility scores? How large are the correlation coefficients?

H9 The correlation coefficients between the sizes in Kbytes of websites and the numbers of known problems by the AChecker/the Nibbler accessibility scores are zero. The alternative hypothesis is that the correlations differ from zero; that is, the sizes and the accessibility are dependent.

3.5.2 Correlation between the economic aspects and barriers

Q10 Are there any links between the GNI per capita of a country and the known problems by AChecker/the Nibbler accessibility scores of the web pages? How large are the correlation coefficients?

H10 The economic character of a country does not affect the web accessibility versus the alternative hypothesis: the wealthier a country is (GNI per capita), the fewer barriers will be found on its websites.

3.5.3 Correlation between the number of elderly citizens and barriers

Q11 Are there any correlation between the size of elderly population and the known problems by AChecker and the Nibbler accessibility scores? How large are the correlation coefficients? Does higher ratio of the elderly population produce accessible websites?

H11 Ratio of the elderly population is independent of the number of known problems versus these quantities are dependent.

Table 5 Average values and dispersions in case of groups of countries and groups of questions

	V4		N–W-EU		Total	
	Average	Dispersion	Average	Dispersion	Average	Dispersion
Accessibility	8.845	1.823	8.923	1.933	8.886	1.88
Experiences	7.125	3.734	7.329	3.536	7.232	3.631
Marketing	7.369	3.5	7.222	3.467	7.292	3.482
Technology	7.995	2.982	8.09	3.136	8.045	2.903

3.6 The applied statistical methods

This subsection presents the applied statistical methods and the references in which they can be found in greater detail. During the analysis of the data, we computed not only the values of the descriptive statistics, but we tested some hypotheses as well. In most cases, we apply two-sided alternative hypotheses with two-tailed tests. We computed the appropriate test statistics and the p -values belonging to them. The p value is the probability of error type one, namely the probability of having a value that is larger in absolute value than the value of the test statistics, assuming null hypothesis holds [51] (p. 334). Decisions were made on the standard significance level 0.05: if the p -value is less than 0.05, we rejected hypothesis H_0 , and in the opposite case, we accepted it.

Applying Nibbler and AChecker evaluating tools, we get numbers for the websites; therefore, we can compare the expected values of these numbers. Nibbler provides point values for the website from different aspects; consequently, we can apply two-sided two sample tests for comparing the expectations of the points [51] (p. 346). If the equality of the dispersion does not hold, the Welch test is applied for checking the equality of the expectations [51] (p. 347). In these cases, hypothesis H_0 was the equality of the expectations and the alternative hypothesis was its opposite. The AChecker test counts the number of errors; ergo, the equality of the expectations can be tested in the same manner.

As for the data gained by questionnaires, we investigated the equality of the ratios of the websites satisfying certain conditions. In these cases, hypothesis H_0 was the equality of the two proportions and the alternative hypothesis was supported by all other values [51] (p. 364). For investigating the link between two quantities, we applied Pearson correlation coefficients [51] (p. 432) and hypothesis H_0 was confirmed with a value of zero [51] (p. 436) as then the independence of the variables is true in terms of their Gauss distribution. We investigated rank correlations as well by testing the zero value of the Spearman rank correlation coefficient [51] (p. 691). Finally, we made clusters from the countries to see the similarity in groups. For this purpose, the generally accepted k -means clustering algorithm was applied [52]. All statistical computations were carried out by the statistical program package R (R version 3.5.0.) [53].

Table 6 p -values of the tests

	V4: N–W-EU
Accessibility	0.646
Experiences	0.357
Marketing	0.467
Technology	0.649

4 Results

4.1 Results Based on the Nibbler Tests

Answer to Q1 Grouping the questions by properties, we cannot find significant differences between the results of the groups in case of V4 and N–W-EU countries. If we compare the V4 and N–W-EU countries by the groups of questions, there are significant differences between the expectations of the question groups except the areas of Experiences and Marketing.

Table 5 shows that the largest averages belong to accessibility and the smallest ones to the experiences in almost all cases.

We performed the equality test in each question group. The p -values are displayed in Table 6.

Table 6 confirms that the expectations of the V4 countries can be considered the same as the expectations of the N–W-EU countries in all groups of questions. If we make the pairs of question groups, we have $\binom{4}{2} = 6$ pairs. Performing the tests of equalities of the expectations for all pairs, we have the following results (Table 7).

Table 7 shows that there is no significant difference between Experiences and Marketing scores, but all other pairs present significant differences. The statement is true for the V4 countries, for the N–W-EU countries and for all countries too. The best performing area is Accessibility, it is followed by Technology and the last areas are Marketing and Technology.

Answer to Q2 Investigating the correlations between the areas, we found the following correlation coefficients (Table 8). The scores of the different groups of questions are not independent except Marketing and Technology. Higher scores in one area imply higher scores in another one. Marketing and Technology are independent in the case of V4

Table 7 *p*-values belonging to the tests of equalities

	V4	N–W-EU	Total
Accessibility–experiences	1.18E – 16	2.48E – 16	2.09E – 31
Accessibility–Marketing	2.15E – 14	1.58E – 19	2.63E – 32
Accessibility–Technology	1.45E – 05	6.67E – 06	3.91E – 10
Experiences–Marketing	0.269032	0.59758	0.689924
Experiences–Technology	0.000118	0.000187	8.07E – 08
Marketing–Technology	0.003382	1.10E – 05	2.04E – 07

Table 8 Correlations between the results of groups of questions in case of V4, N–W-EU countries and in all countries

	V4	N–W-EU	Total
Accessibility–Experiences	0.602	0.496	0.538
Accessibility–Marketing	0.273	0.25	0.254
Accessibility–Technology	0.564	0.448	0.508
Experiences–Marketing	0.721	0.795	0.756
Experiences–Technology	0.559	0.709	0.641
Marketing–Technology	0.171	0.541	0.37

Table 9 Determination coefficients in the case V4, N–W-EU and all investigated countries

	V4	N–W-EU	Total
Accessibility–Experiences	0.363	0.246	0.289
Accessibility–Marketing	0.075	0.062	0.065
Accessibility–Technology	0.318	0.201	0.258
Experiences–Marketing	0.52	0.632	0.571
Experiences–Technology	0.313	0.502	0.411
Marketing–Technology	0.029	0.293	0.137

Table 10 *P*-values of the test of independence in the case of one-sided and two-sided alternative hypothesis

	Two-sided alternative hypothesis			One-sided alternative hypothesis		
	V4	N–W-EU	Total	V4	N–W-EU	Total
Accessibility–Experiences	0	0	0	0	0	0
Accessibility–Marketing	0.063	0.074	0.011	0.033	0.037	0.006
Accessibility–Technology	0	0.001	0	0	0.0005	0
Experiences–Marketing	0	0	0	0	0	0
Experiences–Technology	0	0	0	0	0	0
Marketing–Technology	0.249	0	0	0.125	0	0

countries, but considering N–W-EU countries or all data the statement does not hold. Testing the independence of the areas, i.e. the scores belonging to the separate aspects, the results are summarized in Table 10.

Table 8 presents usually medium positive correlations. The relation is stochastic, the scores change together in the same direction on average, and largest Accessibility implies largest Experience and Technology. In case of Accessibility and Marketing, Technology and Marketing correlations are rather small.

The determination coefficients are shown in Table 9.

Table 9 demonstrates that determination of one area by another is slight and usually is under 50% in most cases (Table 10).

Decisions are presented in Table 11.

In the columns concerning the two-sided alternative hypothesis, we can see that independence is acceptable in the case of Accessibility and Marketing if we investigate V4 countries or N–W-EU countries. But the *p*-values 0.063 and 0.074 are close to the standard 0.05 value. The number of web pages included in the study is 48 and 51, respectively. If we investigate all of them together, the total number is 99, and the dependence is demonstrable on the level of significance

Table 11 Decisions about the independence of areas (D = dependent, I = independent)

	Two-sided alternative hypothesis			One-sided alternative hypothesis		
	V4	N–W-EU	Total	V4	N–W-EU	Total
Accessibility–Experiences	D	D	D	D	D	D
Accessibility–Marketing	I	I	D	D	D	D
Accessibility–Technology	D	D	D	D	D	D
Experiences–Marketing	D	D	D	D	D	D
Experiences–Technology	D	D	D	D	D	D
Marketing–Technology	I	D	D	I	D	D

Table 12 Average values in the case of different error types

	V4	N–W-EU	Total
Known	45.426	47.686	46.602
Likely	0.894	1.902	1.418
Potential	710.681	487.941	594.765
Total	757.1	537.529	642.785

Table 13 Significance levels testing the equality of the numbers of errors in the web pages of V4 and N–W-EU countries

Error type	Known	Likely	Potential	Total	
				Two-sided	One-sided
<i>p</i> -value	0.886	0.297	0.005	0.082	0.041

Table 14 Number of guideline's subsections (error type) in case of significant differences between the V4 and N–W-EU countries

Number of error type	Interpretation of the success criterion	Error type	Sign. level
1.1.1.	Non-text content (level A)	Potential	0.02
1.3.1	Info and relationships (level A)	Potential	0.035
1.4.1.	Use of colour (level A)	Potential	0.028
1.4.5.	Images of text (level AA)	Potential	0.019
2.4.3.	Focus order (level A)	Potential	0.016
2.4.4.	Link purpose (in context) (level A)	Potential	0.030
3.1.2	Language of parts (level AA)	Potential	0.001
3.2.1	On focus (level A)	Potential	0.014
3.3.2.	Labels or instructions (level A)	Known	0.015

0.011. The independence of Marketing and Technology is acceptable in the case of one-sided and two-sided tests on the level of significance 0.249 and 0.125, respectively, in the case of V4 countries.

Table 15 Most frequent known errors in the case of different groups of countries and in the case in all countries

Type of known error	V4	N–W-EU	Total	Sign. level
1.1.1. Non-text content	11.208	29.686	20.727	0.1389
1.4.4 Resized text	10.437	5.392	7.838	0.258
1.4.6 Contrast (enhanced)	10.563	3.255	6.79	0.209

4.2 Results Based on the AChecker Tests

Answer to Q3 There are no significant differences between V4 and N–W-EU countries in the expected number of known and likely errors, but there is in the number of potential errors. The total number of the errors is significantly higher in V4 countries than in N–W-EU countries, but the difference is due to the potential errors. If we compare the V4 and N–W-EU countries grouping the test questions based on type, we get the following average values (Table 12). In the case of both V4 and N–W-EU countries, the most frequent **known** errors are “Non-text content”, “Contrast (Enhanced)” and “Resize text”. This emphasizes the similarity between the properties of the V4 and N–W-EU countries. The average values are given in Table 15.

If we compare the V4 countries and N–W-EU countries by grouping the errors by types, i.e. by the known, likely and potential error numbers, we get the results given in Table 13.

Table 16 Average values of the groups in case of known errors

Success criteria (level of conformance)	V4	N–W-EU	Total	<i>p</i> value
A	1.094	1.60866	1.359	0.350
AA	0.859	0.26	0.687	0.313
AAA	0.463	0.150	0.301	0.271

Table 17 Number of errors violating the different error types

Level	V4			N–W-EU		Comp.		Total	
	Mean	Dispersion	Max	Mean	Dispersion	Max	<i>p</i> value	Mean	Dispersion
A	26.25	34.54	213	38.608	86.81	371	0.35	32.616	66.75
AA	11.167	24.82	155	6.843	12.532	68	0.82	8.939	19.501
AAA	10.646	42.677	275	3.451	12.415	63	0.266	6.939	31.068

Table 18 Number of errors per error type

Known error type	Level of conformance	Number of errors
1.1.1. Non-text content	Level A	2052
1.3.1 Info and Relationships	Level A	376
1.4.4 Resize Text	Level AA	776
1.4.6 Contrast (Enhanced)	Level AAA	715
3.3.2 Labels or Instructions	Level A	260

Fig. 4 Data of the investigated websites

The data found Table 13 confirm that in the case of known and likely errors there are no significant differences between the groups of V4 and N–W-EU countries, but in the case of potential errors there is. If we consider all types of errors, we can state that the number of the errors of websites in V4 countries is significantly higher than that of N–W-EU countries provided we use one-sided tests (*p*-value is under the level 0.05). The difference, however, is not significant if

Fig. 5 Average data of the countries

we test non-equality by two-sided test. The last result can be explained by the large dispersions values.

Investigating the number of errors in the websites of V4 and N–W-EU countries, we can state that there are usually no significant differences between the V4 and the N–W-EU countries even from the perspective of questions. Performing the Welch test for every question separately (183 tests), we found significant differences in nine cases, and these are presented in Table 14. It should be highlighted that all errors except 3.3.2 are potential errors and only one known error can be found among them (Table 15).

Although the maximal number of the average error is higher in the group of N–W-EU countries than in the group of V4 countries, the opposite is true for the second and the third values in the rank. Moreover, the differences are not significant, as it can be seen from the values in the last column, a consequence of the large values in the standard deviations.

Answer to Q4 We cannot ascertain significant differences between the V4 countries and the N–W-EU countries by the level of conformances of known errors. Grouping the

questions by level of conformance (A, AA and AAA), we performed the equality test for the expectations of the number of errors in the case of known errors. The average values of the groups and the p-values are presented in Table 16.

Table 16 shows that in the case of the level A group V4 has better values, but in the case of levels AA and AAA the N–W-EU group has fewer errors. Performing the test of equality, we do not see significant differences between the groups of countries. The number of errors violating success criteria (level of conformance) A, AA and AAA is given in Table 17. We can conclude that there is no significant difference between the V4 countries and the N–W-EU countries. The average number of N–W-EU is larger in the case of error type A. The reverse is true in the case of error types AA and AAA. The maximal values are large as compared to the average, which causes large dispersions.

Table 19 Covariance, correlation and p-values between the results of Nibbler and AChecker tests investigating separate websites and countries

	Cov.	Corr.	p value
Separate websites	− 0.12	− 0.059	0.561
Countries	− 0.037	− 0.118	0.762

Fig. 6 Plotted ranks

Fig. 7 Plotted cluster numbers

Answer to Q5 Table 18 shows the five highest number of errors in the case of known errors based on the AChecker tests.

4.3 Comparison of the Nibbler and AChecker Tests' Results

Answer to Q6 Nibbler and AChecker data tests are independent in ranks and values by countries for each individual website. We compared the results of the Nibbler and AChecker tests for the same websites. No dependence can be found. We plotted the average values of the websites by Nibbler and AChecker tests, and the results are shown in Figs. 4 and 5. These figures demonstrate that no correlation can be seen between the results of the different methods.

The same conclusion can be drawn based on the calculations of the tests (see Table 19) (Figs. 6, 7, Table 20).

Table 21 demonstrates that the correlations are low, and we accept the hypothesis of the independence of both the rank and clusters on very high levels of significance.

4.4 Results Based on the Experts' Tests

Answer to Q7 Following Google Chrome SEE web application tests, we can ascertain that we found no lost information.

Table 20 Rank and the clusters of the countries according to the Nibbler and AChecker tests

	FIN	PO	HU	SWI	SWE	SLO	CZR	GER	AU
Nibbler average values	7.348	7.395	7.476	7.595	7.827	7.838	7.881	8.011	8.044
Rank	9	8	7	6	5	4	3	2	1
Nibbler cluster	3	3	2	2	1	1	1	1	1
AChecker average values	2.403	5.649	4.505	2.517	2.809	3.823	3.388	5.021	2.624
Rank	1	9	7	2	4	6	5	8	3
AChecker clusters	1	3	3	1	1	2	2	3	1

We can state that these websites can be used equally by users who have colour deficiency problems. The zero value in Table 22, column 3 confirms that there was no information loss when simulating four types of colour deficiencies (see Table 4, question 3).

By testing the websites with Variantor special glasses, we can state that we found no information loss, and these websites can be used equally well by users with colour deficiency problems.

Answer to Q8 Based on our questionnaires, there is no usual significant difference between the V4 and the N–W-EU countries. Results focusing on countries and groups of countries are found in Table 4. The results are summarized in Table 22.

Creating blind website versions has turned out to be a less convenient solution for providing accessibility. Usually the blind version is not updated frequently enough and contains old information. The applied aspects of measurement include the following: each link should be determined from

the link text. For example, “click here” or “here” is not understandable; that is, the user could ask where they are. “Success Criterion 2.4.4 Link Purpose (In Context), that is the purpose of each link can be determined from the link text alone or from the link together with its programmatically determined link context, except where the purpose of the link would be ambiguous for users in general (Level A)” [15]. Therefore, the values of the first and second columns of Table 22 are considered with a negative sign. Column six is not subject to analysis because of its redundant nature. Native language users of a country do not need the English version of a site for accessibility. Websites with a WAI logo were scored separately.

Only one Austrian healthcare-related governmental website contains a WAI accessibility logo among the 99 websites included in the study. This poor result may have two underlying reasons: either designers do not know about it or they do not consider it important. The fact that the WAI logo is rarely found on the investigated websites disproves the state-

Table 21 Results of testing the independence of ranks and clusters of countries provided by Nibbler and AChecker tests

	Covariance	Correlation	<i>p</i> value
Ranks	− 0.625	− 0.083	0.831
Clusters	0.083	0.104	0.791

Table 22 Ratio of web pages satisfying special requirements in the different countries and groups of countries

Question number	1 Blind version	2 Link purpose	3 Colour deficiency	4 CAPTCHA	5 Keyboard	6 English version	7 Responsive	8 Accessibility logo	Σ
HU	0	0.882	0	0.176	0.529	0	0.824	0	0.65
AU	0	0.875	0	0	1	0.125	1	0.125	1.25
CZR	0	0	0	0	0.875	0.125	0.875	0	1.75
FIN	0	0	0	0	0.7	0.9	0.8	0	1.5
PO	0	1	0	0.125	0.5	0	0.625	0	0.25
GER	0	1	0	0	1	0	1	0	1
SWI	0.125	1	0	0	0.688	1	0.5	0	0.06
SLO	0	0.615	0	0	0.769	0.077	1	0	1.15
SWE	0.125	1	0	0	0.625	0.25	0.75	0	0.25
V4	0	0.674	0	0.087	0.652	0.043	0.848	0	0.91
N–W-EU	0.059	0.784	0	0	0.784	0.549	0.765	0.019	0.73
Total	0.031	0.732	0	0.041	0.722	0.309	0.804	0.01	0.81

Table 23 Comparison of V4 and N–W-EU countries from the different perspectives (E = there is no significant difference, D = different)

Question number	1 Blind version	2 Link purpose	3 Colour deficiency	4 CAPTCHA	5 Keyboard	6 English version	7 Responsive
<i>p</i> value	0.095	0.220	1	0.032	0.147	7.49E – 08	0.303
Decision	E	E	E	D	E	D	E

Fig. 8 Box-plot of the AChecker known values in the function of web page sizes (large, medium, small) **Fig. 10** Box-plot of Nibbler accessibility scores in the function of the web page sizes **Fig. 9** Diagram of the AChecker known values in the function of the web page sizes (Kbytes) **Fig. 11** Diagram of the Nibbler accessibility scores in the function of the web page sizes (Kbytes)

ment by Goodwin et al. [33]: “website with WAI logo have better website accessibility scores”. We can state that it is not true that a website with a WAI logo has better accessibility scores.

Based on the last column of Table 22 (total score), it can be established that the best performing countries are the Czech Republic (total score: 1.75), Finland (total score: 1.5) and Austria (total score: 1.25); they are followed by Slovakia (total score: 1.15), Germany (total score: 1.0) and Hungary (total score: 0.65). Although the total score 0.91 of the V4

countries was higher than the total score 0.73 of the N–W-EU countries, there is no significant difference between them. Altogether, based on both the statistical analyses and the Nibbler results, the first three countries were Germany, the Czech Republic and Slovakia and statistics showed that the order in AChecker test result was Finland, Switzerland and Austria. This suggests that the Nibbler results are more similar to the findings of the questionnaires. This also demonstrates that for

Fig. 12 Box-plot of the AChecker known problems in the function of the GNI per capita

Fig. 13 Box-plot of Nibbler scores in the function of the GNI per capita

Fig. 14 Box-plots of numbers of the AChecker known problems in the function of the elderly population's ratio

a thorough evaluation automated tests and human feedback must complement and confirm each other.

Comparing the V4 and the N–W-EU countries, there are no usual significant differences between the ratios. The

Fig. 15 Box-plot of Nibbler accessibility scores in the functions of the elderly population's ratio

exceptions are the CAPTCHA function and English version availability as we take the previously mentioned calculation in Table 22 into account. The total score of the V4 countries is 0.91, whereas that of the N–W-EU countries is 0.73 (Table 23).

4.5 Results for External and Internal Characteristics

Answer to Q9 AChecker tests

Figures 8 and 9 show that in the case of the AChecker test a larger size does not imply more errors. The estimated value of the correlation coefficient is 0.027, and it is very close to 0. The p -value of the test is 0.791; therefore, we accept the hypothesis that these quantities do not correlate. This result is the opposite of what was concluded by Goodwin et al. [33].

Concerning the Nibbler accessibility test, the opposite results are shown in Figs. 10 and 11. The trend that is suggested by the values is that smaller web pages have slightly lower Nibbler scores on average as compared to larger websites. The categories of small/medium/large were further defined by the trisection size values.

The estimated value of the correlation coefficient is 0.226. Although it is not high, it is still significantly higher than in the case of AChecker scores. The p -value is 0.025, which is less than the usual 0.05 significance level; therefore, there is a slight correlation between size and Nibbler scores. The correlation is positive with larger websites having a higher Nibbler score. This shows that the creators of larger websites pay more attention to accessibility, but the number of mistakes does not increase with the size of the web page.

Answer to Q10 Fig. 12 presents the number of AChecker known errors in the function of the GNI per capita.

The correlation coefficient is estimated to be -0.097 ; thus, the hypothesis of independency can be accepted on a significance level of 0.169.

In the case of Nibbler tests, we can notice the opposite phenomenon. The estimated value of the correlation coefficient is 0.217, which proves a slight positive correlation between the economic potential and accessibility. The hypothesis of independence is rejected; in the case of one-sided alternative hypothesis, the p -value is 0.016, which confirms Goodwin et al. [33] too (Fig. 13).

Answer to Q11 The relationships between the ratio of elderly people and accessibility were investigated by AChecker and Nibbler as well. We have not found correlations. The box-plots are shown in Figs. 14 and 15.

The correlation coefficients are 0.166 and -0.039 , respectively, so the hypothesis of independency can be accepted on the usual significance levels.

5 Discussion

Access to healthcare-related websites is of utmost importance to citizens all over the world, yet the present research demonstrates that people with disabilities are often excluded. Many of the accessibility problems identified in this study can be fixed relatively easily and do not require huge redesigns of the site, for example, adding appropriate ALT attributes for images. ALT tags convey the text equivalent of the image on display and are necessary for those with disabilities as they may not have the ability to physically see the image.

Providing adequate access especially to healthcare-related information is inarguably a vital task in the field of information technology. Our present research demonstrates, nevertheless, that people with disabilities are often excluded. For instance, a surprisingly large number of errors detected by AChecker (see Table 18) concern a major area of improvement, namely Alt attributes. ALT tags convey the text equivalent of the image on display and are necessary for those users who due to some visual impairment are not able to physically see the actual pictures themselves on the site. Alt attributes are responsible for what the screen reader can read in such cases as an alternative text about an image. The guidelines for implementation, handling and problem resolution are clearly stated in WCAG 2.0. [15], which already classifies Alt tags as a Level A area. Identifying and remedying Alt tag issues, as accessibility problems, is a relatively simple task and does not require huge redesigns of the site.

“Success Criterion 1.1.1 Non-text Content (Level A): All non-text content that is presented to the user has a text alternative that serves the equivalent purpose” [15]. “Success Criterion 1.3.1 Info and Relationships (Level A): Information, structure and relationships conveyed through presentation can be programmatically determined or are available in text” [15]. “The intent of this Success Criterion is to ensure that information and relationships that are implied by visual or auditory formatting are preserved when

the presentation format changes. For example, the presentation format changes when the content is read by a screen reader or when a user style sheet is substituted for the style sheet provided by the author” [15]. “Success Criterion 1.4.4 Resize text (Level AA): Except for captions and images of text, text can be resized without assistive technology up to 200% without loss of content or functionality” [15]. “The intent of this Success Criterion is to ensure that visually rendered text, including text-based controls can be scaled successfully so that it can be read directly by people with mild visual disabilities, without requiring the use of assistive technology such as a screen magnifier. Users may benefit from scaling all content on the Web page, but text is most critical” [15]. “Success Criterion 1.4.6 Contrast (Enhanced) (Level AAA): The visual presentation of text and images of text has a contrast ratio of at least 7:1” [15].

From the data collected and analysed on the basis of both our questionnaires and the figures produced by the applied automated tools, we accept hypotheses: H1, H4, H5, H6, H7, H8 and H11.

The answer is mixed for H3. For two subproblems, the first part of H3 is accepted; for another two, it is rejected. A few questions of the second part of H3 are rejected, but in the most cases they are accepted. H9 and H10 are mixed again.

5.1 Accepted Hypotheses

Taking into account our response to research question Q1, documented in Tables 6, 7 and 8, it can be claimed that there is no significant difference between the expectations in the groups of countries. So, we formulate Thesis T1. T1: The expectations of the scores of the V4 countries and the N–W-EU countries regarding the aspects separately are equal. There is no significant difference. H4 is also acceptable. We formulate T4: The average number of errors is equal comparing V4 and N–W-EU countries in case of each conformance level versus they are different. The most frequently violated success criterions are 1.1.1 Non-text content, 1.3.1 Info and Relationships, 1.4.4 Resize Text and 1.4.6 Contrast (Enhanced) (see Table 18), which are all known problems, and they belong to Principle 1. So, H5 is true. T5: Most errors originated when web developers violated Principle 1—that is, web developers do not provide a text alternative to any non-text content. We accept H6 because the independence of the two evaluations can be accepted in case of the both websites and countries. It is proved by Figs. 4, 5, 6, 7 and Tables 19, 20, 21. T6: There is no correlation between Nibbler and AChecker evaluations, either in ranks or in values. Regarding the answers to questions Q7 and concerning H7 hypothesis, we can state the following thesis about the colour design of the investigated websites. T7: Users with colour deficiencies are able to use the investigated websites the same way as people with no vision impairment (here we applied

SEE and Variantor glasses). T8: There is no significant difference between the N-W-EU and V4 countries based on the evaluation of the experts' questionnaire (see Table 22). T11: There is no connection between the ratio of elderly populations and web accessibility, even though elderly people need more accessible web pages (see Figs. 14, 15).

5.2 Mixed Cases

We agree with the first part of H2 because the scores show medium positive correlations numerically in the case of V4, N-W-EU countries and all countries. (See Tables 8, 9.)

We reject the second part of H2. In most cases, the correlation between the different aspects cannot be considered zero; therefore, the scores gained at different aspects are not independent. In the case of the two-sided alternative hypothesis, the only exception is the Marketing–Technology pair in the case of the V4 countries. Therefore, in the case of any other pairs we can state that on average a larger score for a certain aspect implies larger score for another one.

Consequently, we formulate T2 as follows: Better technology results in higher accessibility in numerical values. Although numerically the average number of errors of the N-W-EU countries is higher in the case of level A than that of V4 countries and the opposite is true for level AA and AAA, statistically there are no significant differences between the expectations. (Tables 16, 17 show the results.) Larger score for one aspect implies larger score for another one, except for Marketing and Technology in the V4 countries.

T3: There are no significant differences in the case of known and likely problems. Considering the numbers of errors of all types, we can state that the expectation is larger in case of V4 countries than in the case of N-W-EU countries. The most frequently affected guidelines are Non-text content, Resized text and Contrast. Non-text content means that there is no verbal explanation for the figures. Success Criterion “1.4.4 Resize text (Level AA)” and Success Criterion “1.4.6 Contract (Enhanced) (Level AAA)” were detailed in the first paragraph of Discussion. The expected numbers of the above errors can be considered the same in case of V4 and N-W-EU countries.

The scores are not country specific and not country group specific. There is a high degree of variation among individual web pages in the same countries. This implies that there are very well and there are particularly poorly performing sites in each country. There is no correlation between Nibbler and AChecker results. Although they do not measure the results based on the same algorithms, it is necessary to compare them since they complement each other and they allow us to ascertain which one yields similar results to the applied questionnaires. AChecker tests investigate mainly programming solutions, while Nibbler takes user experience into account, and therefore correlates more with the conclusions drawn on

the basis of user feedback. This comparison helps web developers decide which test serves their goals better, depending on what they want to test.

We also performed the tests designed by Goodwin et al. [33] for investigating the relation between size and accessibility. The result is mixed: we accept the hypothesis of the independence of size and accessibility if we measure accessibility by AChecker, but if independence is measured by Nibbler we reject the previous statement and accept the notion that size and accessibility go hand in hand. T9: The size of the web page does not correlate with the number of the known errors indicated by AChecker, and larger websites do not contain more errors than smaller websites. Larger websites, however, have better accessibility if the accessibility is characterized by Nibbler scores.

Q10 is motivated again by Goodwin et al. [33]. We investigated the relation between economic potential and accessibility. We found that H0 is accepted if accessibility was measured by AChecker, and H0 is rejected if the accessibility was characterized by Nibbler scores.

T10: The effect of the economic potential of a region cannot be demonstrated in AChecker tests, but it can be presented in the results of the Nibbler test. The countries with higher GNI per capita have more accessible websites by Nibbler.

5.3 Recommendations

Unfortunately, in the identified known, likely and potential problems, errors have not changed for the last 10 years compared to our earlier research [38, 41]. Our recommendation still holds [38, 41]:

1. Provide alternative short text for all non-text elements (for example, image), if one is not able to write a short description, use long texts,
2. Use relative positioning, rather than absolute,
3. The content of the site should be accessible without using the mouse (the appearance of the content) and should not depend on JavaScript event handlers/modal windows,
4. Use <label> tag defining the elements of form and where it is not possible use “title” attributes,
5. The texts of the references have to be understandable without their contexts,
6. In the <html> tag, identify the primary natural language using the “lang” attribute, and specify the base direction of directionally neutral texts using “dir” attributes,
7. Provide summaries for tables using “summary” attributes in <table> tags,
8. Put some separating characters between links,
9. Check <title> tags whether they identify the subject of the web page correctly,

10. Html tags must be closed correctly so that assistive technologies can parse the content accurately [38, 41].

We are hereby extending our earlier recommendations by adding Point

11. All websites must be made responsive, so that they can be accessed by any device or platform independently of screen size.

The remaining part of this section describes how some of our own recommendations conform to the WCAG 2.0. The first recommendation is relevant for users who need a screen reader. It conforms to the WCAG 2.0 1.1.1 success criterion.

The third one is also necessary if somebody is not able to use a mouse. The first part fits the 2.1 guideline of WCAG 2.0, which suggests making all functions available for keyboard input.

Our fourth recommendation fits the 2.4.2 and 2.4.6 success criteria. Those help navigate the user. Both describe the topic of the page.

The fifth one conforms to the 2.4.4 and 2.4.9 criteria of WCAG 2.0. Both describe how to identify links.

Our sixth recommendation meets 3.1.1 and 3.1.2 criteria of WCAG 2.0. Those explain use of language. It is also important for screen readers to identify the human language of the page or the human language of a part of the text.

The ninth one fits the 2.4.2 criterion of WCAG 2.0. It defines the purpose or the topic of the page.

Our tenth recommendation meets Guideline 4.1 of WCAG 2.0 for the purpose of using the web easily for future technologies.

6 Conclusions

The element of innovation this research presents is that it looks at some relevant accessibility issues concerning health-related websites in different historical and economic regions in Europe. Such a detailed statistical comparison has not yet been made in Europe for testing healthcare-related websites, even though their regular revision seems necessary owing to the demands of the growing elderly population. The authors would also like to draw attention to the fact that the two applied automatic testers yielded different accessibility results for the same sites; therefore, statistical evaluation based on user response has proved inevitable for thorough analysis.

This research demonstrates that a vast majority of healthcare-related websites in the North–West-European (N–W-E) and V4 countries (Czech Republic, Hungary, Poland, Slovakia) do not meet industry-related accessibility guidelines. Most of the websites in this research violated

the same guidelines because their designers did not take the needs of people with disabilities into account. Consequently, these websites contain almost the same, very common errors. Accessibility errors should be identified by software testing tools based on WCAG 2.0 and user response. We did not find differences between the results of N–W-EU and V4 countries, in spite of the fact that the GDP and GNI of the N–W-EU countries are much higher than the GDP and GNI of V4 countries.

First of all, the findings in this paper might help web designers by providing a better understanding of their websites and can also be used to facilitate assessing web accessibility. Second, if service or software development has been carried out without taking the requirements of accessibility into consideration, it is certain that the result will not be an accessible solution and accessibility problems will emerge. The accessibility problems will remain until an expert in accessibility gives instruction to web designers; otherwise, web designers will have to familiarize themselves with all accessibility guidelines. There is no doubt that users with special needs are not involved in the design process and the testing of the usability of websites. Furthermore, it is clear that the cost of developing accessible websites from the first step of the design process is much cheaper than to redesign or transform the badly designed website and make it accessible.

All our measurements and research are based on the WCAG 2.0 guidelines; meanwhile, the newest version WCAG 2.1 [26, 27] has just been released. In the future, this research should be repeated based on WCAG 2.1, but currently there is a lack of effective automatic test tools based on WCAG 2.1.

In conclusion, a site that meets accessibility requirements opens the market to a wider range of customers. If we look at the growing number of elderly people, their demand for accessible websites can be expected to increase. If companies do not focus on these requirements, they will lose a high number of their customers. The overall benefit is that of increased value. The issues identified in this study can also be considered to be independent of financial considerations, since they require only a bit of attention and the willingness to identify the needs of people with disabilities.

Acknowledgements Open access funding provided by University of Pannonia (PE). The first author would like to thank the financial support of the “King Salman Award for Disability Research” in 2018. The “King Salman Award for Disability Research” was initiated by His Royal Highness, Prince Sultan bin Salman bin Abdulaziz, Chairman of the Board of Trustees of the King Salman Center for Disability Research.

Compliance with ethical standards

Conflicts of interest The authors declare no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Sik-Lányi, C.: Colour-Fidelity and Barrier-Free Design of Virtual Worlds and Games. Habilitation theses, University of Pannonia, Veszprem, Hungary (2017)
- European Commission (2003): A Europe accessible for all, Brussels (2010). http://www.accessibletourism.org/resources/final_report_ega_en.pdf. Accessed 17 Nov 2018
- European Commission (2007): Ageing well in the information society, an i2010 initiative action plan on information and communication technologies and ageing, COM (2007)332 final. Brussels (2007). <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2007:0332:FIN:EN:PDF>. Accessed 17 Nov 2018
- Szabo, J.; Pall, A.; Sik Lányi, C.; Pataky, I.: Development of a computer controlled cognitive diagnostics and rehabilitation method for the use on the Internet. In: Sharkey P, Sik Lányi, C.; Standen, P.J. (eds) Proceedings of the 4th International Conference on Disability, Virtual Reality and Associated Technologies ICDVRAT 2002, Veszprém, Hungary, Sept. 18–20. 2002, pp. 149–154. The University of Reading, Reading, UK (2002)
- OECD data. <https://data.oecd.org/pinboard-editor/>. Accessed 27 Aug 2018
- GDP Growth Rate, Europe. <https://tradingeconomics.com/country-list/gdp?continent=europe>. Accessed 27 Aug 2018
- Population Europe. <http://www.worldometers.info/population/europe/>. Accessed 27 Aug 2018
- Europe, Country, Demographics, Age structure: https://www.indexmundi.com/site_map.html. Accessed 15 Mar 2019
- The World Bank, GNI per capita, Atlas method (current US\$) in 2017: <https://data.worldbank.org/indicator/NY.GNP.PCAP.CD?view=map>. Accessed 15 Mar 2019
- The World Bank, GNI per capita, PPP (current US\$) in 2017: <https://data.worldbank.org/indicator/NY.GNP.PCAP.PP.CD?view=map>. Accessed 15 Mar 2019
- Karner, C.; Horváth, O.: A stroke megbetegedés gazdasági és társadalmi jelentősége. (The economic and social significance of stroke disease). *IME* **1**, 47–50 (2014)
- Universal Desing. <http://www.universaldesign.com/>. Accessed 17 Nov 2018
- Accessibility for the Disabled, A Design Manual for a Barrier Free Environment. <http://www.un.org/esa/socdev/enable/designm/>. Accessed 17 Nov 2018
- WCAG 1.0: <https://www.w3.org/TR/WAI-WEBCONTENT/#priorities>. Accessed 17 Nov 2018
- WCAG 2.0: <https://www.w3.org/TR/WCAG20/>. Accessed on 17 Nov 2018
- Government of Canada. <https://www.canada.ca/en/treasury-board-secretariat/topics/government-communications.html>. Accessed 17 Nov 2018
- WAIC.JP (in Japanese). Web Accessibility Infrastructure Commission. https://waic.jp/guideline/jis-x-8341-3_2010/. Accessed 27 Aug 2018
- Stephen, T: Springer, SSB BART Group (2010). <http://www.ssbartgroup.com/reference/laws-and-standards/section-508-of-the-rehabilitation-act/>. Accessed 21 Aug 2018
- WebAIM (Web Accessibility in Mind): World laws introduction to laws throughout the world. <https://webaim.org/articles/laws/world/>. Accessed 27 Aug 2018
- Digital Accessibility Laws Around the Globe. Law Office of Lainey Feingold. <https://www.lflegal.com/2013/05/gaad-legal/> (2013) accessed 27 August 2018
- Nakata, K.: Deadlines Loom for Canada's Web Accessibility Law. HiSoft (2013). <http://blog.hisoftware.com/2013/accessibility/deadlines-loom-for-canadas-web-accessibility-laws>. Accessed 27 Aug 2018
- ADA.gov. <https://www.ada.gov/archive/t3highlight.htm>. Accessed 27 Aug 2018
- Americans with Disabilities Act Title III Regulations. Nondiscrimination on the Basis Of Disability in Public Accommodations and Commercial Facilities. https://www.ada.gov/regs2010/titleIII_2010/titleIII_2010_regulations.htm. Accessed 27 Aug 2018
- News European Parliament. MEPs vote to make online public services accessible to everyone. 26 February 2014. <http://www.europarl.europa.eu/news/en/press-room/20140220IPR36573/meps-vote-to-make-online-public-services-accessible-to-everyone>. Accessed 17 Nov 2018
- EUR-Lex. Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies. Published in the 'Official Journal of the European Union' on 2.12.2016. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016L2102>. Accessed 17 Nov 2018
- WCAG 2.1. <https://www.w3.org/TR/WCAG21/>. Accessed 15 Mar 2019
- What's new in WCAG 2.1. <https://www.w3.org/WAI/standards-guidelines/wcag/new-in-21/>. Accessed 15 Mar 2019
- Kuzma, J.M.: Accessibility design issues with UK e-government sites. *Gov. Inf. Q.* **27**, 141–146 (2010). <https://doi.org/10.1016/j.giq.2009.10.004>
- Yu, D.S.; Parmanto, B.: US state government websites demonstrate better in terms of accessibility compared to federal government and commercial websites. *Gov. Inf. Q.* **28**, 484–490 (2011). <https://doi.org/10.1016/j.giq.2011.04.001> accessed 26 August 2018
- Olalere, A.; Lazar, J.: Accessibility of U.S. federal government home pages: section 508 compliance and site accessibility statements. *Gov. Inf. Q.* **28**, 303–309 (2011). <https://doi.org/10.1016/j.giq.2011.02.002>
- Youngblood, N.E.: Revisiting Alabama state website accessibility. *Gov. Inf. Q.* **31**, 476–487 (2014). <https://doi.org/10.1016/j.giq.2014.02.007>
- Wentz, B.; Lazar, J.; Stein, M.; Gbenro, E.H.; Ramsey, A.: Danger, danger! Evaluating the accessibility of Web-based emergency alert sign-ups in the northeastern United States. *Gov. Inf. Q.* **31**, 488–497 (2014). <https://doi.org/10.1016/j.giq.2014.02.010>
- Goodwin, M.; Susar, D.; Nietzio, A.; Snaprud, M.; Jensen, C.S.: Global web accessibility analysis of national government portals and ministry web sites. *J. Inf. Technol. Polit.* **8**, 41–67 (2011)
- Zeng, X.; Parmanto, B.: Evaluation of web accessibility of consumer health information websites. *AMIA Annu. Symp. Proc.* **2003**, 743–747 (2003)
- Oliveira, A.D.A.; Eler, M.M.: Strategies and challenges on the accessibility and interoperability of e-government web portals: a case study on Brazilian Federal Universities. In: 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), Turin, pp. 737–742 (2017). <https://doi.org/10.1109/compsac.2017.222>
- Kaoun, F.; Almourad, M.B.: Accessibility as an integral factor in e-government web site evaluation: the case of Dubai e-government. *Inf. Technol. People* **27**, 208–228 (2014)

37. Sik-Lányi, C.; Szucs, V.: Are the WEB-pages of the Hungarian Universities accessible for students with color deficiency? *Óbuda Univ. e-Bull.* **8**(2), 53–58 (2018)
38. Sik Lányi, C.; Forrai, S.; Czank, N.; Hajgató, A.: On developing validator software XValid for testing home pages of Universal Design. In: *Universal Access in HCI, PART I, HCII 2007, Lecture Notes in Computer Science, LNCS 4554*, pp. 284–293 (2007)
39. Sik Lányi, C.: Multimedia software interface design for special needs users. In: Mehdi Khosrow-Pour, D.B.A. (ed.) *Encyclopedia of Information Science and Technology*, 2nd edn, pp. 2761–2766. IGI Global, Hershey (2008). <https://doi.org/10.4018/978-1-60566-026-4.ch440>
40. Mátrai, R.; Kosztyán, Z.T.; Sik Lányi, C.: Navigation methods of special needs users in multimedia systems. *Comput. Hum. Behav.* **4**, 1418–1433 (2008)
41. Czank, N.; Sik, A.; Sik Lányi, C.: Testing the accessibility of WEB sites. *Int. J. Knowl. Web Intell. (IJKWI)* **1**, 87–98 (2011)
42. European Commission, 2014: Priority, Digital Single Market, Bringing down barriers to unlock online opportunities (2014). http://ec.europa.eu/priorities/digital-single-market_en. Accessed 27 Aug 2018
43. European Commission, Press release, 2015: Commission proposes modern digital contract rules to simplify and promote access to digital content and online sales across the EU (2015). http://europa.eu/rapid/press-release_IP-15-6264_en.htm. Accessed 27 Aug 2018
44. European Commission, Press release, 2016: Commission proposes new e-commerce rules to help consumers and companies reap full benefit of Single Market (2016). http://europa.eu/rapid/press-release_IP-16-1887_en.htm. Accessed 27 Aug 2018
45. Online Kiskereskedelmi Index-1 (Online Retail Trade Index-1). <http://www.gkidigital.hu/2016/05/26/22-millioszor-vasaroltunk-tavaly-az-interneten>. Accessed 27 Aug 2018
46. Online Kiskereskedelmi Index-2 (Online Retail Trade Index-2). <http://www.gkidigital.hu/oki/>. Accessed 27 Aug 2018
47. AChecker. <https://achecker.ca/checker/index.php>. Accessed 17 Nov 2018
48. Nibbler. http://nibbler.silktide.com/en_US. Accessed 17 Nov 2018
49. Google Chrome SEE app. <https://chrome.google.com/webstore/detail/see/dkihcccbkkakbpikjmbnmbamkgbjfcdn>. Accessed 27 Aug 2018
50. Variantor—an experience based tool to aid universal design. <http://variantor.com>. Accessed 27 Aug 2018
51. Walpole, R.E.; Myers, R.H.; Myers, S.L.; Ye, K.: *Probability and Statistics for Engineers and Scientists*. Pearson Prentice Hall, London (2007). <https://www.csie.ntu.edu.tw/~sdlin/download/Probability%20&%20Statistics.pdf>. Accessed 17 Nov 2018
52. Hartigan, J.A.; Wong, M.E.: Algorithm AS 136: a K-means clustering algorithm. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1**, 100–108 (1979)
53. R version 3.5.0. The R Foundation for statistical computing (2018). <https://www.r-project.org/>. Accessed 17 Nov 2018