Accessibility in senior-oriented websites: a comparative study

Andreas P. GIANNAKOULOPOULOS

Senior Researcher
New Technologies Laboratory
Faculty of Communication and Media Studies
University of Athens
5 Stadiou str., 105 62, Athens, Greece
tel: + 30 210 3689426
email: agsp96@media.uoa.gr

Keywords: *seniors, accessibility, usability, web development, design for all.*

Introduction

Starting off in eEurope's 2002 action plan [2], the European Commission has adopted the World Wide Web Consortium's (hereafter Consortium) Web Accessibility Initiative (WAI) [14] and especially the set of guidelines known as Web Content Accessibility Guidelines (version 1.0) which concern the construction of accessible websites [15]. The objective of this decision was to avoid the exclusion of any citizen group in the raising Information Society and was triggered with respect to such social groups as those of the disabled persons and the elderly [2]. The scope of implementation of the guidelines expands gracefully from public sector websites to the commercial ones and, hopefully, to all websites developed from and for the European population.

In this paper the results and conclusions of a comparative accessibility evaluation study are presented. The study was conducted in the year 2005 and focuses to the comparison of Greek and International senior-oriented websites.

The results of the study reveal that Greek websites fall short of the International websites as far as the implementation of WCAG is concerned, which is in compliance with other aspects of the digital landscape of Greece [8]. Beyond the comparison between categories of websites, the sample as a whole was evaluated by the use of the ordinal scale which was constructed from the numerical data obtained during the research. The outcome of this overall evaluation was that contemporary senior-oriented websites do not exploit the advantages offered by the web development accessibility techniques. This conclusion is in accordance with other recent surveys in the field [3, 16] and attests the widely spread opinion that a large proportion of the publicly available hypertextual content exhibits accessibility and usability problems, particularly for special social groups such as the senior citizens.

Objective

The objective of this study on accessibility is firstly to create and express in measurable terms the degree of compliance of websites of the sample with the internationally accepted patterns [17] and accessibility guidelines and secondly to allow the comparison between the degree of compliance of Greek social services websites (especially those websites which concern senior citizens) with the respective degree of International websites.

Studies on accessibility are a relatively new ground of implementation of scientific methods and this fact leads to two crucial points: (a) although various

patterns of measuring and assessing have been developed, none of them has been entirely accepted as the general pattern, which means that an accurate scale does not exist, and (b) a significant increase in studies on accessibility is observed, which are conducted with various tools, methodologies and –respectively– assessment scales, which in turn means that these measurements are not homogenous and therefore not comparable to each other. As it has characteristically been noted, «*there are different views about the best evaluation method, yet there are relatively few studies that have been conducted comparing Accessibility Evaluation Methods. A review of the small number of papers available suggests that comparisons are difficult because different methods measure different variables*» [7].

Research Tool

For the needs of this study on accessibility, a special tool -which provides threefold web accessibility valuation [5]- was developed and implemented . This tool follows to a great extent the Consortium's guidelines regarding the proposed methodology for checking website accessibility [13] and incorporates measurement techniques proposed in international literature [9, 12].

The tool consists of three measurement procedures which give arithmetic values as results, whereas a unifying equation gives a final grade for each website assessed. The three measurement procedures concern: (a) the automatic code evaluation, where the code compatibility with the objective criteria of the accessibility guidelines is checked, (b) the markup language and style sheets validity check, where the code syntax validity is checked according to the standards proposed by the Consortium, and (c) the grading of each website under inspection according to the subjective criteria by one or more analysts, where the compliance with accessibility guidelines for which an automatic check is not technically feasible is checked. The results of each measurement are assessed and aggregated to give the final grade.

The first trial, which concerns the automated accessibility evaluation, is utilized by the use of the automated tool Ocawa (Operational Control and Analysis for Web Accessibility) [4]. The selection of this particular tool is based upon the following reasons: (a) it is a European-based project, (b) it has been developed by reliable organizations, including France Télécom, (c) it has been recently developed (after the year 2000), thus it embodies all the latest advancements with respect to web accessibility issues, (d) it provides multiple pages evaluation per request, (e) it evaluates the HTML code against a rather rich "pool of rules" taking advantage of an algorithm based on Artificial Intelligence concepts, (f) it was proved rather stable during the beta-testing period and (g) it outputs an informative results report which provides useful statistical data. The mechanism returns both the number of tags per evaluated page and the number of remarks (i.e. number of errors found). The value of the composite variable "error rate" is calculated from the division of the total number of remarks (sum of remarks) by the total number of tags (sum of tags).

The second trial includes the validity checks with respect to the formally established grammars for markup and style sheets, known as Web standards [17]. The semi-automatic mechanisms proposed for those tests are the ones provided as services by the W3C, namely the Markup Validation Service and the CSS Validation Service, respectively. The final output of the second trial is the value of the variable named "multiplier", the range of which is the set {1, 2, 4}. The final value is calculated by multiplying the values of the Boolean variables "grade of markup validity" and "grade of style sheet validity", the range of values for

each one of them being the set {1, 2}. Validity value equal to 1 denotes approved usage of the respective grammar and validity value equal to 2 denotes erroneous syntax with respect to the corresponding standards.

The third trial covers the proposed in the relevant bibliography manual evaluation by experts [9, 12, 15]. The human review within the threefold evaluation framework turns into a measurable trial by the assignment of grades to websites with respect to subjective criteria for content accessibility. The checklist of criteria for the third trial stems from the subjective guidelines and can be found in the checklist proposed by the W3C, along with relevant usability criteria [6, 11], classified under five broad categories for analysis and examination. Each website under evaluation receives a grade in each of the five categories, reflecting the extend to which the particular website fulfils the criteria falling under the specific category. The grade scale for each category is the set $\{n \in N: 0 \le n \le 5\}$, where a grade equal to 0 corresponds to the ideal status, whereas a grade equal to 5 corresponds to the worst case. The five categories for the subjective evaluation are the following: (a) navigation, (b) content, (c) structure, (d) requirements and (e) support.

Recalling that the objective of the threefold evaluation platform is to produce a numerical value reflecting the accessibility level of each website, in order to promote comparability and gradation among websites, a unifying equation has been developed which integrates the outputs of the various trials and produces a single numeric grade for each website (in the form of penalty degrees). The formula of the equation reads as follows:

$$F = \left(ERR_{rate} \cdot M \cdot W_1 + E_{avg} \cdot W_2\right) \cdot 100$$

Where:

F (final score): The final score (penalty grade) outputted by the integrated evaluation tool.

ERRrate (error rate): The value of the error rate variable, as received from the first trial.

M (multiplier): The value of the multiplier variable, as received from the second trial.

W1 (weight 1): Weighting factor for the conjunct outcome of the first and second trial. The predefined value for this factor is equal to 0,7.

Eavg (expert average grade): The average grade of all five categories for each website, assigned by the expert analysts as part of the third trial.

W2 (weight 2): Weighting factor for the outcome of the third trial. The predefined value for this factor is equal to 0,3.

The output of the research tool is a final grade for each evaluated website which denotes degrees of error, i.e. the greater the final grade, the more accessibility problems for that particular website.

Methodology

The stages of developing this accessibility study were the following: initial planning, tool selection, pilot implementation, special tool developing, sample selection, measuring, statistical analysis of the results, conclusions.

Selecting the sample was one of the most difficult parts of the study. Since the main objective of the study was to record the guidelines implementation in the Greek part of the Web and compare it with the international situation, the

composition of a sample consisting of foreign websites which are strictly focused on the third age and Greek websites indirectly related to senior citizens (e.g. insurance funds websites) would make the comparison between the two samples unequal. Bearing this in mind, it was decided to use a mirror-sample, where each category of the sample (International websites, Greek websites) is the symmetrical image of the other, as far as the main sampling characteristic (i.e. degree of popularity of the websites of these specific categories) is concerned.

The technical criterion used for the aforementioned purpose was the degree of popularity of each website, according to the data of the international web company Alexa, the measurements of which are still a dependable source of the degree of popularity of websites and have often been accepted by the academic community.

The popularity criterion by itself was not adequate to select the websites of the sample, as, firstly, the measurements of Alexa concern all websites visited by the owners of its toolbar –which is a quite large part of the Web– and, secondly, the company gives access in list-form data only for the 500 websites with the highest grades. The combination of these two constrains made necessary the mingling of the popularity criterion with the content criterion, so as to eventually compose a sample which would represent the websites indirectly related to senior citizens. Given the fact that the content criterion is used to create directories, the problem was restricted to selecting the categories and consequently selecting the most appropriate directory among those available.

The directory finally selected was the Open Directory [10] for three reasons: (a) the directory Alexa uses for the navigation service in the database with the statistic data is the Open Directory, (b) the Open Directory is an acknowledged directory with high degree of acceptance from reliable institutions (for example, it is the basis of Google Directory), and (c) The Open Directory is the product of collective and volunteer work of a huge number of contributors, which up to a point guarantees not only polyphony but also its independence from advertising and commercial interventions.

Through the combination of these two criteria, seven categories of websites indirectly related to senior citizens were selected, which were further divided into categories of high and low interest, with a ratio 2/1 in favor of the high interest categories. The high interest categories included Health, Home, News, and the low interest categories included Shopping, Recreation and Sports, Business and Society. For each of the categories of the first group the first 20 websites according to the Alexa measurements were selected, whereas for each of the categories of the second group the first 10. The number of websites in each category was equally divided between Greek and International websites. The final sample consisted of 100 Greek and 100 International websites indirectly related to senior citizens.

All the websites in the sample were checked analytically with the Threefold Web Accessibility Evaluation tool. By using the equation of calculating the final grade, an arithmetic value (the grade of each website in the sample) was assigned. This grade expresses the size of accessibility problems and therefore the higher the final grade of a website, the bigger the accessibility problems it faces. The results of each individual check, as well as the final grades that arose, were processed statistically in order to lead to inferences.

Results

The results of the first trial, which concerns the code evaluation by the mechanism of Ocawa, were recorded in three variables: (a) the total tags variable,

which represents the number of tags used by the web developer for the construction of the evaluated pages, (b) the total remarks variable, which represents the number of remarks (i.e. found errors) returned by the evaluator and (c) the error rate variable, which represents the error-per-tag rate and is calculated from the division of variable [b] by variable [a]. The following table depicts the mean and the standard deviation of all three variables for the total sample with respect to the first trial.

	N	MEAN	STD. DEVIATION
Total Tags	200	5373,98	4255,621
Total Remarks	200	1767,43	1646,188
Error Rate	200	,338104037410	,1696751955133
Valid Cases	200		

(Table 1) – Trial One – Total Sample

Respectively, the following two tables depict the results of the first trial for the International and the Greek websites.

	N	MEAN	STD. DEVIATION
Total Tags	100	6364,65	3862,475
Total Remarks	100	1764,57	1433,840
Error Rate	100	,271582138900	,1245790091706
Valid Cases	100		

(Table 2) – Trial One – International websites

	N	MEAN	STD. DEVIATION
Total Tags	100	4383,31	4416,125
Total Remarks	100	1770,28	1841,556
Error Rate	100	,404625935920	,1827853823251
Valid Cases	100		

(Table 3) – Trial One – Greek websites

The variable which reveals the results of the first trial is the “error rate”. The means of the variable are 0,271 in the case of International websites and 0,404 in the case of the Greek ones, leading to the conclusion that the first group outranks the second with respect to the code evaluation trial.

Given that the two samples are mutually independent (i.e. the performance of each group is neither related to nor affected by the performance of the other), the appropriate statistic for the testing of the hypothesis that the difference in means should be attributed to the performance of each group (alternative hypothesis) and not to circumstantial factors (null hypothesis) is the two independent samples t-test. Since the t-test p-value (0,000) is lower than 0,05 (0,000 < 0,05) and further supported by the fact that the zero value is not included within the 95% confidence interval, we accept that the difference of means is statistically significant and is attributed to the performance of the websites of each group.

The second trial concerns the validity checks with respect to (a) the markup language and (b) the cascading style sheets (CSS). The results of the second trial were recorded in three variables: (a) the validity grade with respect to the markup language, (b) the validity grade with respect to the style sheets and (c)

the combined result of the previous two checks calculated by the multiplication of the validity grades and stored in the variable “multiplier”. The following table depicts the mean and the standard deviation of all the three variables for the total sample with respect to the second trial.

	N	MEAN	STD. DEVIATION
HTML validity grade	200	1,96	,196
CSS validity grade	200	1,66	,475
Multiplier	200	3,26	,999
Valid Cases	200		

(Table 4) – Trial Two – Total Sample

Respectively, the following two tables depict the results of the second trial for the International and the Greek websites.

	N	MEAN	STD. DEVIATION
HTML validity grade	100	1,94	,239
CSS validity grade	100	1,75	,435
Multiplier	100	3,40	,953
Valid Cases	100		

(Table 5) – Trial Two – International websites

	N	MEAN	STD. DEVIATION
HTML validity grade	100	1,98	,141
CSS validity grade	100	1,57	,498
Multiplier	100	3,12	1,028
Valid Cases	100		

(Table 6) – Trial Two – Greek websites

From the previous tables it is evidenced that the mean of the multiplier is higher in the case of the International websites, compared to the case of the Greek ones. Since the t-test p-value (0,047) is lower than 0,05 (0,047 < 0,05) and further supported by the fact that the zero value is not included within the 95% confidence interval, we accept that the difference of means is statistically significant and is attributed to the performance of the websites of each group. Due to the fact that the range of values for all three variables of the second trial is rather limited -in the case of validity checks the set {1, 2} and in the case of the multiplier the set {1, 2, 4}-, the tables of descriptive frequencies provide a more appropriate tool for further analysis of the results obtained by the second trial. From the descriptive frequencies the following conclusions may be drawn:

The ideal status, under which a given website is proven valid against both checks, is limited to only 2% of the total sample, with no significant variation between the two groups of websites.

Accordingly, the percentage of websites which fail on both checks -i.e. each validity grade is equal to 2 and the value of the multiplier is equal to 4 (2*2=4)- is disappointingly high: 64% as far as the total sample is concerned, 71% in the case of International websites and 57% in the case of Greek websites.

The frequencies of the second trial also reveal that the markup validity check is of increased difficulty compared to the style sheets validity checks. The percentage of websites which are constructed with valid markup language varies

from 2% in the case of the Greek websites to 6% in the case of the International ones, thus accumulatively accounting for the 4% of the total sample. Conversely, the percentage of websites which employ valid style sheets is significantly higher than the percentage of websites with valid markup and varies from 25% in the case of the International websites to 43% in the case of the Greek ones, thus accumulatively accounting for the 34% of the total sample.

The third and last trial of the threefold accessibility evaluation process concerns the subjective evaluation by human experts (analysts). The experts assigned an accumulative mark with respect to all of the aspects of each given category (namely navigation, content, structure, requirements, support). Finally the average of the experts grade in each category was calculated for every website of the sample. The range of values available to the experts was the set {0, 1, 2, 3, 4, 5}, where 0 denoted the absence of accessibility problems and 5 denoted the highest penalty value for severe accessibility inadequacies. The average value of the subjective evaluation is stored in the "experts grade" variable. The following table depicts the minimum value observed, the maximum value observed, the mean and the standard deviation of the "experts grade" variable for the total sample with respect to the third trial.

	N	MINIMUM	MAXIMUM	MEAN	STD. DEVIATION
Experts Grade	200	,00	5,00	2,1700	,97810
Valid Cases	200				

(Table 7) – Trial Three – Total Sample

Respectively, the following two tables depict the results of the third trial for the International and the Greek websites.

	N	MINIMUM	MAXIMUM	MEAN	STD. DEVIATION
Experts Grade	100	,00	4,40	1,8320	,93235
Valid Cases	100				

(Table 8) – Trial Three – International websites

	N	MINIMUM	MAXIMUM	MEAN	STD. DEVIATION
Experts Grade	100	,20	5,00	2,5080	,90717
Valid Cases	100				

(Table 9) – Trial Three –Greek websites

From the previous tables it is evidenced that the mean of the experts grade is lower in the case of the International websites, compared to the case of the Greek ones. Since the t-test p-value (0,000) is lower than 0,05 (0,000 < 0,05) and further supported by the fact that the zero value is not included within the 95% confidence interval, we accept that the difference of means is statistically significant and is attributed to the performance of the websites of each group.

After the completion of the three trials the value of the "final grade" variable was calculated for each website of the sample according to the unification formula. It is worth noting that –as in the case of independent trials- the final grade estimates degrees of penalty and therefore websites which obtain higher final grades are the most problematic in terms of accessibility and vice versa. In particular, the range of values for the final grade is the subset $\{x \mid x \in [0, +\infty)\}$ of the set of real numbers $\mathbb{R}$, given that (a) there is no negative factor within the

unification formula (i.e. the final grade will always be a positive number or zero) and (b) both the experts grade and the error rate may be equal to zero. If the latter case occurs concurrently for both values the whole expression is set to zero which is the ultimate excellent grade assigned by the research tool.

The following table depicts the minimum value observed, the maximum value observed, the mean and the standard deviation of the "final grade" variable for the total sample with respect to all the trials of the threefold web accessibility evaluation tool.

	N	MINIMUM	MAXIMUM	MEAN	STD. DEVIATION
Final Grade	200	12,06	317,49	147,721	55,44267
Valid Cases	200				

(Table 10) – Final Grade – Total Sample

Respectively, the following two tables depict the final results with respect to all three trials for the International and the Greek websites, after the unification process.

	N	MINIMUM	MAXIMUM	MEAN	STD. DEVIATION
Final Grade	100	21,99	255,92	119,423	48,61398
Valid Cases	100				

(Table 11) – Final Grade – International websites

	N	MINIMUM	MAXIMUM	MEAN	STD. DEVIATION
Final Grade	100	12,06	317,49	164,020	53,01886
Valid Cases	100				

(Table 12) – Final Grade – Greek websites

As expected the mean of the final grade is lower in the case of the International websites, compared to the case of the Greek ones. Since the t-test p-value (0,000) is lower than 0,05 (0,000 < 0,05) and further supported by the fact that the zero value is not included within the 95% confidence interval, we accept that the difference of means is statistically significant and is attributed to the performance of the websites of each group.

Conclusions

From the study data, an accessibility ordinal scale was created which reveals that only 4 out of 10 websites respond satisfactorily to the accessibility guidelines standards. Of course, together with this assessment, two factors should be taken into grave consideration: (a) the accessibility ordinal scale was the product of processing the study data, which means that it does not allow assessments that have to do with the absolute degree of compliance with the accessibility guidelines (overestimation factor), and (b) the sample consisted of websites selected according to how much their content interests senior citizens and therefore, by inference, the sample consisted of websites that appear to be more sensitive to accessibility and usability issues [6], even though the prime motivation may be limited to the will for increased popularity (underestimation factor).

The validity percentages that arise from the frequencies of the second check can only be regarded as disappointing. A scant percentage (2%) of the websites appear to meet the basic validity standards –according to the Consortium

patterns– concerning the websites' code [1].Yet, even when the case is the validity of one out of the two basic technologies that were checked, the percentage does not exceed a relatively low 34%. The factors that should also be taken into consideration in relation to the conclusion drawn in this paragraph, are, firstly, the issue of cascading style sheets which present particular difficulties with regard to the automated validity assessment (overestimation factor), and, secondly, the rigidity of the software used for the check, as it is the case of document type declaration, where the mistakes in declaration lead to the denial of the validity of the whole document (underestimation factor). In the case of validity checks, the factors that might increase or decrease the reliability of the conclusion cannot be used in any type of controversy: 2% validity is an alarmingly low percentage...

The error rate, as it was recorded in the first trial, rises up to 0,33 (or 33%) for the whole sample; this percentage is considered particularly high. It is worth noting that the error rate is one of the most reliable variables of the research tool, and that is because it calculates errors-per-tag – that is, it analyses the basic structural elements of the World Wide Web, namely the HTML tags. The fact is that one in three tags appears to have a compliance with the guidelines problem, which shows that the websites indirectly related to senior citizens are not even close to the desired levels of usability and user-friendliness [6, 9, 12].

From the performance in the individual checks and the final grade, it is clear that the Greek websites are worse that the respective International ones as far as the accessibility guidelines implementation goes. The difference in the average final grades between the two categories is an adequate indication for this conclusion, as the average penalty grade for the Greek websites rises up to 164,02 whereas the respective average for the International websites is 119,42.

From the individual checks, it is evident that the one that is responsible for the poor performance of the Greek websites is the automated code evaluation, where the error rate goes up to 40%, whereas the respective measurement for the International websites goes up to 27%. If one would like to interpret this, in combination with the fact that in the validity checks considerable divergences are not observed, she could surmise that the high error rate could be attributed to lack of information of Greek designers concerning the Consortium's initiative. In a similar study in Great Britain, low accessibility was attributed to restricted knowledge, the web designers' indifference and obstacles of financial nature [3].

Epigrammatically, the main conclusions of the accessibility study are that the extent of the guidelines implementation is far from the desired level, and that the Greek senior-oriented websites are less accessible than the respective international ones.

Bibliography

[1] Berners-Lee, T. & M. Fischetti, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*, New York: HarperCollins, 1999.

[2] European Commission, *eInclusion & eAccessibility - Design for all*, http://europa.eu.int/information_society/policy/accessibility/deploy/dfa/index_en.htm, Accessed: June 2006.

[3] Disability Rights Commission, *The Web Access and Inclusion for Disabled People - A formal investigation conducted by the Disability Rights Commission*, London: TSO, 2004.

[4] France Télécom Recherche & Développement, *OCAWA: Accessibility Validator*, http://www.ocawa.com/, Accessed: June 2006.

[5] Giannakoulopoulos, A., “Threefold Web Accessibility Evaluation by the Use of an Integrated Tool”, in *Proceedings of the 2nd International Conference on Social and Organizational Informatics and Cybernetics*, July 20-23, 2006 - Orlando, Florida, USA (to appear).
[6] Kirkpatrick, C., “Getting two for the price of one: accessibility and usability”, *Computers in Libraries*, v23 i1, January 2003, pp. 27–31.
[7] Lang, T., “Comparing website accessibility evaluation methods and learnings from usability evaluation methods”, in *Peak Usability*, http://www.peakusability.com.au/articles/, Accessed: June 2006.
[8] Meimaris, M., “Greece Going Digital. An Overview of the Digital Media in Greece” in *The Integrated Media Machine, Aspects of Internet Culture, Hypertechnologies and Informal Learning*, Rovaniemi: Edita Helsinki, University of Lapland, 2001.
[9] Mueller, J-P., *Accessibility for everybody: understanding the Section 508 accessibility requirements*, Berkeley: Apress, 2003.
[10] Netscape Communications, *Open Directory Project*, http://www.dmoz.org/, Accessed: June 2006.
[11] Nielsen, J. & M. Tahir, *Homepage Usability: 50 Websites Deconstructed*, Indianapolis: New Riders Publishing, 2002.
[12] Slatin, J. & S. Rush, *Maximum accessibility*, Boston: Addison–Wesley, 2003.
[13] World Wide Web Consortium, *Core Techniques for Web Content Accessibility Guidelines 1.0*, http://www.w3.org/TR/WCAG10-CORE-TECHS/, Accessed: June 2006.
[14] World Wide Web Consortium, *Web Accessibility Initiative*, http://www.w3.org/WAI/, Accessed: June 2006.
[15] World Wide Web Consortium, *Web Content Accessibility Guidelines 1.0*, http://www.w3.org/TR/WCAG10/, Accessed: June 2006.
[16] Zaphiris, P. & G. Zacharia, “Website Content Accessibility of the Cyprus Domain” in Manolopoulos, Y., S. Evripidou & A. Kakas (eds.), *Lecture Notes in Computer Science: Advances in Informatics*, Heidelberg: Springer Verlag, 2003, pp. 248-261.
[17] Zeldman, J., *Designing with Web Standards*, Indianapolis: New Riders Publishing, 2003.