

Semiautomatic Evaluation of Websites Usability

Eva García, Antonio García, Luis de-Marcos, Salvador Otón, José-Ramón Hilera

Computer Science Department

University of Alcalá

Alcalá de Henares, Spain

eva.garcial@uah.es, a.garciac@uah.es, luis.demarcos@uah.es, salvador.oton@uah.es, jose.hilera@uah.es

Abstract - This paper presents a semiautomatic evaluation method of usability in websites. Expert review is done in this case using a template based on the ISO 9241-151 guidelines that provides numerical and graphical results. Such results permit to observe the aspects of usability that are missed on the evaluated website. This template has many applications as it can be used both to evaluate websites and to teach heuristic evaluation method for students who are studying subjects of usability.

Keywords - Usability evaluation; usability guidelines; heuristic evaluation; web usability

I. INTRODUCTION

Usability is one of the most important features to consider when making systems that have large audiences, which need to operate in an intuitive system, without prior training or direct support [1]. In addition, usability is the second most important problem (just after security) for the acceptance of systems by users [2]; so, it must not be omitted when developing computer systems. However, currently there are few tools to help automating the process of usability evaluation.

The aim of this paper is to show a new way of semiautomatic heuristic evaluation [3] of websites. For this, we have developed a spreadsheet template with guidelines to evaluate, which are organized into groups, and each of them is given a value based on a certain scale. When the scores of all guidelines are filled, a result will be obtained showing the more problematic aspects in relation to the usability of the evaluated website.

In this case, the heuristics are based on guidelines provided by ISO 9241-151 [4]. This standard is specifically a part of the ISO 9241 family of standards, which began publication in 1997 under the title "Ergonomics of human-system interaction" and provides requirements and recommendations related to the attributes of the hardware, software and environment that contribute to usability and ergonomics underlying principles.

Part 151 in particular, is titled "Guidance on World Wide Web user interfaces" and provides guidance for user-centered design of web interfaces in order to increase usability. Web user interfaces are aimed to all Internet users and to closed user groups such as members of an

organization, customers or suppliers of a company or other specific user communities.

User interfaces of different types of user agents such as web browsers or additional tools such as web authoring tools are not directly aimed in this part of ISO 9241, although some of the guidelines could be applied to these systems as well.

Web user interfaces can be presented in a personal computer, in a mobile system or in any other device connected to the network. While the recommendations contained in this part of ISO 9241 are applied to a wide range of technologies available in front-end, web interfaces design of mobile devices or smart devices may require additional guidance.

A new way of semiautomatic heuristic evaluation of websites is proposed here. To do this, Section II describes the design of the template, i.e., groups that are designed to bring together the guidelines, the scale that has been used to evaluate each guideline, the content of the template and how can results be interpreted. In Section III, template is applied to several websites and obtained results are presented in each case, as well as a comparison of them. Finally, some conclusions and future research are presented.

II. TEMPLATE DESIGN

To evaluate in a semiautomatic way the usability of a website, a spreadsheet template has been created from an existing one created by the company Userfocus, available at [5]. The original template consists of 247 guidelines created by the company, grouped in nine categories: home page usability, task orientation, navigation and IA, forms and data entry, trust and credibility, writing and content quality, page layout and visual design, search usability, and help, feedback and error tolerance.

The new template (Figure 1) proposes a significant improvement over the previous one, including new guidelines from ISO 9241-151 standard. These guidelines have been put together in new groups, also different from the original template, and also taking into account the sections of ISO 9241-151 standard.

In addition, the rating scale for each guideline has also been modified to take more realistic the scoring system.

A. *Designed groups*

As mentioned earlier, the guidelines of the template are based on ISO 9241-151. This standard consists of the following sections:

1. Scope
2. Normative references
3. Terms and definitions
4. Application
5. A reference model for human-centered design of World Wide Web user interfaces
6. High-level design decisions and design strategy
7. Content design
8. Navigation and search
9. Content presentation
10. General design aspects

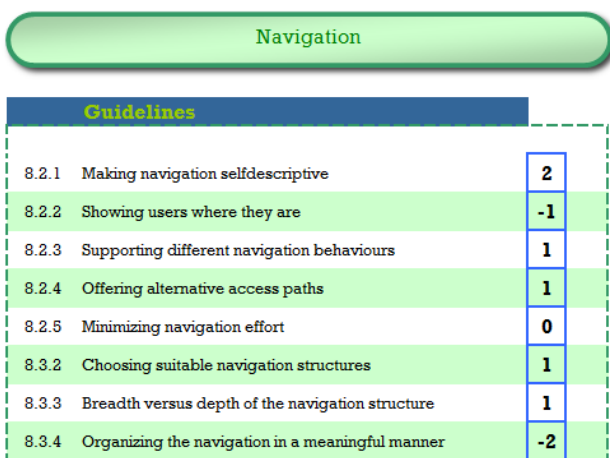


Figure 1. Overview of the template showing a category with scored guidelines.

Only Sections 7 to 10 have been taken into account for the template, because they are the only evaluable guidelines for a website already developed. This is because Sections 1 to 5 contain general information about the ISO and Section 6 contains high level decisions and design strategies, so they will only be applicable to websites that are still in development but not for those that have already been developed.

Guidelines for evaluating usability have been divided into several groups according to the different sections of the ISO 9241-151 listed above, but with an exception: Section 8 has been divided into two parts, as in the ISO the aspects of navigation and search are put together in the same section, and it is preferable to evaluate separately the navigation and the search of the website.

Therefore, sections that the template finally contains are as follows:

1. Content design: it is about everything related to the conceptual model of content, content objects and functionality.
2. Navigation: it deals with issues related to the structure and components of the navigation.
3. Search: it is about the aspects related to search terms on a web page.

4. Content presentation: it treats the aspects related to the design of a web page, links design, interaction objects and text design.

5. General design aspects: the guidelines of this group aim to be designed for a culturally diverse and multilingual use, and provide help and make web user interfaces error tolerant. In addition, it cares about the names of the URLs, the time of downloading, designing with independence of input devices, etc.

Each one of these groups will be a category in the spreadsheet template, and each category includes the guidelines for the subject matter that group.

B. *Rating scale*

Original template had a scale of integer values between -1 and +1, i.e., the possible values that could be given to each guideline were:

- 1: Guideline is not satisfied.
- 0: Guideline is not totally satisfied (it is partially satisfied).
- +1: Guideline is satisfied.

If guideline is not satisfied, in the original template the value should be left blank. This way is not very intuitive, and it does not allow intermediate values when a guideline is almost completely satisfied or it hardly satisfies.

To improve this system of evaluation of the guidelines, our template uses the Likert scale [6]. This is reasonable since guidelines consist of a title (which is what appears in the template and corresponds to a section of the ISO), but also an explanation (available within the ISO), which sets out statements that should be satisfied.

To do this, it is necessary that the scale has an odd number of levels. It should be noticed that the guideline may not be applicable. Due to this reason, a 5-level scale (with 5 categorical variables) has been chosen, i.e., the possible values that can be given to each guideline are:

- 2: Strongly dissatisfied.
- 1: Dissatisfied.
- 0: The guideline is not applicable.
- +1: Satisfied.
- +2: Strongly satisfied.

C. *Template content*

The template consists of 7 categories or sections, of which the last five correspond to designed groups that have been explained in Section 2.1. The categories that make up the template are:

- Instructions: it contains instructions for using the spreadsheet workbook and to properly fill the template. To do this, the values that can be assigned to each guideline are explained, i.e., the rating scale discussed in Section 2.2.
- Results: after filling in all the scores of the guidelines, it will be necessary to see this category for interpreting the obtained results. This category consists of a table summarizing the results for each set of guidelines, as well as a chart where results of the previous table can be graphically seen.

- Content design: it is the category corresponding with the design group called “Content design”.
- Navigation: it is the category corresponding with the design group called “Navigation”.
- Search: it is the category corresponding with the designed group called “Search”.
- Presentation: it is the category corresponding with the designed group called “Content presentation”.
- General design: it is the category corresponding with the designed group called “General design aspects”.

D. Interpretation of results

In the category “Results” of the template, results of the evaluation of the website can be interpreted. This will have a table that identifies, for each of the designed groups, the following information:

- Raw score: it is the addition of the scores of each of the guidelines of the group in question.
- #Questions: it is the number of guidelines that are evaluated in the group in question. Guidelines that are not rated are not taken into account (but there should not be any, because all values must be filled) nor those that contain a 0 (because in that case guideline is not applicable).
- Score: it is the score in percentage terms, i.e., it gives an idea of the degree of compliance with a given set of guidelines. Its value can vary between -100% (if none of the guidelines of designed group is satisfied) and +100% (if all guidelines of the designed group are completely satisfied). It is calculated as follows:

$$\frac{\sum_{i=1}^n v_i}{2 \cdot n} \cdot 100 \tag{1}$$

where n is the number of guidelines of the group in question, and vi is the rated value to each of the guidelines.

- Overall score: it is the overall score in terms of percentage, i.e., it gives an idea of the degree of general compliance with the guidelines and, therefore, of the ISO 9241-151. It is the average of the scores obtained in the score categories.

In addition to the results table, this section contains a radar chart that helps getting a quick idea of the general usability of the website evaluated:

- The larger the shaded area, best usability. In a general way, a website has a good usability when the score of each section is at least 50%. A 30-50% range means a reasonable usability and a score under the 30% means that the website has a poor usability.
- The higher the scale of the graph, best usability average score.
- The angles in the chart represent each of the designed groups as described in 2.1, so that the greater the score has an angle, better usability will have the corresponding group.

III. REAL APPLICATION

The tool is created for helping the heuristic evaluation, since several evaluators are involved in the heuristic evaluation but the template can only be filled by a single evaluator at a time. The idea is that each assessor completes the template on his own, and then pooling the results to issue a final report. This report is the only document the user will receive, because it is written in an easy to understand vocabulary for the user, so the evaluation sheets will not be shown to the user, because they contain technical vocabulary that he or she could not understand.

To demonstrate the operation of the template as a heuristic evaluation method, we have chosen two websites where it has been applied: a website that has been evaluated negatively and one that has been evaluated positively. To check the capabilities of the template, evaluations have been carried out by an usability expert who has experience in web interfaces (as recommended by Nielsen [7] to identify most of the problems).

A. Example of website evaluated positively

The site chosen for this section has been the London transport website [7]. After conducting the heuristic evaluation with the template, results obtained can be seen in Figure 2.

Summary of Results			
	Raw score	# Questions	Score
Content design	23	15	77%
Navigation	46	26	88%
Search	19	17	56%
Content presentation	42	37	57%
General design aspects	19	13	73%
Overall score		108	70%

Figure 2. Table of results of the heuristic evaluation of the London transport website.

As it can be seen in that figure, the main conclusions can be drawn are:

- The best evaluated group is the navigation group, with 88%, indicating that the website navigation is very good.
- The worst rated group is the search group, with 56%, indicating that there would be desirable to improve the website search system.
- The final average score is 70%, so that we can deduce that the website usability is quite good.

B. Example of website evaluated negatively

The site chosen in this case was that of an Atlanta restaurant guide (<http://www.restaurantguideatlanta.com/>). After conducting a heuristic evaluation with the template, we got the results that can be seen in Figure 3.

As shown in the figure below, the main conclusions obtained are:

- The best evaluated group is the general design aspects group, with 28%, indicating that these aspects can be even improved.

- The worst rated group is the content presentation group, with -41%, indicating that in this respect the website is quite poor.
- The final average score is -12%, so that we can deduce that the usability of the website can (and should) be improved significantly.

Summary of Results			
	Raw score	# Questions	Score
Content design	-3	12	-13%
Navigation	-16	22	-36%
Search	1	19	3%
Content presentation	-26	32	-41%
General design aspects	5	9	28%
Overall score		94	-12%

Figure 3. Table of results of the heuristic evaluation of the website of Atlanta restaurant guide.

C. Comparison of the results of example

To compare the results of the example, it is best to observe the charts of both websites generated by the template (Figure 4) after performing the heuristic evaluation of them.

As shown in the figure below, there are considerable differences in charts:

- The shaded area is higher in the first chart than in the second one, indicating a better general usability of the website.
- The scale of the first chart varies between 0% and 100%, while the second chart scale varies between -50% and 50%. This indicates that in the first chart the average score of usability is greater than in the second chart.
- In the first chart, navigation is highlighting, followed by the content design and general design aspects, and it can be seen that the aspects that can be improved are the search and the content presentation.
- In the second chart highlights the low overall score, especially as regards the content presentation.

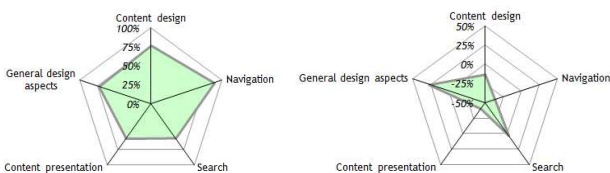


Figure 4. Graphic results of the usability evaluation of the London transport website (left) and Atlanta restaurants guide (right).

Comparing these results with those obtained in Sections III.A and III.B, same conclusions can be drawn when observing both the table and the chart obtained from the template, although it is easier and faster to get a general idea of the usability of a website by looking at the chart.

IV. CONCLUSION AND FUTURE WORK

Once proposed and introduced the modifications in the original template, and having considered some websites to compare the results obtained, we can deduce that this is a

valid method for semiautomatic evaluation of websites usability. We will carry out a comparative experiment with this template and the original, so we will be able to deduce the real benefits of the improvement.

One of the advantages of using this template is that it is fulfilling a very important standard of web usability; and another advantage is that it is a semi-automated method, since the expert introduces values he deems appropriate and the template calculates the percentages of compliance of the guidelines and returns results with which to draw conclusions easily. However, it also has the disadvantages inherent to the heuristic evaluation, among which are the following: evaluators should be experts in usability and, if possible, be familiar with the type of interface that is being evaluated [8], and must have multiple evaluators [9] (which is expensive).

As future research we plan to adapt the guidelines of the ISO 9241-151 to mobile devices, i.e., research what guidelines can be used to heuristically evaluate the usability of websites viewed from mobile devices. Once this is done, the template will be modified and adapted to be used to evaluate the usability of mobile websites according to ISO 9241-151. We will also design a method to collect the results of the evaluation of all experts and achieve a common result.

Finally, we will try to automate the evaluation of some guidelines and analyze some user's behavior patterns, like extracting the link structure of a website and obtaining the link graph according to a predefined ruleset, and then analyzing the behavior of the user while navigating.

ACKNOWLEDGMENT

This research is funded by the FPI research staff education program of the "Junta de Comunidades de Castilla-La Mancha". Authors also want to acknowledge support from the TIFyC research group and the Computer Science Department of University of Alcalá.

REFERENCES

- [1] R. Gafni, "Quality Metrics for PDA-based M-Learning Information Systems", in *Interdisciplinary Journal of E-Learning and Learning Objects*, vol. 5, 2009, pp. 359-378.
- [2] J. Buranatrived and P. Vickers, "An Investigation of the Impact of Mobile Phone and PDA Interfaces on the Usability of Mobile-Commerce Applications", *Proc. IEEE 5th International Workshop on Networked Appliances*, 2002, pp. 90-95.
- [3] J. Nielsen, "Heuristic Evaluation". <http://www.useit.com/papers/heuristic> (last access: 2011/09/06).
- [4] ISO 9241-151:2008. "Ergonomics of human-system interaction. Part 151: Guidance on World Wide Web user interfaces". Organization for Standardization. Geneva, Switzerland, 2008.
- [5] Userfocus, "Expert Review Checkpoints". <http://www.userfocus.co.uk/pdf/ExpertReviewCheckpoints.xls> (last access: 2011/09/06)
- [6] R. Likert, "A Technique for the Measurement of Attitudes", in *Archives of Psychology*, vol. 140, pp. 5-53. New York, 1932 (cited by J.T. Cacioppo and G.G. Berntson, "Relationship Between Attitudes and Evaluative Space: A Critical Review, With Emphasis on the Separability of

- Positive and Negative Substrates”, in Psychological Bulletin, vol. 115, No. 3, pp. 401-423, 1994).
- [7] “Transport for London”, <http://www.tfl.gov.uk> (last access: 2011/09/06)
- [8] J. Nielsen, “Finding usability problems through heuristic evaluation”, Proc. SIGCHI conference on human factors in computing systems, pp. 373-380, Monterey, CA, USA, 1992.
- [9] J. Nielsen and R. Molich, “Heuristic evaluation of user interfaces”, Proc. SIGCHI conference on human factors in computing systems: Empowering people, pp. 249-256, Seattle, WA, USA, 1990.