

Usability Evaluation of Web Portals

Andrina Granić, Ivica Mitrović* and Nikola Marangunić

Faculty of Natural Sciences, Mathematics and Kinesiology, University of Split

Nikole Tesle 12, 21000 Split, Croatia

E-mail: {andrina.granic, nikola.marangunic}@pmfst.hr

**Arts Academy, University of Split, Glagoljaska bb, 21000 Split, Croatia*

E-mail: ivica.mitrovic@umas.hr

Abstract. *The paper reports on the experience concerning the design of a "discount evaluation" of broad-reach web portal usability. The methodology employs a range of assessment methods, providing a comprehensive quantitative and qualitative evaluation feedback. It advocates a number of usability test methods along with specialists' inspection. The results indicate that chosen research instruments, measures and methods for usability testing were consistent. Conversely, the results of the guideline inspection did not conform to the ones obtained through the user testing. Although showing respectable potential, it raised some concerns.*

Keywords. usability evaluation, guideline inspection, end user testing, broad-reach portals

1. Introduction

An improvement of *usability* as a quality of use in context [1] is nowadays perhaps the most important goal of current research in the field of human-computer interaction (HCI). As usability becomes defined within a relationship between task, user and system purpose, there is no simple definition or meaningful single measure of usability. Most used assessment methods are grouped into two categories e.g. [11, 13]: (i) *usability test methods*; user-based which involve end users, hence including user testing, focus groups, interviews, questionnaires and surveys as well as (ii) *usability inspection methods*, without end users embracing heuristic evaluations and cognitive walkthroughs as frequently used ones. Recent research has had a tendency to bring together those two basic approaches, cf. [7].

When considering usability of a *web portal*, a site that functions as a point of access to information on the Web [20], it should be noted that current research on usability evaluation is

mostly connected with focused portals such as enterprise or corporate portals [2], travel portals [5], library web portals [3], healthcare web portals [16] and similar. This is the result of the global trend of portal specialization. In such context it is difficult to find any studies related to broad-reach web portal assessments.

This paper reports on our experience with the design of a "discount" methodology for usability evaluation of *broad-reach web portals*. The main motivation for our research came from reports stating that the most visited Croatian web sites are broad-reach web portals, thus implying that they would be familiar to end users and designers. We conducted the experiment in order to evaluate how easily used and efficient those portals are. The complementing usability test methods proved to be consistent. Conversely, results of the designed guideline-based evaluation were not in agreement with the ones obtained from the usability testing, raising some concerns which will be addressed in our future work. The results go in hand with the assertion that we should not rely on isolated evaluations. Instead, usability assessment methods should be combined, obtaining different kinds of usability improvement suggestions.

The rest of the paper is structured as follows. Section 2 presents the experimental methodology for web portal usability evaluation. Section 3 brings achieved results, while analysis and interpretation of findings is offered in the successive. Section 5 concludes the paper.

2. The experiment

We conducted a controlled experiment which advocates scenario-guided user evaluations involving a number of usability testing methods cf. [15] used to collect both quantitative data and qualitative "remarks" (refer to [6] as well). In

addition, user-based testing is supplemented with less strict heuristic evaluation [14] i.e. guideline inspection. This follows the literature that suggests combination of usability inspection and usability test methods, e.g. [7, 19].

In the following sections we describe the experimental methodology adopted to assess the effectiveness, efficiency and satisfaction as well as expert evaluation of selected Croatian broad-reach portals and the results obtained.

2.1. Participants and portals

The study involved thirty participants with basic computer literacy. According to their practical experience in web design, they were classified in two different groups composed of fifteen participants. The "practitioner" group was composed of three independent sub-groups of randomly chosen participants including computer science experts, marketing experts familiar with Internet issues and students of web design. The "non-practitioner" group was consisted of three independent sub-groups of randomly chosen young, middle aged and elderly participants.

Moreover, a particular group of ten "instant experts" [22] for guideline-based evaluation was formed. These were web design practitioners who learnt principles of good user-centred designs in addition to the evaluation approaches and provided usability expert assessment of the selected portals.

We included four *broad-reach portals* in our experiment, the most visited and also the first-established: Index portal (www.index.hr), Net portal (www.net.hr), Vip portal (www.vip.hr) and T-Portal portal (www.tportal.hr).

2.2. Instruments and measures

User assessment was conducted individually, with Internet access and a screen capturing software for tracing and recording users' actions and navigation. We measured task time and achievement. End user testing was based on criteria expressed in terms of few measures [10]:

- objective performance measurement of *effectiveness* (percent task completion),
- objective performance measurement of *efficiency* (time on task) and
- users' *subjective assessment*.

The System Usability Scale (SUS), a simple standard, ten-item *attitude questionnaire* with five-point Likert scale [4], was used for the

subjective valuation. As an additional subjective feedback, answers to the *semi-structured interview* were collected.

In order to perform a less formal heuristic evaluation, mainly guideline-based inspection, an *evaluation form* consisting of a set of adapted principles augmented with portal-related auxiliary guidelines was prepared. Individual expert's marks and comments were collected. The score for every portal was calculated as an average mark on a seven-point Likert scale.

2.3. Experiment procedure

In order to understand the effect of web portal design in a sample work situation, we elaborated a work scenario, a sequence of typical tasks and user actions. To test assigned tasks and time interval, clarity and unambiguousness of measuring instruments for subjective assessment and adequacy of hardware and software support, *pilot testing* was performed.

We chose three typical tasks whose structure and location on the portals was not changed over time. The tasks covered different topics, offering to diverse groups of involved participants a similar opportunity for finding task-related information. For each selected portal, undertaken tasks were the same and the probability of their completion was similar.

The evaluation procedure was carried out independently with each test user, using a personal computer with Internet access in addition to a software and hardware support for recording results of task completion. Within each evaluation session all the portals were assessed, with the order of their evaluation randomly selected. The allocated session's average time per every participant was 45 minutes. An evaluation procedure consisted of the following steps:

- task-based end user testing,
- usability satisfaction questionnaire,
- semi-structured interview and
- guideline-based inspection.

Task-based end user testing involved a scenario-guided user assessment with tasks selected to show the portal basic functionality. It enabled us to determine user *efficiency* and *effectiveness* while working with the web portal. A user's objective accomplishment measure, labelled as *fulfilment*, was calculated as an average time spent on all allocated tasks weighted with successfulness of task completion. For each user, the time limit for all assigned tasks was 15 minutes per portal.

A *usability satisfaction questionnaire* enabled the assessment of the users' subjective satisfaction with diverse interaction aspects. We used the SUS questionnaire, as it is argued that it yields the most reliable results across sample sizes [18]. Its questions addressed different aspects of the user's reaction to the portal as a whole, providing an indication of a level of statement agreement on a five-point Likert scale. The feedback from the questionnaire was augmented with the users' answers in a *semi-structured interview*. In this interview we asked the participants to rate and comment on the portal's *visual attractiveness* as well.

End user testing of web portal usability combined with guideline-based evaluation provided a more precise broad-reach portal usability assessment. In order to overcome the problem of not having enough usability experts who could be involved in the portal evaluation process, we had the *guideline inspection* performed by "instant experts", web design practitioners familiar with the HCI principles. A detailed evaluation form with Nielsen's usability heuristics [14], as a set of ten key principles, was adapted to portal context and augmented with a series of auxiliary guidelines, as additional explanations of web portals design *cf.* [12], [21]. In the provided evaluation form experts had to specify a level of their agreement with the principle/guideline on a seven-point Likert scale and to provide comments in order to justify the assigned mark. Besides, remarks and observations concerning the overall assessment procedure were more than welcome.

The guideline-based evaluation coupled with the task-based end user testing, the attitude questionnaire and the interview, provided a sufficient quantitative and qualitative assessment feedback. In the following we present experimental results and findings.

3. Results

Results acquired through the usability test methods along with main findings from guideline inspection are addressed in what follows.

3.1. Results of test-usability methods

Descriptive statistics of the objective accomplishment measure *fulfilment*, including arithmetic means, standard deviations and significance levels of Kolmogorov-Smirnov coefficient is presented in Table 1. Result of the

distribution for measure *fulfilment* on the T-Portal portal differs significantly from normal distribution (K-S = 0.008). Accordingly, Friedman's test as a non-parametrical procedure was performed. Statistically significant value of chi square ($\chi^2 = 49.4$, $df = 3$, $p < 0.01$) indicates the existence of differences in the objective accomplishment measure among portals.

Table 1. Results of the *fulfilment* measure (note that lower M score means better result)

fulfilment	M	SD	K-S
Index portal	59.77	38.726	0.292
Net portal	108.40	46.300	0.720
Vip portal	62.13	17.211	0.656
T-Portal portal	171.64	168.143	0.008

Descriptive statistics of results acquired for subjective satisfaction measure *SUS* is shown in Table 2. No statistical difference in the distribution of the results from the expected normal distribution was found ($K-S_{1,2,3,4} > 0.05$). In order to test the difference among portals, the analysis of variance as a parametric procedure was applied. Significant F-ratio ($F = 746.94$, $df = 29$, $p < 0.01$) indicates the existence of differences among the portals in the results related to the obtained subjective measure.

Table 2. Results of the *SUS* measure (note that higher M score means better result)

SUS	M	SD	K-S
Index portal	75.33	18.820	0.819
Net portal	56.00	25.194	0.902
Vip portal	77.83	15.821	0.319
T-Portal portal	51.75	23.836	0.961

We also considered all accomplished experimental results related to the two groups of participants – the practitioner group and the non-practitioner one. The differences in the user's objective accomplishment and subjective satisfaction usability measures between the two groups were tested with t-tests for small independent samples. Statistically significant difference between the groups was found for results of the *fulfilment* measure ($t = 2.95$, $p < 0.01$). The group of practitioners showed better results on mean values (mean = 308.4, $SD = 57.217$) than the non-practitioners group (mean = 495.46, $SD = 238.479$).

On the contrary, a statistically significant difference was not found for the results related to the subjective satisfaction measure *SUS* ($t = 1.95$, $p = 0.062$) between the practitioner group of participants (mean = 243.17, SD = 51.317) and the non-practitioner one (mean = 278.67, SD = 48.531).

Pearson's correlation coefficients for the participants' results in the achieved usability objective and subjective measures showed no significant correlation between overall *SUS* and overall *fulfilment* ($r = 0.14$). However, significant correlation was found between overall *SUS* and score for overall *visual attractiveness* ($r = 0.41$).

3.2. Results of usability inspection method

Obtained experts' marks and comments from the evaluation forms are presented in Table 3. Arithmetic means of marks showed the following results. Highest mark was given to Vip portal (mean = 5.38), followed by Net portal (mean = 4.85), T-Portal portal (mean = 4.64) and Index portal (mean = 4.01).

Overall results could be further related to experts' comments obtained by means of the form. For Vip portal experts emphasized well-adjusted and consistent layout, simple navigation and feeling of control. Portal scored the best for

guidelines 8 and 2. Follows Net portal with highest marks obtained for guidelines 2 and 4, while lowest for 10. It is described as consistent with good quality of information structure, but with poor and old fashion visual appearance. T-Portal complies very well with guideline 5, but conversely doesn't comply at all with guideline 8. Identified problems are related to employed diverse types of navigation and too extensive initial page. Lack of consistency and aggressive "visual noise" were the main reasons why Index portal got severely bad marks. The worst marks were obtained for guideline 4. Identified problems include ambiguous home page, lack of consistency and navigation overload.

The evaluation form analysis included the assessment of adapted guidelines themselves and judgment of the quality of experts' evaluation. Qualitative analysis criteria were expressed in terms of mark span and value of comments (see columns *Mark Span* and *Info* in Table 3).

The guidelines were "horizontally" examined through expert's comments and observations, assigning low (L), medium (M) and high (H) values according to the quantity and the level of details of provided comments (*Info* column). The range of marks expresses lowest and highest marks given by the experts (*Mark span* column).

Table 3. Guideline analysis

Portals	Index portal		Vip portal		Net portal		T-portal portal	
	Mark span	Info	Mark span	Info	Mark span	Info	Mark span	Info
Guidelines								
1. Portal is actively informing user about its' processes (information about what is going on is always present).	4 – 6	L	2 – 6	L	3 – 6	L	2 – 6	L
2. Concept of portal is well adjusted to the user context.	5 – 6	M	5 – 7	M	5 – 6	L	2 – 6	M
3. While working with portal users have feeling of control, safety and navigation freedom.	3 – 6	H	4 – 7	M	2 – 6	H	2 – 6	H
4. Portal respects media standards and usual practice/usage/routine.	1 – 4	H	5 – 7	M	4 – 6	M	2 – 7	M
5. Portal prevents possible user errors.	2 – 6	L	3 – 7	L	3 – 7	L	2 – 7	L
6. User is intuitively getting information on portal, i.e. user doesn't have to remember information path but recognize it.	4 – 5	M	4 – 7	L	3 – 6	M	2 – 6	L
7. Portal is adjusted for efficient use by novice users as well as by experts.	3 – 5	H	4 – 6	M	3 – 6	H	2 – 6	M
8. Portals' design is clear, understandable and transparent, i.e. most needed information are at the same time most visible.	2 – 5	M	5 – 7	M	4 – 6	M	2 – 6	L
9. Portal enables user recognize and help recover from errors.	2 – 6	L	3 – 7	L	2 – 7	L	3 – 7	L
10. Portal offers help while working on it.	2 – 5	M	3 – 6	M	3 – 6	M	2 – 7	M

"Vertical" analysis comprised inspection of specialist's answers to the guideline compliance related to the four assessed portals. The same information quality criteria were used while analyzing experts' work: number, percentage and quality of provided comments as well as number of additional observations.

4. Interpretation of the results

The results of the task-based end user testing showed statistically significant differences among the assessed portals according to the measure of user's objective accomplishment. This suggests that portals could be ranked by mean values. The results of the subjective satisfaction measure also showed differences among portals and their ranking by mean values. The measures of user's objective accomplishment and her/his subjective satisfaction were not significantly correlated. This is in accordance with the results of the meta-analytic research report on correlations among usability measures calculated from the raw data of 73 studies [8].

The overall achieved results could be further related to the most frequent statements from the interviews. Participants felt especially pleased and comfortable working with the portals where their objective achievement was high. They considered them as broad-reach web sites with good quality of information structure, respectable layout and straightforward navigation.

Correlation between overall SUS results and overall visual attractiveness indicates that a pleasant appearance influences the subjective perception of portal usability. The interview statements also support this finding. Such assumption is in line with related studies which, along with the related HCI issues, also address aesthetics aspects of design, *cf.* [17].

The results of the objective accomplishment measure revealed expected differences between non-practitioners and practitioners, the latter being faster and more successful in the tasks' achievement. This indicates that the selected tasks and the objective accomplishment measure were consistent. Conversely, the measure of subjective satisfaction did not show any statistically significant difference between these two groups. Such finding indicates that the questionnaire itself and its translation to Croatian language could be considered as an appropriate instrument for user subjective assessment. Moreover, our experience suggests that the choice of the sample size in addition to the

structure of engaged end users is also in line with the outcomes of related studies, *cf.* [8].

Comprehensive analysis of data obtained through ten evaluation forms was performed. The achieved result of usability inspection did not conform to the ones obtained throughout applied usability testing. The highest ranked web portal in the end user testing scored as the lowest one in the "instant expert" usability evaluation. There are two possible reasons for such an outcome – the designed evaluation form and/or the selection of usability specialists involved in the web portal guideline-based assessment.

Concerning the first reason, some of adapted Nielsen's principles showed poor applicability in the web portal context. Namely, the information gained through individual specialist comments differs a lot. Remarks like "there is no mistake" or "not good at all" represent comments of low information quality. Conversely, detailed ones which list specific observations related to page layout, fonts, navigation and graphics, are classified in medium or high quality category. It is interesting that guidelines which stir poor information have wide mark span, which could imply both limitations in their understanding and/or their vague formulation. Consequently, a number of guidelines should be more comprehensible and auxiliary guidelines revised and redundant ones excluded. A new set of guidelines is needed, the one which is not so strictly based on Nielsen's heuristics.

Regarding the choice of usability "instant experts", significant difference in acquired information suggests non-homogeneity of the group concerning their HCI knowledge and usability expertise. Such problem was hard to prevent due to inadequate number of resident HCI specialists as well as the high costs of possible foreign experts' engagement.

5. Conclusion and future steps

The objective of the research is the design of "discount evaluation approach" to web portal assessment. Aiming to achieve the goal, design of most visited Croatian broad-reach web portals was assessed both through a number of usability test methods and the usability inspection method. The designed evaluation methodology, as an approach which advocates specialists' assessment along with scenario-guided user evaluations used to collect both quantitative data and qualitative remarks, provided comprehensive valuation feedback. The experience reported on in this

paper indicates that the chosen research instruments, measures and methods for user-based evaluation were consistent. Conversely, the results of the designed inspection method, as a less formal heuristic evaluation, were not in agreement with the ones obtained from the user test methods. Consequently, the guideline-based evaluation, even though showing respectable potential, raised a couple of concerns. This is in line with recent research on heuristic evaluation which is focused on improving its effectiveness and efficiency with respect to user testing, cf. [9]. Future work will consider the following issues:

- first, in order to upgrade the applied usability inspection method (i) the instant experts selection and the evaluation form issues should be revised according to the quantitative and qualitative analysis of the obtained results and (ii) the redesigned methods should be applied for the assessment of more specialized Croatian web portals;
- second, in order to improve the applicability of the methodology to practice and to achieve its broad generalization, an inclusion of a cross-cultural sample should be considered.

6. Acknowledgements

This paper describes the results of research being carried out within the project 177-0361994-1998 *Usability and Adaptivity of Interfaces for Intelligent Authoring Shells* funded by MZOŠ of the Republic of Croatia. The experiment was conducted at Dept. of Visual Communication Design, Arts Academy.

7. References

- [1] Bevan, N., Macleod, M., 1994. Usability measurement in context. *Behaviour & Information Technology*, 13, pp. 132-145.
- [2] Boye, J., 2006. Improving portal usability. *CMS Watch*. <http://www.cmswatch.com/Feature/151-Portal-Usability>.
- [3] Brantley, S., Armstrong, A., Lewis, K.M., 2006. Usability Testing of a Customizable Library Web Portal. *College & Research Libraries* 67, 2.
- [4] Brooke, J., 1996. SUS: a "quick and dirty" usability scale. In Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, A.L. (Eds.): *Usability Evaluation in Industry*. Taylor and Francis, London.
- [5] Carstens, D.S., Patterson, P., 2005. Usability Study of travel websites. *Journal of Usability Studies*, 1, 1.
- [6] Granić, A., Mitrović, I., Marangunić, N., 2008. Experience with Usability Testing of Web Portals. 4th Int. Conf. on Web Information Systems and Technologies. 4-7 May, 2008, Madeira, Portugal (accepted for presentation)
- [7] Hornbæk, K., 2006. Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Man-Machine Studies* 64, 2.
- [8] Hornbæk, K., Law, E.L.-C., 2007. Meta-Analysis of Correlations Among Usability Measures. *CHI 2007 Proceedings*, April 28–May 3, San Jose, California, USA.
- [9] Hvannberg, E., Law, E.L.-C., Larusdottir, M., 2007. Heuristic evaluation: Comparing ways of finding and reporting usability problems. *Interacting with Computers* 19, 225–240.
- [10] ISO/IEC 25062:2006, 2006. Software engineering – Software product Quality Req. and Evaluation (SQuaRE) – Common Industry Format (CIF) for usability test reports.
- [11] Lewis, J. R., 2005. Introduction to Usability Testing. Tutorial given at HCI International 2005. July 22-27. Las Vegas, Nevada.
- [12] MIT Usability Guidelines, 2004. <http://web.mit.edu/is/usability/selected.guidelines.pdf>.
- [13] Nielsen, J., 1993. *Usability Engineering*. Academic Press Limited. Boston.
- [14] Nielsen, J., 1994. Heuristic Evaluation. In Nielsen, J., Mack, R. (Eds.): *Usability Inspection Methods*. John Wiley and Sons Inc, New York.
- [15] Shackel, B., 1991. Usability – context, framework, design and evaluation. In Shackel, B., Richardson, S. (eds.): *Human Factors for Informatics Usability*, Cambridge University Press, Cambridge, 21-38.
- [16] Theng, Y.L., Soh, E.S., 2005. An Asian Study of Healthcare Web Portals: Implications for Healthcare Digital Libraries. *Proceedings of the 8th international conference on Asian digital libraries*, ICADL. Bangkok.
- [17] Tractinsky, N., Katz, A., Ikar, D., 2000. What is beautiful is usable. *Interacting with Computers* 13, 2.
- [18] Tullis, T.S., Stetson, J.N., 2004. A Comparison of Questionnaires for Assessing Website Usability. *Proceedings of UPA Conference*. Minneapolis. <http://home.comcast.net/~tomtullis/publications/UPA2004TullisStetson.pdf>.
- [19] Uldall-Espersen, T., Frøkjær, E., Hornbæk, K., 2007. Tracing Impact in a Usability Improvement. *Interacting with Computers*.
- [20] Waloszek, G., 2001. Portal Usability – Is There Such A Thing?. *SAP Design Guild*, Edition 3. http://www.sapdesignguild.org/editions/edition3/overview_edition3.asp.
- [21] Wood, J., 2004. Usability Heuristics Explained, *iQ Content*. http://www.iqcontent.com/publications/features/article_32.
- [22] Wright, P., Monk, A., 1991. A cost-effective evaluation method for use by designers. *Int. Journal of Man-Machine Studies* 35, 6, 891-912.