

A usability tool for web evaluation applied to digital library design

Yin Leng Theng, Norliza Mohd-Nasir, Harold Thimbleby
Interaction Design Centre, Middlesex University, UK
{y.theng, norliza1, h.thimbleby}@mdx.ac.uk

Introduction

This paper describes a usability tool implemented to demonstrate that meaningful results could be obtained from small user studies. Digital libraries (DLs) are chosen as the illustrative example of web documents in our study. They are examples of web (or browser-like) interfaces to large digitised and organised collections of information.

Contrast with other tools [e.g., 4, etc.], this prototype tool incorporates questionnaire techniques (an example of quantitative evaluation) and heuristic evaluation (an example of qualitative evaluation) employing real users to understand the usability of web documents. Insights from qualitative evaluations are beneficial in helping one understand reasons why problems occur. Quantitative evaluations help designers to compare and evaluate the effectiveness of systems using robust, quantifiable metrics [1].

The tool contributes in these areas: (i) it is a research tool for analysing, comparing and delivering usability results to designers; (ii) it makes user testing less cumbersome, less time-consuming and more cost-effective; and (iii) it can be used early in the design cycle to influence design, thus closing the design loop without overwhelming designers with unnecessary detail. The inputs of the tool capture users' responses to questionnaire and usability heuristics, as well as attributes such as name and type of systems (see <http://www.cs.mdx.ac.uk/tool/>). Subjects' responses from the questionnaire and heuristic evaluation can be entered into the databases using the web-based questionnaire (see <http://www.cs.mdx.ac.uk/dl/>) and heuristic form (see <http://www.cs.mdx.ac.uk/heuristics/>) provided by the tool. Outputs are the deliverables produced by the tool such as analysis of users' responses compared with other studies, and display of usability problems detected.

The tool applied to digital libraries

We describe the tool in use by exhibiting how two sets of data could be compared to get various usability insights. A demonstration of the tool will be provided at the conference. The first set of data came from a study conducted with 30 subjects. One group of 15 evaluated the

Networked Computer Science Technical Reference Library (NCSTRL, see <http://www.ncstrl.org/>) and another group of 15 evaluated the ACM Digital Library (ACMDL, see <http://www.acm.org/>). The second set of data came from a study that involved 15 subjects evaluating the New Zealand Digital Library (NZDL, see <http://www.nzdl.org/>). All the three sample DLs (studied in the two experiments) were *similar*, in that they contained computing technical materials. Please refer to <http://www.cs.mdx.ac.uk/dl/> for the questionnaire categories. It is debatable but we made the assumption that if an area scores 75% and above for ratings in the "5 and above" category on a 7-point scale, it implies that, the area is well-implemented in the DL in question. The subjects in the studies were also asked to evaluate the DLs using established usability heuristics [1]. Subjects' responses from the questionnaire and heuristic evaluation (<http://www.cs.mdx.ac.uk/heuristics/>) were entered into the databases using the web-based questionnaire and heuristic form provided by the prototype usability tool.

There are many ways to analyse and display usability results. A sample usability report of NZDL generated by the prototype tool for a sample size of 3, randomly selected from the 15 subjects who evaluated NZDL, can be seen at <http://www.cs.mdx.ac.uk/tool/results.html>. The report for NZDL shows for each design category (G1-G9) how well NZDL (indicated by "+ve" and "-ve") performed when compared with, a benchmark score of 75%, and an average score obtained from the first set of data on the usability of ACMDL and NCSTRL. Row R5 shows all nine design categories were not well implemented in NZDL, indicated with "-ve" symbol when compared with a benchmark score of 75%. Except for G2, G8 and G9, subjects rated NZDL less favourably when compared with ACMDL and NCSTRL, as indicated by "-ve" symbol in row R4. As n increases (n=5, n=7), the "+ve and -ve" symbols indicating NZDL's usability in design categories G2 - G9 when compared with ACMDL and NCSTRL (see R4, G2-G9), seemed to converge to a certain "pattern", illustrating that small studies could also give similar impressions of usability compared to larger studies.

A list of usability problems, detected by the subjects completing the heuristics evaluation, which were compiled to correspond with the design categories, can be displayed by clicking onto P1-P6. For example, a list of usability problems for design category G4 could be obtained by clicking P3. Information about usability problems identified could be meaningfully used by designers to carry out, for example, severity ratings of the problems with the same group of test users [3]. Designers could strategically allocate the most resources to fix the most serious problems, or to obtain a rough estimate of the need for additional usability design or evaluation investment in terms of cost and time.

Conclusions and future work

This preliminary work. In this paper, we proposed that useful insights could be obtained from results of small-scale user studies by comparing them against results obtained from larger, representative user studies of *similar* systems. Some critics may no doubt say that this way of comparison is less than perfect methodology. The realistic alternative is to do something as opposed to do nothing — so that some kind of usability results can be obtained with minimal cost and time [2]. Further work will validate the tool and the approach it represents with further different types of DLs and numbers of subjects, as well as strengthening the analysis,

display and reliability of the usability results.

References

1. Dix, A., Finlay, J., Abowd, G. and Beale, R. (1995), Human-Computer Interaction, Prentice-Hall.
2. Nielsen, J. (1997), Discount usability testing for the Web, http://www.useit.com/papers_discount_usability.html.
3. Nielsen, J. (1995), Severity Ratings for Usability Problems, <http://useit.com/papers/heuristic/severityrating.html>.
4. Management, N. and Kirakowski, J. (2000), Web usability questionnaire, <http://www.nomos.se/wammi/>.