Cox, D. and Greenberg, S. (1998) Dealing with Heuristic Evaluation Data. In Proceedings of the UPA '98 Usability Professionals' Association Conference, Poster presentation.

Dealing with Heuristic Evaluation Data

Donald A. Cox and Saul Greenberg
Department of Computer Science, University of Calgary
Calgary, Alberta Canada T2N 1N4
+1-403-220-6087
coxd *or* saul@cpsc.ucalgary.ca

INTRODUCTION

The goal of performing a heuristic evaluation, like any other usability evaluation, is to influence the development of the product [4]. This influence is primarily enacted by both formal and informal conversations based on *problem reports* given to the developers by the evaluators. These problem reports summarize the findings of the heuristic evaluation and its implications.

As a discount usability method, one often seeks maximum efficiency from the heuristic evaluation process. It may be tempting to seek efficiency by giving the raw problem descriptions generated by the evaluators to developers as the final problem reports. This does not work because these raw problem descriptions can not be used effectively by the development team. The detail provided by individual problem descriptions is lacking [1]. Also, the problem descriptions are not organized in a way that is relevant to developers addressing them. Rather than identifying root causes and how they may be addressed, they often identify widely separated or even apparently contradictory symptoms.

The question of how raw problem descriptions are transformed into a coherent final report has not been answered adequately in the literature. In this paper, we consider what happens in this transformation, a process, which we call *results synthesis*. We propose that results synthesis is accomplished by having a group deal with the raw data such that their understanding and interpretation of it emerges naturally. We argue that this activity can be supported in both paper and computer environments that allow the easy expression of complex and dynamic relationships.

THE PROBLEM

Heuristic evaluation [3] may be divided up into four stages.

1. *Preparation.* All the necessary resources are gathered and arrangements made.
2. *Inspection.* Three to five evaluators make their independent assessments of the interface.
3. *Results synthesis*. Disparate views of the evaluators are combined into a single coherent picture of the interface's problems and possible solutions.
4. *Communication*. Results are documented in a final report and communicated to the developers.

Except for results synthesis, each of these stages has been addressed extensively. Nielsen [3], for example, details the qualifications of evaluators and how many should be used; the sorts of interface that can be evaluated; and how the inspections should be carried out. Jeffries [1] describes what should be in the report given to developers. In contrast, results synthesis is sparsely described. Nielsen merely recommends that it may done by a single person or as a group debriefing [3]. Although we know the desirable properties of the final report [1,4] that is the outcome of results synthesis, no process for obtaining these properties has been described.

RESULTS SYNTHESIS AND EMERGENT STRUCTURES

The inputs to results synthesis are the raw problem descriptions. These are often very brief, to the point of being indecipherable to all but the original evaluator [1]. The output from results synthesis is a series of detailed problem reports. Problems and their solutions are justified in detail and expressed at the right level of abstraction; severity is assessed; and tradeoffs explained and weighed. Given the poverty of the raw problem descriptions and the richness of the final problem reports, there is quite a bit going on within results synthesis. The relationships between problems must be examined so that their root causes and best remedies can be identified. This can only be achieved by considering the entire collection of problems as a whole.

While there may be many ways to go about creating the final problem reports from the raw problem descriptions, we suggest that results synthesis is best approached through a process supporting emergence [5]. The understanding of what is wrong with the interface and what should be done about it emerges out of an extended consideration of the raw problem descriptions. This understanding is not a simple aggregation or union of the data. Rather, it is the result of making additional connections between individual observations, and of creating new abstractions. This often goes beyond what is in the data; it brings in the experience of the evaluators both in the inspection of the particular interface as well as from their general background as usability specialists.

Researchers have argued that people can best handle emergent structure through a spatial medium. This is because items are laid out in the space, and the developing relationships between items—even ambiguous or partial ones—are expressed via spatial proximity and visual cues [2,5]. People performing results synthesis, for example, can record raw problem reports on PostIt™ notes, stick them onto a large whiteboard used for annotation, and move them about as the process unfolds.

In our studies of results synthesis we observed several overlapping stages. Initially, individual evaluators have only seen the particular problem descriptions they recorded. Consequently, in the first stage participants get acquainted with the entire data set. They spread out the problems onto the workspace. A useful initial organization is to cluster individual problems by heuristic. This provides an overview of the whole problem set. Even when focussing on a particular item, the participants are always aware of its place in the larger context.

As the participants' attention moves between seemingly disparate parts of the collection, they begin to notice connections between their own contributions and those of others. This leads into the second stage, where a single understanding of the whole collection emerges, though its initial stages are characterized by uncertainty and volatility [5]. This emerging understanding can be easily expressed and naturally perceived through the use of spatial proximity [2]: the distance between items expresses an indication of the strength of their relationship. For example, items that are stacked atop one another indicate that they are considered a single element. The raw problem reports are gradually rearranged into new groups that reflect new organizations, such as by interface element, by solution, or by user task.

In the third stage, participants create the final problem reports. The new groupings, problem descriptions and annotations on the workspace form the basis for the final reports. Creating the wording and form of the problem reports are an occasion for further consideration, discussion, and consensus building.

RESULTS SYNTHESIS AS A GROUP ACTIVITY

We advocate a group, rather than an individual, perform results synthesis. We suggest that the group should include at least the original evaluators, end users, and developers. The evaluators who inspected the interface are necessary because the raw problem reports often contain only fragmentary information that is difficult for others to understand [1]. The inspectors are reminded of their original thoughts about the interface as they review the raw problem reports with the others. This creates a natural opportunity for further clarification and elaboration as they create a common understanding. Both end-users and developers contribute valuable information and perspectives to the process as part of the discussions.

Groups that perform results synthesis need to strike a balance between the extra information provided by more group members and the extra effort needed to operate in a larger group. A larger group comprising evaluators, developers and end users are much more likely to be productive if they have been involved in usability activities from the start and are part of the team.

PAPER OR COMPUTER?

A paper-based approach to results synthesis provides a familiar and somewhat flexible medium. However, it has several disadvantages. The raw problem reports may have to be rewritten onto PostIts, which is tedious. Similarly, items and structures generated during the emergent process have to be transcribed into a form suitable for the final report. Another disadvantage is that the group has to be convened at a common location, which could be difficult if its members are located across different sites. Working around these problems takes extra time and effort, which could be enough to dissuade practitioners from using this technique when they are on tight schedules.

We are creating a prototype designed to address these limitations while still providing similar opportunities to a paper-based approach. It uses spatial hypertext mechanisms developed to support emergent phenomena [2]. We allow participants to exploit spatial proximity by presenting elements in a space. Elements are easy to manipulate. Visual cues hinting at the emergent structures are easy to create and modify in a computational environment. The structures that are created become the structure for the final report. To make it easier to get the raw problem reports into the system, we provide a separate program that allows evaluators to record and categorize problems during the evaluation. The output of this program is the input to our prototype. The prototype is also groupware, which means that it can be used in a participatory process, even if all the participants are not present.

CONCLUSION

Results synthesis is the process by which the assessments of independent inspectors are turned into a single picture of the interface's problems and solutions. We view results synthesis as a process characterized by emergence, where the understanding or interpretation of the data and what it means emerges out of extended consideration of the raw problem reports. People can pursue this process through environments that allow the free and easy expression of subtle and uncertain relationships. Traditionally, this can be done by rearranging PostIt notes on a whiteboard. Moving this spatial environment onto a computer system can provide benefits by making the process more efficient while creating new opportunities.

1. Jeffries, R. (1994). Usability Problem Reports: Helping Evaluators Communicate Effectively with Developers. In J. Nielsen and R.L. Mack (Eds.), *Usability Inspection Methods.* New York NY: Wiley.
2. Marshall, C.C., and Shipman, F.M. (1995). Spatial Hypertext: Designing for change. *Communications of the ACM*, 38(8), 88-97.
3. Nielsen, J. (1994). *Heuristic Evaluation.* In J. Nielsen and R.L. Mack (Eds.), *Usability Inspection Methods*. New York NY: Wiley.
4. Sawyer, P., Flanders, A., and Wixon, D. (1996). Making A Difference - The Impact of Inspections. *Proc ACM CHI '96 Conference*. pp. 376-382.
5. Shipman, F.M., and Marshall, C.C. (1994). *Formality Considered Harmful: Experiences, Emerging Themes, and Directions.* Xerox PARC Technical Report ISTL-CSA-94-08-02.