[1]An evaluation of a Wizard approach to Web design

Karl W. Sandberg
Mid Sweden University and Luleå University of Technology, Sweden

Joel Palmius
Mid Sweden University

Yan Pan
Luleå University of Technology

Abstract

It is necessary to understand to what degree does the interface help users accomplish their task(s) and to what degree are the users satisfied with their experience. Cognitive design is a systems approach to designing a human-computer interface that draws from a combination of the behavioral sciences, software engineering, and computer science to design an expandable Web system that is useful and easy to use. Despite the great effort involved in developing user interfaces, many user interfaces are not evaluated for their usability or acceptability to users. In this paper, we report a continuing research in the design and evaluation of Web systems. Our design process which integrates distinct techniques, knowledge, and experience from fields of cognition, human computer interaction, systems engineering, and usability engineering is used to formulate new Web based ventures.

Introduction

In recent years much information has become available in digital form, and

[1] Third International Conference on Engineering Psychology and Cognitive Ergonomics, 25-27 October, Edinburgh, Scotland

new techniques of information organization and retrieval have become both possible and necessary. Now that the Internet and the World Wide Web (WWW) have become indispensable research tools for locating and disseminating information, issues of information organization and retrieval have become of interest to a very broad constituency (Sandberg, Pan, and Palmius, 2000).

Many Web sites have failed to realize their potential (e.g. ease of publishing, relatively low cost, low training needs, link-based navigational model, ease in updating information, platform independence, and the potential for reaching a wide audience) because they disregard users' needs and requirements (Sang, 1996). Evidence of this disregard is obvious in the pervasiveness of the following usability-compromising situations: Slow response time or speed, difficulty finding information, and poor design and implementation.

There are several reasons for the current state of affairs: Many organizations rush to be on the WWW without spending sufficient time to understand the needs of potential users or to analyse or judge the scope of the content. Other reasons are the disregard for, and inadequacy of methodology and guidelines for designing and organizing information on the WWW.

In designing information retrieval engines it is critically important to be sensitive to human factors; how people search, how they make decisions, and how they use the information they retrieve (Pan and Sandberg, 1999).

As a part of a larger project, IT-infrastructure, designed to improve knowledge transfer between Mid Sweden University and Small Enterprises (SME). The project described here was started to develop a web-based interface to facilitate the exchange. After initial planning discussions, it was decided that the larger project’s existing web pages were not satisfactory. Felt was that even users with long web search experience were unsuccessful in finding the desired information. A problem that is a common problem with web pages (Spool, Scanlon, Schroeder, and Snyder, 1999).

It was decided that two new web interfaces should be developed. One would follow the traditional category/hierarchical layout (e.g., Yahoo). The second would limit cognitive demand as much as possible to facilitate inexperienced users being able to locate desired information.

Drawing on earlier experiences with "wizards"(e.g., Install Shield), and recommendations from current web design literature (Spool et al, 1999) a web-based "wizard" was developed (Figure 1). This interface incorporated the traditional wizard design of asking the user questions and proving a manageable amount of possible answers to each question. It was felt that

this approach would best help novice users who would normally suffer from information overload when confronted with a traditional hierarchical design.

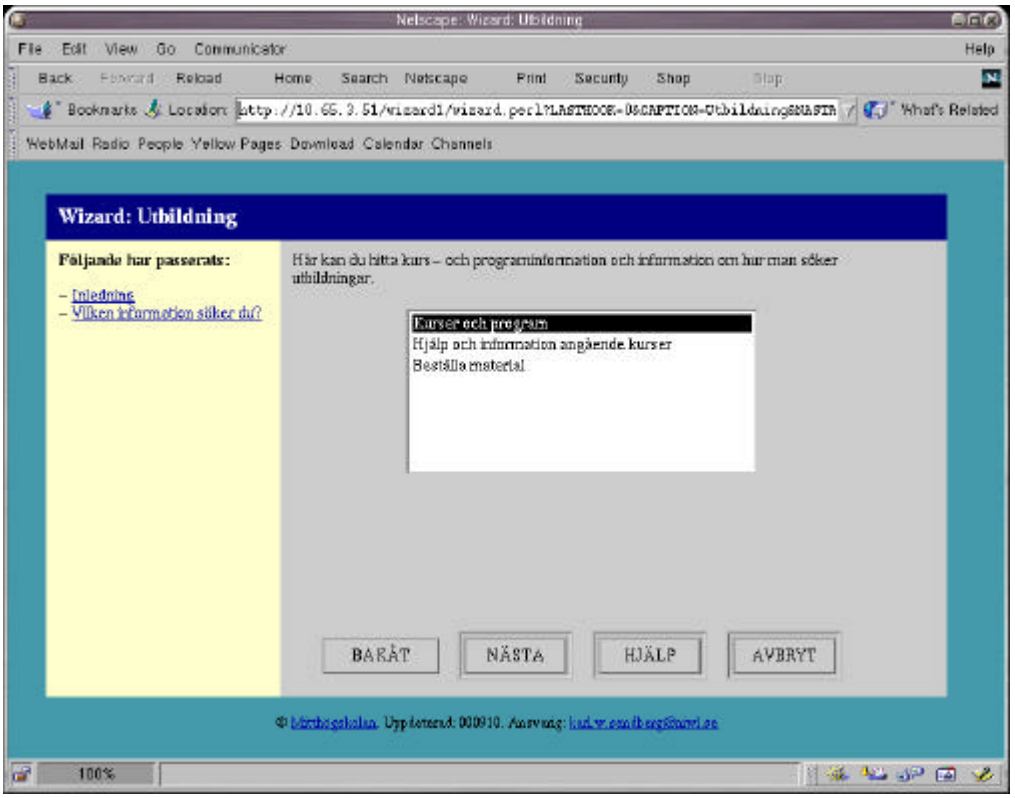

Figure 1. Web-based "wizard".

During the developmental process it was decided that the web interface contact surfaces should not be limited to serving university staff and SME users. Instead, the interfaces was generalized enough to be usable for three target groups: Students, staff and SME users. As part of the assessment design, each of these groups would include users experienced inexperienced in computer use.

To ensure the usability of the developed interface, a usability testing methodology was developed as a part of a master thesis (Alander, 2000) and then used in a bachelor thesis (Ell and Sundberg, 2000). This methodology borrowed most of its key components from literature on usability (Rubin, 1999; Nielsen, 1993) and human factors (Wickens, 1998).

At the time of the writing of this paper, the interfaces have been developed and are being tested. Most of the intended sample from the student target group (both experienced and inexperienced) has been through testing of the interfaces.

This paper describes the preliminary results from that evaluation, with the intention of demonstrating that the "wizard" approach to the design of contact surfaces of this kind scores at least as high on usability evaluations as does the traditional hierarchical approach.

Method

Test plan

From the project plan of the developmental project six key tasks were identified. These included things that were seen as basic for the functioning of the contact surface such as "Find the phone number to the contact administrator Roger Oscarsson who works at…"

The tasks were put through a pilot test to ensure that they were possible to solve within all three interfaces (The two new ones, i.e. the wizard and the hierarchical, and the version currently used at Mid Sweden University). A second goal was to identify the most efficient way the tasks could be accomplished.

Acquiring test subjects

Ten subjects have been tested. As students are readily available, all of the test persons have so far come from the "student" target group.

As each target group was supposed to be divided further into experienced and inexperienced users, students were randomly selected from different parts of the university. The "experienced" users were selected from the Information Systems Science Program (Department of Informatics). The "inexperienced" users were psychology students from the Department of Human Resources, Management and Environment.

It was not possible to find students with no experience in using web interfaces at Mid Sweden University. Therefore, ”inexperienced” students were defined as "users who rate themselves as inexperienced". Test subjects were asked to fill out a questionnaire asking them how often and to what ends they use the web, how they rate themselves in regard to experience of computers, and how they prefer to reach information.

The actual test

The test persons were recorded on a video with two inputs, one showing the screen as they saw it (by having connected the VGA output via a scan

converter to a mixer board) and one from a video camera showing the keyboard and the mouse.

The test subjects were given the tasks one by one, and were asked to solve the task within the given interface. Each test subject tested the same six tasks in all three interfaces, but in order to compensate for learning from the interaction of the previous interfaces, the test order was rotated.

For each task, there was a time limit of five minutes. From our own experience we judged that if the task could not be solved within that time span that the average user would give up. This test strategy was adopted to simulate actual conditions of use. Maximum test time was 90 minutes (questionnaires and instructions excluded); the test subjects were allowed a pause and a soda between each interface to avoid fatigue.

After all the six tasks had been performed within an interface, the test subject was asked to fill out a questionnaire and rate the usability of the product.

The questions on the questionnaire were in the form of statements with five-step Likert responses. The questions fell into three categories. The first category, "Satisfaction", was meant to measure how satisfied the test subject was with the interface. This category included statements like "I liked this product" and "I got angry because I couldn't find the answer". The second category, "Efficiency", was meant to measure how efficient the test subjects thought the product was. This category included statements like "This product is nothing to have if you're in a hurry" and "One has to click too many times to get to the answer". The third category was meant to measure "Learnability". Here the statements included "It was easy to learn how to use this product" and "The structure of the product feels logical".

Apart from the cases where the questions had content validity the questions of both questionnaires were borrowed from literature on usability testing (Rubin, 1994).

Results

All results described here are from the target group as a whole. That group consists of ten test subjects. The results consist of the objective measures of the actual performance, and the subjective measures of how the test subjects rated the products.

Performance

As the tasks had been extracted from the same project plan that formed the basis for the development of the interfaces, it is hardly surprising that most

test subject completed most tasks in the two new interfaces. The Table 1 describes the average number of tasks (out of six) that the test subject failed to complete, per interface.

Table 1. Subject failure of tasks.

	Product								
	Wizard			Hierarchical			Current		
	Mean	Median	Std Deviation	Mean	Median	Std Deviation	Mean	Median	Std Deviation
FAIL	,30	,00	,48	,60	,00	,84	2,30	2,00	,67

A "failure" consists of one of three possible outcomes. The first outcome is that time ran out before the task was completed. The second outcome is that the test subject gave up before finding the correct answer. The third "failure" outcome is that the test subject gave the wrong answer. From the table, it may be seen that fewer failures were recorded with the wizard than with the hierarchical interface. The usual case for both was that all tasks were completed. Table 2 shows average time (in seconds) spent on each task and the average.

Table 2. Task completion time.

	Product								
	Wizard			Hierarchical			Current		
	Mean	Median	Std Deviation	Mean	Median	Std Deviation	Mean	Median	Std Deviation
TASK1	71,60	29,00	93,39	14,30	9,50	12,75	266,00	266,00	,
TASK2	64,30	59,50	33,61	37,20	32,50	24,12	6,20	5,00	4,18
TASK3	26,50	25,50	14,31	50,80	39,50	36,20	234,00	234,00	,
TASK4	39,10	33,00	18,08	52,20	36,00	39,91	57,30	41,50	47,37
TASK5	90,86	78,00	51,15	36,43	35,00	9,03	145,60	126,00	70,09
TASK6	64,90	45,00	61,76	120,29	133,00	60,70	11,00	8,50	5,16
AVGTIM	57,02	45,25	22,61	48,00	51,83	19,00	50,55	42,50	26,71

Here it is apparent that the mean time for solving a task was longer in the wizard than in the hierarchical interface, but that the opposite was true for the median time. The figures for the "Current" interface, tasks one and three, may be ignored as only one subject managed to solve those tasks.

Usability Indexes (subjective)

After each interface, the test subjects were asked to fill out a questionnaire to rate their impression of the product. Table 3 and Figure 2 describe the calculated usability indexes from those questionnaires.

Table 3. Usability indexes.

	PRODUCT								
	Wizard			Hierarchical			Current		
	Mean	Median	Std Deviation	Mean	Median	Std Deviation	Mean	Median	Std Deviation
Satisfaction	4,18	4,38	,86	3,85	4,13	,92	2,05	2,00	,57
Efficiency	4,33	4,50	,80	3,90	4,00	,89	1,73	1,38	,82
Learnability	4,34	4,60	,87	3,88	4,20	,98	2,14	2,20	,52

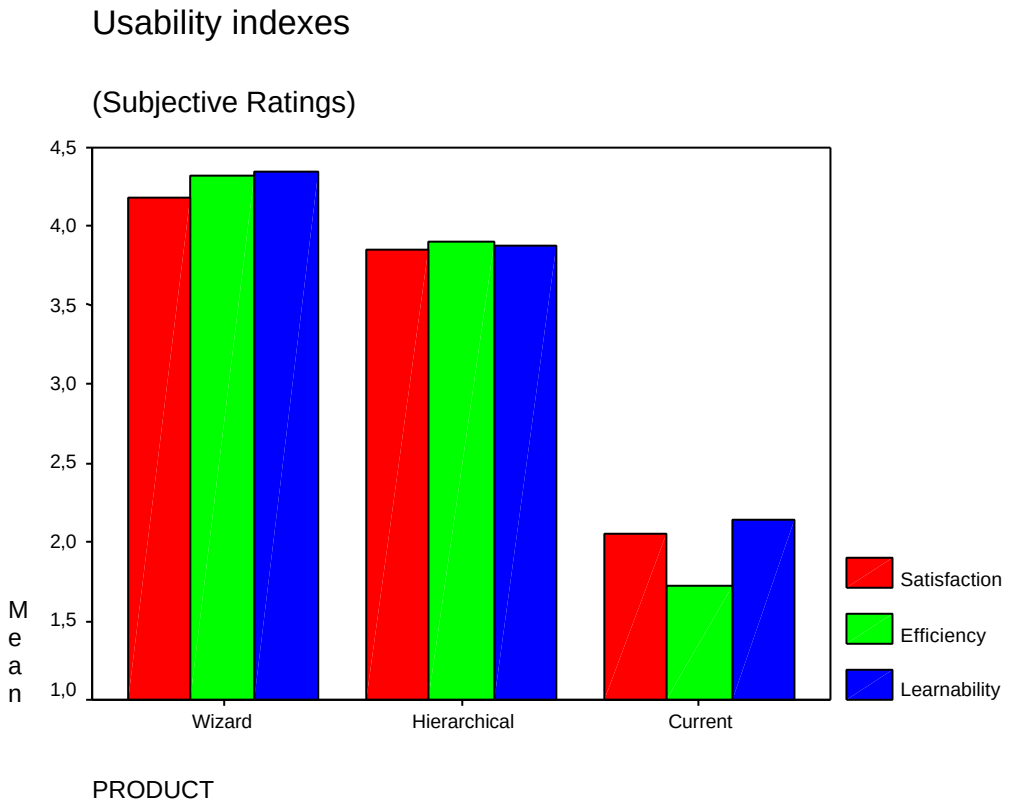

Figure 2. Usability indexes.

Here it's possible to see that the test subjects were generally favorable in their responses for both the wizard and the hierarchical interface (for all indexes the maximum score was 5). They also rated the interfaces similarly.

It should be noted here that the questionnaire did not include questions about aesthetic impressions of the interfaces. However, comments from

female subjects often described the wizard as "ugly" or "boring", while several of the male subjects described it as being "interesting" and "cool".

Discussion

In the above we have described the development of a "wizard" web interface and the evaluation of the same. The results of the evaluation show that the "wizard" approach scores equal or higher in both subjective and objective "usability" when compared with a traditional hierarchical approach to design. There are points that should be noted. The first and most important is that the data presented here is based on a small sample (ten test subjects) and should be considered as preliminary. The second is that we have not as of yet fully analyzed the influences of gender. This is an issue for further research.

References

Alander, R. (2000). *Användbarhetsförluster vid utveckling av ett webbaserat system* (Usability issues in the development of a web-based system). Department of Informatics, Mid Sweden University: Östersund, Sweden.

Ell, M., and Sundberg, A-K. (2000). *Är WAP användbart?* (Is WAP Useable?). Department of Informatics, Mid Sweden University: Östersund, Sweden.

Nielsen, J. (1993). *Usability engineering*. Academic Press: Boston, USA.

Pan. Y., and Sandberg, K. W. (1999). *Usability evaluation of personal security codes based on portraits of human faces*. In D. Harris (ed.) Engineering Psychology and Cognitive Ergonomics, Volume Four, pp 387-393. Ashgate: Aldershot, England.

Rubin, J. (1994). *Handbook of usability testing*. Wiley: New York.

Sandberg, K. W., Pan, Y, and Palmius, J (2000). User model of information retrieval on Internet: In frame of informatics usability engineering. *Behavior research methods, instruments, & computers* (in press).

Sano, D. (1996). *Designing large-scale websites, a visual design methodology*. Wiley: New York.

Spool, J., Scanlon, T., Schroeder, W., and Snyder, C. (1999). *Web site usability - A designer's guide.* Morgan Kaufmann: San Francisco.

Wickens, C., and Gordon, S. (1998). *An introduction to human factors engineering*. Longman: New York.