DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF TORONTO

CSC 428F/2514F: **HUMAN-COMPUTER INTERACTION**

Fall Term, 1997-8

Assignment 1

USEFULNESS AND USABILITY ANALYSIS AND EVALUATION OF EXISTING SYSTEMS

HANDED OUT: Thursday, September 11, 4:00 p.m.
DUE BACK IN: Thursday, October 2, 4:00 p.m.
VALUE: 15 points

One valuable strategy to apply to the analysis and evaluation of an interactive system and its user interface is to attempt a comparative description and evaluation of several similar systems. The purpose of this assignment is to give you some experience in applying this strategy.

THE PROBLEM DOMAIN: INTERNET SEARCH ENGINES

The Internet contains an ever-increasing and amazingly large collection of information. This information exists in millions of Web sites, has been developed with varying degrees of expertise and care, is named and organized in as many different ways as there are authors of the information, and exists in such volume and in such disarray that nobody knows what is there and whether or not the information they desire is available.

One solution that has emerged to help individuals find what they need is that of Internet search engines and Web Guides (meta-indexes). A Web Guide provides a structured hypertext allowing one to navigate through a search space of information collections in order to find the needed information. An Internet search engine allows one to type in a series of text names or descriptions of what one is looking for in order to find the needed information.

There are many Internet search engines and Web Guides, of which Altavista and Yahoo are perhaps the best known. Netscape 3.0, for example, provides immediate access to Yahoo, Excite, Infoseek, Lycos, Search.COM, HOTBOT, LookSmart, WebCrawler, and AOL NetFind on a page directly reachable from its home page, and provides access to others such as Altavista indirectly from some of the pages.

In this assignment you will compare and contrast the usefulness and usability of **three** of these software products and write a report summarizing your findings. You must include **at least one search engine** and **at least one Web Guide (meta-index).**

THE ANALYSIS AND EVALUATION METHOD

Begin by calling up each system and “playing with” it for a little while, until you have a feeling for its functionality and for its interface. Usage seems to be self-evident; the systems come with no “user's manuals.” Yet the systems are different in their goals and scope, and in the way in which each has been realized. A careful analysis and evaluation of each is therefore instructive.

It is important to distinguish between usefulness and usability. *Usefulness* refers to a system's functionality, to the extent to which it allows individuals to carry out the tasks that they need and want to do at work, at school, and at play. *Usability* refers to the extent to which a system facilitates the carrying out of the tasks that are allowed to do, that is, the extent to which it is easy to learn and easy to use, and the extent to which it enables pleasant, comfortable, and safe usage.

You may want to structure this work by beginning with an examination of functionality and usefulness and then proceeding to two different methods for judging usability.

USEFULNESS

Begin by considering and developing **very brief** answers to the following questions:

Goals What is the goal of each system? State each as **succinctly** as possible.

Task Analysis The goal of task analysis is to construct a description of how people carry out their "work," both with and without computers. Describe **very succinctly** the steps involved in using search engines and/or Web Guides to look for information.

Conceptual Model A conceptual model is a description of the key ideas, terminology, operations, and relationships with which a user thinks about a system. Describe **very succinctly** the conceptual models which the designers of the systems seem to intend for the users.

Next, to carry out the analysis, you may want to develop a set of benchmark tasks with which you can compare and contrast the systems. These tasks will consist of on the order of 6-12 retrieval problems that differ significantly from one another and together seem to span the space of retrieval problems. Here is the beginning of a plausible list (you will want to start your own list in your own way, and not use what appears below):

Find out the most recent closing price of Apple stock.
Find out Jean Chretien's date of birth.
Find out how to drive to the Baseball Hall of Fame.
Find the opera schedule for La Scala for 1997-98.
Find out how to deliver a child.
…..

Finally, you will want to observe and describe how well your 3 chosen systems allow your users to carry out the set of benchmark tasks. You will have 4-5 users:
a) Two users chosen from outside your group (class members or friends) who will engage in **user testing** for you under your directions and observation.
b) Each member of your group who will try some of the tasks while being observed by the other two or three persons, leading to a discussion and **heuristic evaluation** of the systems.

JUDGING USABILITY WITH USER TESTING

One way to evaluate a system and its interface is with the help of evidence obtained by observing systems being utilized by real users, which is known both as *usability testing* and *user testing*.

Most usability testing involves experimental tasks that reflect important or frequent uses of the system, which should be selected through contact with real users and work situations, not by intuition. Participants are recorded as they engage in the tasks; problems and "errors" are identified. Interface changes to make these problems and errors less likely can then be proposed. Features that take particularly long to identify or operate can be identified (to do this, you must keep track of the amount of time taken to carry out specific tasks). Overall user satisfaction (or frustration) levels can be monitored, either informally or through questionnaires.

Generally, a number of people are asked to carry out the experimental task, in order to pick up a wider range of problems and to get some sense of the frequency and consequences of each. Nielsen and Landauer (1993) analyze usability problems described in 11 published projects and conclude that the maximum cost/benefit ratio for a medium-large software project would have been obtained through the use of three test users. Again, this is not a "magic number," but an indication that usability testing can be successfully carried out with only a modest budget. In your case, you will make do with only 2 test users.

A useful set of guidelines for usability testing appears in Nielsen (1992b), and is adapted and extended in the table below. Because of the limited time available for this assignment, you will be skipping or ignoring some of these suggestions, the ones that appear in small italic type. For example, you will not be using video (this will be introduced in Assignment 4), although you are free to tape record the sessions if your subjects permit.

Practical study design
- Reflect on the participants' backgrounds and how they might affect the study
- Be aware of problems that arise when experimenters know the users personally
- Prepare for the study carefully (avoid last minute panic)
- Select the tasks carefully to be representative and to fit the allotted time
- In general, start with an easier (but not frivolous) task
- Write down features of the system that are *not* being tested as well as those that are!
- Define the start-up state for the study precisely
- Define precise rules for when and how users can be helped during the study
- Plan the timing and cut-off procedure (if subject gets stuck) for each part of the study
- *Include reasonable provisions for data collection (e.g., tape or video recorder, keystroke capture where appropriate)*
- Plan data analysis techniques in advance
- *Carry out a pilot study (important but often overlooked)*

Written materials
- *Participant release form*
- *Questionnaire covering prior experience etc. (if relevant)*
- *Introduction to the study for users, including scenario of use*
- Checklist for experimenters
- *Evaluation survey (if relevant)*

Carrying out the study
- Let users know that complete anonymity will be preserved
- Let them know that they may quit at any time
- Stress that the system is being tested, not the participant
- Indicate that you are only interested in their thoughts relevant to the system
- Demonstrate the thinking-aloud method by acting it out for a simple task, such as figuring out how to load a stapler
- Hand out instructions for each part of the study individually, not all at once
- Maintain a relaxed environment free of interruptions
- Occasionally encourage users to talk if they grow silent
- If users ask questions, try to get them to talk (e.g., "What do you think is going on?" and follow predefined rules on when to help or interrupt to help.
- Debrief each user after the experiment

Improving the study
- *The pilot study should "debug" the study. This minimize changes during the study, allowing quantitative data analysis. But improvements may be warranted.*
- *Experimenters' role can be improved*
- *Tasks given to participant can be improved*
- *Written materials can be improved*

Figure 2.8 (BGBG, p. 85). Guidelines for user testing with thinking aloud (based on Nielsen, 1992a).

JUDGING USABILITY WITH HEURISTIC EVALUATION

Recently, Jakob Nielsen and his colleagues (see BGBG, pp. 82-84) have developed a usability inspection method known as *heuristic evaluation*. A few usability specialists are asked to look for problems in an interface or interface design by judging its compliance with a small set of very general design guidelines or "heuristics." Evaluators are trusted to use their experience and intuition to identify whether a guideline makes sense or not in a particular context and how to apply it. Heuristics focus the evaluator's attention on aspects of an interface that are often sources of trouble, making the detection of problems more likely.

Nielsen and Landauer (1993) analyze usability problems described in 11 published projects and conclude that the maximum cost/benefit ratio for a medium-large software project would have been obtained through the use of four evaluators. Although this is a not a reliable magic number, it does indicate that heuristic evaluation satisfies Nielsen's goal of being a "discount usability engineering method" that can be successfully applied on a modest budget. In your case, you will be making do with 3 or 4 evaluators, the number of members of your project team.

The table below lists ten design heuristics derived by Nielsen (1994a,b) from a factor analysis that suggests which guidelines provide the best coverage of published usability problems. They are "common-sense" guidelines that usability specialists are familiar with, but keeping them in mind while examining interfaces seems to be help evaluators in identifying problems.

"Visibility of system status. The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

Match between system and the real world. The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

User control and freedom. Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

Consistency and standards. Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

Error prevention. Even better than good error messages is a careful design which prevents a problem from occurring in the first place.

Recognition rather than recall. Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

Flexibility and efficiency of use. Accelerators—unseen by the novice user—may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

Aesthetic and minimalist design. Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

Help users recognize, diagnose, and recover from errors. Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

Help and documentation. Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large."

FIGURE 2.7 (BGBG, p. 83). Ten usability guidelines derived from a factor analysis of 249 usability problems encountered in 11 projects (Nielsen, 1994a, p. 30; see also Nielsen, 1994b).

Yet even ten guidelines result in conflicts and tradeoffs. It is always necessary to think about the context and how sensibly to apply any guidelines, and how to modify the guidelines for a specific situation, such as Internet search engines. Even more important is to carry out empirical testing of designs built with the aid of guidelines. Guidelines and heuristic evaluation should be used *together with* empirical user testing, not as a replacement for user testing.

ANALYSIS AND EVALUATION METHOD

1. Familiarize yourself and members of your team with some of these systems and pick 3 of them.
2. Find 2 user testing subjects, each of whom you plan to observe for 30-45 minutes.
3. Carry out the user testing and discuss what you have observed among members of your group.
4. Work with the systems yourselves and conduct the heuristic evaluation.
5. Say goodbye to your team members (for now) and work by yourself on your report.

YOUR TASK

Your task is to describe, compare, and evaluate these three systems in the context of their goals, environments, and users, and to document your findings in a report.

Your report should contain discussions and presentations of results from the various steps above and end with brief answers to the following .

- Write a one-sentence description of the differences between a search engine and a Web guide. (**REQUIRED OF ALL STUDENTS**)
- Overall, how good a job have the designers and implementers of each system done? What are the major strengths and weaknesses of each system? If you had to pick one system, which would you pick and why? (**REQUIRED OF ALL STUDENTS**)
- What have you learned about the strengths and weaknesses of user testing and heuristic evaluation? (**REQUIRED ONLY OF GRADUATE STUDENTS**)

You will carry out the research and evaluation in a team consisting of 3 or 4 individuals. You will then do the analysis and write the report **BY YOURSELF**.

In Assignments 2-4, you may continue with the same team or find a new team. You must then keep the same team for Assignments 2-4, so use Assignment 1 to get to know potential collaborators for the term and to decide if your goals, interests, work styles, and personalities are compatible.

WHAT YOU SHOULD HAND IN

All you need to hand in is your report. All reports must be typed and submitted on 8.5"x11" paper. **Structure and organization, spelling, grammar, word usage, and document appearance will count for roughly 20% of your grade.**

Each submission must include a title page with a meaningful title, your name, your student ID#, the names of the other students on your team, your tutor's name, the course name and number, and the date. The second page should contain a one paragraph executive summary of the document, and a table of contents. **The body of your report should be roughly 6 to 12 pages long, with roughly a 6-12 page Appendix containing supporting material** (e.g., raw data, partial transcripts of user testing sessions, tables of evaluations with respect to heuristic evaluation guidelines). This page count assumes a double-spaced presentation with a 10 or 12 point type face. **Note:** Points will be deducted for assignments that are needlessly long.