DESIGNING AN INFORMATION RETRIEVAL INTERFACE BASED ON USER CHARACTERISTICS[1]

Christine L. Borgman
Donald Case
University of California, Los Angeles

Charles T. Meadow
University of Toronto

Abstract

With the increasing number of information retrieval systems and databases available and the increasing demand of end users to utilize the systems, a need exists for improved interfaces and improved training mechanisms. This paper reports on a project to develop an integrated online instruction and assistance system to be used as a "front end" to the U.S. Department of Energy's RECON retrieval system. The conceptual framework for the interface design is based on individual characteristics of its current and prospective users, predominantly scientists conducting energy research. We are building a prototype based on information gained from interviews with scientists using the system (either directly or through a search intermediary) and interviews with search intermediaries. The paper reports on research in progress, including the results of the interviews and the preliminary design of the interface. The conference presentation will include a fuller description of the interface than can be specified here.

1.0 INTRODUCTION

1.1 Change in Demand for Information Systems and Services

In recent years the number of information systems and databases has grown at a rapid rate. As recently as ten years ago, use was concentrated in two systems (Lockheed DIALOG and SDC ORBIT) with fewer than a hundred databases between them. The most recent counts suggest that more than 2,400 databases can be accessed on some 300+ publicly available2 systems [1]. We assume from the number of systems in the marketplace that a need for the information exists. The question of who is using them remains, however. After a long tradition of use by trained search intermediaries, a number of writers are predicting increasing use of online systems by end users (e.g., [2,3,4,5,6]. Certainly, organizations such as DIALOG and BRS assumed that end user searching was increasing when they introduced systems aimed at this class of users (Knowledge Index and BRS/After Dark, respectively). General-purpose "hobbyist" computing services such as The Source and Compuserve have done well in a related market.

In spite of these expectations, we have yet to see data that confirm a strong movement by end users into online searching. It may be that online searching is a different market than the hobbyist computing market of The Source and Compuserve. As others have noted [e.g., 3,4,6,7], the major barrier to end users taking over the terminal appears to be lack of willingness to invest the effort in learning and maintaining online searching skills. The most direct

way to get end users onto systems may be to lower the effort barrier [3,4].

1.2 Online Assistance Technologies

The effort barriers to accessing information systems are composed of both initial training and skill maintenance requirements. The most common forms of training and skill maintenance used by search intermediaries require a significant time commitment: initial classroom training of one to three days followed by extensive practice and reading of documentation as systems and databases change.

A new technology, variously called "front ends," "microcomputer interfaces," and "gateways," promises to lower the effort barrier by combining training and online assistance in one mechanism. Further, by providing training and assistance in an online mode, they occur in the same medium as the information retrieval task, thus simplifying the learning process. Such technologies are not destined to replace search intermediaries entirely, but may serve to widen the audience for online searching, make it possible for end users to do some or all of their own searching (perhaps leaving the more difficult searches to intermediaries), and may be useful to intermediaries for access to systems that they do not use regularly [8,9].

Several such front end interfaces, both commercial and experimental, already have been developed. Well-known examples of research systems are IIDA [10,11,12] and CONIT [13,14]. Commercial systems include Sci-Mate (Institute for Scientific Information), In-Search (Menlo Corporation), MicroDisclosure (Disclosure Partners), and Search Helper (Information Access Corporation). Their role can be any of the following:

a. Teaching how to use a system,
b. Creating a "virtual" system, easier to learn and use than the target system (for which it serves as a front end).
c. Offering active help to a user in performing a search, such as in composing a search statement or deciding which file to search.

1.3 User Diversity

One problem in designing interfaces lies in the variety of styles with which users approach complex entities such as interactive information retrieval systems. Recent research has shown that users perceive systems differently from the way in which their syntax, database structure, and content are formally specified. Users' perceptions vary by previous experience and training, subject specialty, age, and cognitive style [15,16,17]. While the effects of training and experience are more widely known [18], subject background and cognitive style are less well-researched -- though certain to have a strong effect in use of retrieval systems.

We also find differences in information-seeking by user occupation, for instance in the information needs of scientists and engineers (and, to a lesser extent, behavioral scientists). These differences have been documented in a series of studies over three decades [e.g., 19,20,21,22]. Automation of the office environment has generated much interest in how managers gather information [23,24], and yet, as noted by Mick et al. [25], little of this knowledge has been used in implementing retrieval systems. Studies of individual methods of organizing information have shown wide variation among groups such as scientists and engineers [26] and office workers [27]. Professional online searchers also differ in their actual use of information systems [28,29]. Personal preferences in organizing and retrieving information may be related to cognitive styles in ways that could be adapted to computerized information systems.

1.4 The Need

The existing automated assistance programs, while helpful, treat all users alike. We now know that all users are not alike and suspect that much can be gained by acknowledging these differences in the design of assistance software. This paper reports on a research project to assess group differences in the use of one system, the U.S. Department of Energy (DOE) RECON3 retrieval system, and to design an automated assistance system tailored to these needs. We expect these techniques to

be superior to those that treat all systems and users alike.

We are developing a system of support software that has application well beyond this one project. It could be used as the basis for development of a wide range of interactive programs that require syntactic analysis of both the user's input and another processing system's responses. We believe that most complex systems will eventually have these front ends attached to them for ease of use.

2.0 RESEARCH QUESTIONS

The project attempts to address several research questions concerning the nature of information needs, the types of cognitive difficulties encountered, and methods of overcoming these problems with an integrated online instruction and assistance front end or subsystem. Among the questions addressed in the research are these:

1. What kind of information needs lead people to use the DOE RECON databases?

2. What types of needs does DOE RECON actually fulfill?

3. How do information needs vary by occupation or role (e.g., intermediary vs. end user; physical scientist vs. policy scientist)?

4. How does information seeking style vary by occupation or role?

5. How can a computer interface best alleviate cognitive difficulties in the information search process?

6. Can user information seeking behavior and mental models of a system be articulated sufficiently well to incorporate them into the interface design?

7. Is there an optimal set of interface parameters (for example, amount, type, and format of information displayed) for different types of users?

The first four questions will be addressed by the interviews with end users and search intermediaries of the DOE RECON system. The remaining three questions will be addressed by developing and testing a prototype online assistance and instruction system with experimental subjects.

3.0 RESEARCH METHOD

The research project is being conducted in four steps, as follows:

1. Interview direct end users (searching the system themselves), indirect end users (utilizing the system through a search intermediary), and search intermediaries of the DOE RECON system.

2. Design and implement a prototype integrated online instruction and assistance system for novice or infrequent users (both end users and search intermediaries), based on the information gathered in step 1. Outline a contrasting interface using a different technique (command-driven vs. non-command driven), also based on interview results.

3. Evaluate the prototype with users of energy information.

4. Make improvements to the prototype based on evaluation results.

As of this writing we have completed the interview phase of the project (step 1) and are working on the prototype design (step 2). These results are reported here. The conference presentation will include a description of the working prototype and the evaluation plan.

The four phases to the project, their goals, and their results to date are outlined in the following sections.

3.1 Step 1: User Study

3.1.1 Sample of DOE RECON Users

We selected our sample of respondents from a list of DOE RECON users provided by the Department of Energy. For practical and economic reasons, we limited the sample to those DOE RECON users working in California -- specifically in the Los Angeles and San Francisco Bay areas. We have no reason to suspect that these respondents differ in any way from users in other parts of the country. Phoning all holders of a DOE RECON password in these two metropolitan areas, we interviewed as

many users as would speak to us. We did not include holders of inactive accounts (still valid, but not being used). Most of the account-holders were search intermediaries; a few were direct end users. We used the search intermediaries to identify indirect end users who were heavy users of energy information and who had relied on the searchers for database retrieval.

3.1.2 Interviews

Our interviews gathered background information (education, research specialty, publishing activity, computer experience), and data on various information-use and system use variables such as the projects in which users participated, the uses to which they put the information they gathered; amount of time spent seeking information, both in manual and automated environments; and amount and type of information kept in personal files and the extent to which the files were organized; and questions about their use of the DOE RECON system and other energy-related information systems.

3.1.3 Description of Sample

We interviewed a total of 33 people (5 direct end users, 5 indirect end users, and 23 search intermediaries) from 20 organizations in the Los Angeles and San Francisco Bay areas. These 20 organizations can be broken down into four groups: four research institutes (public or private, non-teaching), three universities (two public, one private), nine energy-related companies (engineering, construction, utilities), and two federal agencies.

We found three of our end users in research institutes, one in a university, and six in energy-related companies. The job title most commonly held was manager (4), followed by technical analyst (2); other titles included computer scientist, engineer, professor, technical editor, and senior consultant. These people had held their jobs for an average (and median) of ten years. One end user held only a bachelor's degree; the rest had graduate degrees (7-PhD, 2-MS). All had research specialties connected to energy research, such as development of organic or synthetic fuels, research management, chemical engineering economic modeling, data processing, and writing.

The search intermediaries were drawn from research institutes (11), universities (2), energy-related companies (7), and federal agencies (3). Six held job titles of manager or administrator, eight had the title librarian, another eight were technical information specialists, and one held the title management assistant. They had held their jobs for an average of eight years (median = 7). One search intermediary had only a high school education, two had bachelor's degrees (two science, one non-science), twelve held MLS degrees, five held other master's degrees (one of whom had two master's degrees), and one held a doctorate (information science). Their subject specialties included energy, physics, chemistry and biomedicine.

3.1.4 Interview Results

Our interview results suggest some distinct differences in view of the DOE RECON system between end users (both direct and indirect) and search intermediaries in their use of the system and in their attitudes toward its role in information gathering.

The search intermediaries were skilled in DOE RECON use and usually several other systems. They were more sophisticated users and could articulate the differences in capabilities between DOE RECON and similar systems. Most searched the Energy Database (EDB) on DOE RECON and on other systems, most often DIALOG. They had specific reasons for choosing which system to search, based on the circumstances of the information sought. The most common tradeoffs were costs vs. features and the degree of overlap between multiple databases on individual systems. While DOE RECON costs slightly less than DIALOG on a connect-time basis and has some unique databases, it has fewer searching features than DIALOG, making DIALOG still a first choice for some.

DIALOG also includes many energy-related databases unavailable on DOE RECON, such as ENERGYNET, and databases on important peripheral subjects like engineering (COMPENDEX). The variety of databases on DIALOG, combined with the ease of transferring strategies from one database to another, was another reason that was often given for

choosing DIALOG over DOE RECON. Simple searches requiring only one database were frequently searched on DOE RECON by these searchers. Some intermediaries avoided DOE RECON primarily to minimize the number of systems on which they needed to maintain searching skills. Even though the RECON retrieval software was the predecessor to DIALOG, the two systems have evolved separately over the last 10 to 15 years. Their interfaces are now distinct, but with many structural features in common.

The search intermediaries provided constructive critiques of the systems, but were less specific about their users' needs. Individual searchers sometimes served a population of several thousand researchers and, not surprisingly, had little knowledge of the information needs of their users beyond the content of specific searches they had performed for them. Users come to them with a variety of types of questions, some wanting a small amount of information, but more typically wanting extensive subject-oriented retrieval (known-item searching was frequently diverted to manual retrieval methods). They had little knowledge of the research projects for which the information was needed or of what users did with the information once they had it. However, the search intermediaries were valuable sources of information about potential improvements in DOE RECON, such as allowing synonymous commands from other systems and improved formatting of search results.

In contrast, the end users were able to articulate their information needs clearly, but typically lacked knowledge of information retrieval and had little constructive criticism of the system or its output.

The end user view of the DOE RECON system was generally vague and uncritical. The indirect end users were frequently unaware of the specific system from which their data were obtained, so had no opinion on the system itself. They tended to deem the output satisfactory (although too voluminous) and were pleased that the system existed -- frequently they were impressed with the technology of Boolean-logic based retrieval per se and had no other system with which to compare the results. Those currently searching the system themselves felt competent to search it; those delegating their searches seemed satisfied to do so -- as one indirect end user put it, "if a specialist is there to do it for me, why should I learn to do it myself?"

End user criticisms tended to be database-oriented -- the system didn't contain enough older information, it lacked information on private energy research, and so on. Searching criticisms from direct end users were of the form "I'm never sure if I got everything." These users appeared to take a "shotgun" approach to searching, putting in several general terms for their area and combining them with a Boolean AND. They did not appear to be taking advantage of the system's more sophisticated features that might have provided more assurance of good retrieval.

The end users brought very specific information needs to the system, however. They typically searched in support of a given research project or for the development of proposals. Most had subject-oriented retrieval needs and high-recall needs ("everything on the topic"). Several end users had highly articulated notions of the information infrastructure of their disciplines -- when to go to journals, to reports, to handbooks, to colleagues, or to an automated retrieval system. Their use of the material once they got it varied considerably. It might go directly into a bibliography, or in the case of an industry proposal, it might be sifted to remove proprietary material. In other cases, the information might form the basis for a technical report, but not necessarily be cited directly. Some users kept elaborate files of the information collected and created subject indexes to it, others put things in folders under very general categories, and still others just tossed it when a project was over, on the assumption that it was easier to retrieve it again than to find a way to store it.

3.2 Step 2: Prototype Development

3.2.1 Lessons from the Interviews

From the interview results we have drawn several conclusions as to the characteristics of the front end interface needed. Our primary conclusion is that needs vary by degree

of experience with the system, rather than on other factors such as educational background or research area. The most help is needed by the inexperienced or infrequent searchers, both end users and search intermediaries. Our efforts are aimed at this group. The system criticisms of the highly-skilled searchers relate to fine-tuning -- getting that "last 2%" efficiency from the system. The investment required in a front end to support these sophisticated needs could not be justified relative to the benefits gained, at least not for a project of this size.

The style of use and the degree of computer experience vary, even among the less sophisticated users. For these reasons, we will develop two versions of the interface based on style and computer-experience preferences. One version will employ a fairly standard command language technique, training the user in the commands to use for searching, and assisting in their use in an online mode. The second version will employ a "non-command" technique, where users will never give explicit commands, but rather will be prompted for a response or asked questions to conduct the search. The non-command language interaction is in the form of a conversation, and does not involve menus, per se. The effectiveness of the two methods will be compared in the final evaluation.

3.2.2 Development Process

The prototype is being developed in several steps, with an iterative process of design and testing. The first step was to develop the formal specification for a teaching/assistance software system. As currently specified, the front end will have the capabilities to do the following:

a. Provide CAI courseware to users or future users. Course content includes the concepts and mechanics of searching, based on the information needs and conceptual framework for the system determined as a result of the interviews.

b. Provide CAI courseware on data structures and content for the databases of DOE RECON, complementary to the search concepts course material.

c. Provide a set of user assistance programs, each able to be selected and used independently of the others and at the user's option, and which collectively serve as a search intermediary for an average end user searcher. The major components of such a system, from the user's viewpoint, will be:

i. Assistance in selecting a file from among those available.

ii. Assistance in formulating the logic of a search.

iii. Assistance in selecting the vocabulary elements -- key words or descriptors -- for use in a search.

iv. Assistance in using a library of prepared skeleton searches, or scripts, which are subject or database specific. These will be common searches for which the user must supply parameters such as the name of an author or of a chemical substance.

v. For all of the above, assistance in iterating a search to improve on early results.

vi. A quick reference capability for the user to look up command and field definitions and other well-defined concepts.

Each of these elements has been implemented in some form in other training systems. We are designing an integrated system that combines the best elements of known technology and theoretical knowledge about user differences, assembled into a uniform entity and capable of significant performance improvement over its individual components.

The interface will incorporate many error detection capabilities. At the lower levels of use, we will capture and correct (if sufficiently identifiable) errors in syntax and logic. At the higher levels of use assistance will be largely heuristic. Generally-accepted and identifiable rules for development of search programs do not yet exist. In lieu of these, our goal is to design programs that "work" most of the time and gradually improve them based on user feedback. We will be capturing online monitoring data from system use which will provide a continuous feedback loop.

3.3 Prototype Evaluation

During the summer of 1985, we will perform a preliminary evaluation of the prototype using experimental subjects drawn from the population of graduate students at UCLA doing energy research. Online monitoring data captured from the user/system interaction will provide the primary data for evaluation. The immediate focus of the evaluation will be to assess how well the prototype meets the needs of naive and infrequent users, as identified in the interviews.

3.4 Implementation of Prototype

We will incorporate the evaluation results and refine the prototype into a working, but not fully implemented, system. If funds are available for a second year of the project, the system would be fully implemented in both the command and non-command modes.

4.0 CONCLUSIONS

Our interview results suggest a diversity of information needs, of conceptual views of the system, and of searching capabilities required. We believe that it is possible to capitalize on this diversity in designing an online instruction and assistance system for DOE RECON. The conference presentation will report on further progress toward that goal.

NOTES

1. The work reported here is funded by a grant from the United States Department of Energy to the University of California, Los Angeles. Grant number DE-FG05-84TC40004 AMD A000.

2. Publicly available is defined as unrestricted access based on direct fees or organizational membership on systems that can be reached through value-added telecommunications networks.

3. DOE RECON is distinct from NASA RECON, which has evolved separately and supports a different set of databases.

REFERENCES

[1] Williams, Martha E.; Lannom, L.; Robins, C., eds. & comps. Computer-Readable Databases: A Directory and Data Sourcebook. Chicago, IL: American Library Association, 1984.

[2] Faibisoff, Sylvia G.; Hurych, Jitka. Is There a Future for the End User in Online Bibliographic Searching? Special Libraries, 72:4 (1981), 347-355.

[3] Meadow, Charles T. Online Searching and Computer Programming: Some Behavioral Similarities (Or ... Why End Users will Eventually Take Over the Terminal). Online, 3:1 (1979a), 49-52.

[4] Meadow, Charles T. Letter to the Editor [Response to an editorial commenting on his previous article]. Database, 6:4 (1983), 49-50.

[5] Sewall, Winifred; Bevan, Alice. Nonmediated Use of MEDLINE and TOXLINE by Pathologists and Pharmacists. Bulletin of the Medical Library Association, 64:4 (1976), 382-391.

[6] Thompson, Benna Brodsky. Future Direct Users of Sci-Tech Electronic Databases [Guest Editorial]. Database, 6:2 (1983), 6-9.

[7] Neway, Julie M. The Role of the Information Specialist in Academic Research. Online Review, 6:6 (1982), 527-535.

[8] Meadow, Charles T. The Computer as a Search Intermediary. Online, 3:3 (1979b), 54-59.

[10] Meadow, Charles T.; Hewett, Thomas T.; Aversa, Elizabeth. A Computer Intermediary for Interactive Database Searching, Part I, Design. Journal of the American Society for Information Science, 33:5 (1982), 325-332, and Part II, Evaluation. 33:6 (1982), 357-64.

[9] Meadow, Charles T. On-line Searching and Computer Programming: Some Behavioral Similarities, Online, 3:1 (1979), 49-52.

[11] Meadow, Charles T.; et al. IIDA Final Report, NSF Grant No. DSI 77-26524. Philadelphia: Drexel University School of Library and Information Science and Franklin Institute Research Laboratory, June 1980.

[12] Meadow, Charles T.; et al. Individualized Instruction for Data Access (IIDA), Final Design Report. Philadelphia: Drexel University Graduate School of Library Science and Franklin Institute Research Laboratory, July 1977.

[13] Marcus, Richard S. An Automated Expert Assistant for Information Retrieval. Proceedings of the 44th Annual Meeting of the American Society for Information Science, 18. White Plains, NY: Knowledge Industry Publications, 1981, 270-273.

[14] Marcus, Richard S.; Reintjes, J. Francis. A Translating Computer Interface for End-User Operation of Heterogeneous Retrieval Systems: Part I, Design; Part II, Evaluations. Journal of the American Society for Information Science, 32:4 (1981), 287-303, 304-317.

[15] Borgman, Christine L. The User's Mental Model of an Information Retrieval System: Effects on Performance. Ph.D. Dissertation, Stanford University, 1984.

[16] Egan, Dennis E.; Gomez, Louis H. Assaying, Isolating and Accommodating Individual Differences in Learning a Complex Skill. In Ronna F. Dillon, ed., Individual Differences in Cognition. New York, NY: Academic Press, 1984.

[17] Paisley, William. Information and Work. In Brenda Dervin & Melvin Voigt, Eds., Progress in the Communication Sciences, Vol. 2. Norwood, NJ: Ablex, 1980, 113-165.

[18] Fenichel, Carol H. Online Searching: Measures that Discriminate Among Users with Different Types of Experience. Journal of the American Society for Information Science, 32:1 (1981), 23-32.

[19] Crawford, Susan. Information Needs and Uses. In Williams, Martha E., ed. Annual Review of Information Science and Technology, 13. White Plains, NY: Knowledge Industry Publications, 1978.

[20] Hagstrom, Warren. The Scientific Community. New York, NY: Basic Books, 1965.

[21] Martyn, J. Information Needs and Uses. In Williams, Martha E., ed. Annual Review of Information Science and Technology, 9. White Plains, NY: Knowledge Industry Publications, 1974.

[22] Menzel, H. The Flow of Information Among Scientists: Problems, Opportunities, and Research Questions. New York: Columbia University, Bureau of Applied Social Research, 1958.

[23] Ackoff, R. Management Misinformation Systems. Management Science, 14:B (1967), 147-156.

[24] Mintzberg, Henry. The Nature of Managerial Work. New York, NY: Harper & Row, 1973.

[25] Mick, Colin; Lindsay, George; Callahan, Daniel. Toward Usable User Studies. Journal of the American Society for Information Science, 31 (1980), 347-356.

[26] Jahoda, G.; Hutchins, R.D.; Galford, R.R. Characteristics and Use of Personal Indexes Maintained by Scientists and Engineers in One University. American Documentation, 17 (1966), 71-75.

[27] Malone, Thomas W. How Do People Organize Their Desks? Implications for the Design of Office Information Systems. Palo Alto, CA: Xerox Palo Alto Research Center, 1982.

[28] Oldroyd, Betty K.; Citroen, Charles L. A Study of Strategies Used in Online Searching. Online Review, 1:4 (1977), 295-310.

[29] Fidel, Raya. Online Searching Styles: A Case-Study-Based Model of Searching Behavior. Journal of the American Society for Information Science, 35: 4 (1984), 211-221.