DOCUMENT RESUME

ED 105 891 IR 001 929

AUTHOR	Martin, Thomas H.
TITLE	A Feature Analysis of Interactive Retrieval Systems. Final Report.
INSTITUTION	Stanford Univ., Calif. Inst. for Communication Research.
SPONS AGENCY	National Science Foundation, Washington, D.C. Office of Science Information Services.
REPORT NO	SU-COMM-ICR-74-1
PUB DATE	Sep 74
NOTE	105p.; Report period October 1972 through October 1974
EDRS PRICE	MF-$0.76 HC-$5.70 PLUS POSTAGE
DESCRIPTORS	*Comparative Analysis; Computer Programs; Computer Storage Devices; Design; Human Engineering; *Information Retrieval; *Information Systems; Interaction; Management Systems; *Man Machine Systems; On Line Systems; *State of the Art Reviews
IDENTIFIERS	*Interactive Retrieval Systems

ABSTRACT

The command language features of 11 on-line information retrieval systems are presented in terms of the functional needs of a searcher sitting at a terminal. Functional areas considered are: becoming familiar with the system, receiving help when in trouble, regulating usage, selecting a data base, formulating simple queries, expressing single concepts, interconnecting concepts, displaying results simply, and controlling the display. Features felt most essential to on-line searching are live help, users' guides, boolean operators, search field control, suffix removal, relational operators, dictionary access, request sets, search review, pre-defined formats, on-line formatting, and off-line printing. (SK)

a report of the

INSTITUTE FOR COMMUNICATION RESEARCH

STANFORD UNIVERSITY

R... NSF COMM ICR-74-3

A FEATURE ANALYSIS OF INTERACTIVE RETRIEVAL SYSTEMS

September 1974

Thomas H. Martin

Institute for Communication Research
Stanford University
Stanford, California 94305

Final Report for Period October 1972 - October 1974

Prepared for

Office of Science Information Service
National Science Foundation
Washington, D.C. 20550

Research supported by the National Science Foundation under Research Grant NSF GN 36160 awarded to Stanford University in October 1972

Principal Investigator: Edwin B. Parker, Ph.D.

BIBLIOGRAPHIC DATA SHEET

SU-COMM-ICR-74-1

SEPTEMBER 74

INTERACTIVE RETRIEVAL

NSF GN 35163

Final: 10/72-10/74

National Science Foundation

Washington, D.C. 20550

The command language features of eleven different on-line information retrieval systems are presented in terms of the functional needs of a searcher sitting at a terminal. Functional areas considered are: becoming familiar with the system, receiving help when in trouble, regulating usage, selecting a data base, formulating simple queries, expressing single concepts, interconnecting concepts, displaying results simply, and controlling the display. Features felt most essential to on-line searching are live help, users' guides, boolean operators, search field control, suffix removal, relational operators, dictionary access, request sets, search review, predefined formats, on-line formatting, and off-line printing. It is concluded that no sharp distinction exists between management information and bibliographic retrieval. The report is intended for use by designers of interactive retrieval systems and by students of system design.

*Information retrieval
*Man-machine systems
*Design
*Information systems
Data storage systems
Psychometrics
Human factors engineering

Bibliographic information
On-line information retrieval

Springfield, VA 22151

194

4

TABLE OF CONTENTS

1.0 INTRODUCTION

1.1 THE INTENDED AUDIENCE

This report is intended for use by designers of interactive retrieval systems and by students of system design. In the report command language features of eleven different online information retrieval systems are presented in terms of the functional needs of a searcher sitting at a terminal.

There are a number of assumptions underlying the report that should be made clear. The first is that there has been enough time for competition in the marketplace since initial implementation for systems that do not respond to user needs to disappear. The second assumption is that command languages are sufficiently easy to change so that reoccurring suggestions from users are reflected in revised versions of systems. A third is that most (but not all) of the systems considered in the report were developed independently of the others. One can see this by noting that systems vary widely in the words used for invoking functions while they vary only slightly in the types of information users must supply when invoking the functions. A fourth assumption is that in the absence of totally new technologies, the functional needs of the searcher are not likely to change significantly. The net effect of the argument is that the juxtaposition of features from relatively successful operational retrieval systems can tell one a great deal about the needs of the interactive searcher. Hopefully, the report will ease the transition to a common vocabulary for talking about searching, and perhaps facilitate the development of a common interface for [illegible] different systems. At the very least, it should [illegible] future designers build new systems in ignorance of the [illegible] their predecessors.

There are other audiences that are likely to think the report responds to their needs. Purchasers of interactive retrieval system service and searchers themselves probably will assume that the report can help them choose between systems. To the extent that it suggests what to look for, it is useful, but it should not be relied upon for more. First of all, the system descriptions are intentionally out of date. Secondly, there are many factors a purchaser should consider that are not discussed here - things like cost, reputation for service, and ability to tailor a contract to the specific needs of a purchaser. Without information about the particular set of data bases to be accessed, the educational background of users, and the usage loads at various times of the day and year, no adequate rating of systems is possible. Purchasers of systems or service should plan to carry out benchmark studies in which they determine what a typical load would look like and would cost if run on each of the systems under consideration. Finally, it should be pointed out that many more systems are available in the marketplace than are considered in this report. The National Bureau of Standards (3) has recently put together an index of interactive information systems.

1.2 THE FEATURE ANALYSIS

In order to assess the value of the report and the reasoning behind the selection of materials, the reader must understand the procedures followed in compiling it. The major reason for carrying out the analysis was to bring system designers together and to focus their attention upon similarities and differences in existing user interfaces. Interpersonal discussion is one of the most effective methods for changing human behavior, and user interfaces will not move toward [illegible] designers toward commonality.

User interfaces are not the major interest of most designers. Designers spend most of their time making sure that their systems are reliable and cost effective. What they would most like to know and are least willing to reveal are strategies for increasing system throughput without degrading performance, or strategies for attracting and retaining a money-making mix of users. We intentionally did not probe too far into system internals or system clientele because these were felt to be privileged information. The user interface was a neutral topic for designers yet of great interest to searchers.

A three-day workshop was held at Stanford University April 23-25, 1973. Representatives of each of the eleven systems, a five-man panel of experts, and a handful of invited guests met for intensive discussions and system demonstrations. The five two-hour discussions focused on 1) the searcher/task environment, 2) the data base environment, 3) & 4) search and display features necessary for information retrieval, and 5) instructional and diagnostic features. At least one forty-minute demonstration was given of every system, and most systems were displayed twice. Video tapes were made of six of the system demonstrations directly off of the terminal and are available on loan from ASIS's SIG/UOI (the user on-line interaction special interest group of the American Society for Information Science).

In prepartion for the workshop, user manuals were solicited from each of the system representatives and each representative was visited for half a day. The manuals were used to extract much of the material that appears in the report, and the site visits were used to fill in missing gaps and to probe representatives regarding the user population. It rapidly became clear that designers knew little about the habits and characteristics of individual users. In most cases clients were known only by an account number that lumped together a number of different searchers. Only gross generalizations were available regarding whether end users were carrying out their own searches, whether searchers were frequently switching between data bases, or whether they tended to use video terminals instead of teletypes. After the site visits and before the workshop, matrices were drawn up summarizing the degree to which each of the eleven systems incorporated various features. At the workshop the matrices were used to structure discussion between designers. Revised versions of the matrices have been included as an appendix to the report. Revisions have been made partially at the request of designers and partially to make the matrices an accurate summary of the total report. System representatives have also reviewed the chapters of the report dealing with their systems and have made revisi so that the report accurately states the status of their systems as of April, 1973.

We would like to thank Donald Black and Robert Katter from System Development Corporation, Mark Radwin and Roger Summit from Lockheed, Lawrence Stevens and Howard Coleman from Informatics, Stanley Friedman from IBM, David Colombo and John Fried from Battelle Memorial Laboratories, Richard Giering from Mead Technology Laboratories, Donald Hillman and Louis Stern from Lehigh University, Benjamin Mittman and Wayne Dominick from Northwestern University, Richard Marcus from MIT, and Charles Goldstein from the NASA Lewis Research Center for their contributions to the workshop and report.

In addition to the written report, a twenty-minute color 16mm. film called "Access" has been prepared for introducing college level students to interactive searching. It is being distributed by the Extension Media Center, University of California at Berkeley. In the film, viewers are informed that searching is like looking for a needle in a haystack and that it involves the entering of a request, revision, supplementing with related terms, browsing, and display of results. They are shown a number of different types of users searching for different kinds of information in different kinds of data bases. Information technology is presented in an historical persepective as a means whereby the special interests of small groups can be served, and will be served more effectively in the future by means of networks, inexpensive storage media, and by human information brokers. The film can be used in conjuction with the report as an introduction to interactive searching.

1.3 HOW TO READ THE REPORT

The report itself is divided into chapters dealing with the various functional needs of interactive searchers. Matters outside the control of retrieval system designers are not included. For example, little is said about interactive terminal features or about operating system login procedures. Issues that are beyond the concern of searchers are not included. No mention is made, for instance, of file structures, parsing techniques, or updating techniques. There are a number of issues that it would have been nice to include but which were not since data was not available. For example, nothing is said about average and worst case delays due to system failure, or about recall/precision measures of system effectiveness.

The appendix to the report acts as an index to the subsections. The rows in the four matrices direct the reader to the related subsections. Subsections generally consist of statements supported by command language syntax. Rather than adhering to a uniform notation for commands, presentation techniques are tailored to the special circumstances of each subsection so that the supporting material will be as readable as possible. System names have been pulled into the margin so that they either can be ignored or so that one can locate the manner in which a particular system has implemented a particular feature. Chapters begin with tutorial material placing the functional need in context and contrasting its use in interactive searching with its role in other types of systems. Then features which are encountered by default are introduced followed by features that must be specifically invoked. In some chapters features are organized in terms of subfunctions rather than in terms of how they are invoked. The report concludes with general recommendations to the designers of future systems.

1.4 RELATED WORK

The report fits into a background of activity that has been taking place over a number of years. In January, 1971 a workshop dealing with the user interface for interactive bibliographic searching (15) was sponsor d by the American Federation of Information Processing Societies (AFIPS). The thirty-eight participants at the workshop represented a number of different interactive information retrieval systems. Many articipants were unfamiliar with each other's systems and many expressed a desire to compare systems so that a basis could be laid for further discussion. Available comparisons of systems by Seidon (14) and Welch (16) were effectively out of date. The participants drew up a list of the functional areas involved in searching that form the basis for chapter divisions in the current report.

One of the participants at the 1971 workshop was William Olle, who headed a similar effort to place features in a functional framework for data base management systems (2) under the auspices of the Conference on Data Systems Languages (CODASYL). While there is an overlap between interactive retrieval langauages and data base languages, Olle indicated that he thought a similar feature analysis should be conducted for information retrieval systems (12). When the designers of the eleven retrieval systems met at the April, 1973 Stanford workshop, one of their conclusions was that no sharp discstinction should be made between retrieval of management and bibliographic information.

The National Bureau of Standards has for a number of years kept track of the state of the art in information retrieval system development (3, 4, 6, 7). As the field has stabilized and pressure for standards has increased, NBS representatives have started laying a groundwork for uniform login and logout procedures.

While the National Bureau of Standards representatives are not attempting to impose standards in functional areas of searching, Reintjes and Marcus (13) at M.I.T. have been attempting to develop a translating interface language that can be used for coupling systems so that a searcher need not worry about which retrieval system he is using. Although they have made some progress the task is vast and much remains to be done.

At the other end of the spectrum, a survey under the direction of Carlos Cuadra at System Development Corporation is underway in an attempt to discover what purchasers, intermediaries, and end users actually think about the systems they have been using. In Europe, D.I. Raitt of the European Space Research Organization is using the same questionnaire to discover how Europeans are reacting to the systems available to them.

There are a number of places one can turn to for more information about interactive searching. A thorough and excellent introduction to the field is Lancaster's Information Retrieval On-line (5). If one prefers collections of papers, Walker's Interactive Bibliographic Search: The User Computer Interface (15), Meadow and Henderson's Interactive Bibliographic Systems (11), or May's Automated Law Research (10) are recommended.

There are not as many places where one can turn for information regarding the design of user/computer interfaces. James Martin's <u>Design of Man-Computer Dialogues</u> (8) is one of the only textbooks. The <u>Annual Review of Information Science and Technology</u> has recently included chapters on "The User Interface in Interactive Systems" (1, 9). The Special Interest Group "User On-line Interaction" of the American Society [illegible] rmation Science periodically holds technical sessions dealing with user interface problems.

1. Bennett, John L. "The User Interface in Interactive Systems". In Cuadra, C. (Ed.) Annual Review of Information Science and Technology, Vol. 7. Washington, D.C.: American Society for Information Science, 1972, pp. 159-196.

2. CODASYL Systems Committee Technical Report. Feature Analysis of Generalized Data Base Management Systems. New York, N.Y.: Association for Computing Machinery, May, 1971, 520pp.

3. Fife, Dennis W. et al. "A Technical Index of Interactive Information Systems". Washington, D.C.: National Bureau of Standards, February 1974, 68p. (NBS Technical Note 819).

4. Fong, Elizabeth. "A Survey of Selected Document Processing Systems". Washington, D.C.: National Bureau of Standards, October 1971, 65pp. (NBS Technical Note 599).

5. Lancaster, F. Wilfred. Information Retrieval On-line. Los Angeles, Ca.: Melville Publishing Co., 1973, 596pp.

6. Marron, Beatrice. "A Study of Six University-Based Information Systems". Washington, D.C.: National Bureau of Standards, June, 1973, 95pp. (NBS Technical Note 781).

7. Marron, Beatrice et al. "A Mechanized Information Services Catalog". Washington, D.C.: National Bureau of Standards, February, 1974, 47pp. (NBS Technical Note 814)

8. Martin, James. Design of Man-Computer Dialogues, Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1973, 559pp.

9. Martin, Thomas H. "The User Interface in Interactive Systems". In Cuadra, C. (Ed.) Annual Review of Information Science and Technology, Vol. 8. Washington, D.C.: American Society for Infromation Scie[illegible] 1973, pp. 203-219.

10. May, Ronald A. (Ed.) Automated Law Research. Chicago, Ill.: American Bar Association, 1973, [illegible]p.

11. Meadow, Charles T. and Henderson, Madeline B. (Eds). Interactive Bibliographic Systems. Oak Ridge, Tenn.: U.S. Atomic Energy Commission, April 1973, 205pp. (CONF-711010).

12. Olle, T. William. "A Comparison between Generalized Data Base Management Systems and Interactive Bibliographic Systems". In Walker, D. (Ed.) Interactive Bibliograph Search: The User/Computer Interface, Montvale, N.J.: AFIPS Press, 1971, pp. [illegible]3-213.

13. Reintjes, J. Francis and Marcus, Richard S. "Interim Report on Research in the Coupling of Interactive Information Systems". Boston, Mass.: Massachusetts Institute of Technology, Feb., 1974, 36pp.

14. Seidon, H. R. "A Comparative Analysis of Interactive Storage and Retrieval Systems." Santa Monica, Ca.: System Development Corporation, Jan., 1970, (Report #TM-4421).

15. Walker, Donald E. (Ed.) Interactive Bibliographic Search: The User/Computer Interface. Montvale, N.J.: AFIPS Press, 1971, 311pp.

16. Welch, Noreen O. "A Survey of Five On-line Retrieval Systems". Washington, D.C.: Mitre Corporation August, 1968, (Report #MTP-322).

2.0 THE SYSTEMS SELECTED

2.1 INTRODUCTION

Eleven systems (Battelle's BASIS, IBM's STAIRS, Lockheed's DIALOG, MIT's INTREX, NASA Lewis Research Center's NASIS, Lehigh's LEADER, Mead Technology Laboratories' DATA CENTRAL, Informatics' RECON, System Development Corporation's ORBIT II, Northwestern's RIQS, and Stanford's SPIRES II) were chosen for inclusion in the comparative analysis. The criteria used for selecting systems were that the system had to be 1) operational, 2) on-line and interactive, 3) able to handle multiple users simultaneously, 4) able to handle multiple data bases, 5) able to process data bases with variable length entries and elements, 6) demonstrable to the public, and 7) primarily oriented toward information storage and retrieval. In only one case was an exception made from these criteria: INTREX does not customarily handle more than one data base.

Many systems met the criteria of the comparative analysis but were not included. A number are close relatives of systems included in the analysis (ELHILL, NASA/RECON, BCN, LEXIS) so could be excluded without prejudicing the generality of the comparison. The versions included in the analysis were chosen because the person making the greatest contribution to the design of the system could represent it. Both DIALOG and RECON were included because Roger Summit of Lockheed and Larry Stevens (of NASA in 1968, of Informatics during the comparative analysis) worked together to establish the specifications for the original NASA/RECON.

Other systems were developed outside of the United States (QUIC/LAW of Canada, LIS of Sweden) and it was felt that since system representatives were being paid to come to the Stanford workshop, that it would be best to limit the analysis to U.S. systems.

Some systems came very close to being included but were not for extraneous reasons. Infodata's INQUIRE system was to be included but bowed out a month before the workshop because of the press of work. The *New York Times'* system would have been included except that it required a terminal that was not available on the West coast, was used with single data base, and was written at IBM Federal Systems Division so would not have been represented by a system designer. Lawrence Livermore Laboratories' MASTER CONTROL would have been included except that it runs on a very unusual computer and can only be made publicly accessible by moving it to a declassified computer. The Defense Documentation Center's system was considered but not included for similar reasons. MRI's System 2000 would have been included except that it was discovered too late. Other systems considered but not used were: University of Washington's SOLAR, University of Pittsburgh's PIRETS, Syracuse's SUPARS, and Stanford Research Institute's Augmented Intellect.

No inference should be drawn that a system is in any way inadequate because it was not included in the analysis. We sought to limit the number of participating systems so that we could conduct a manageable workshop. Hopefully, the items analyzed throughout the report will prove to be sufficiently comprehensive so that other systems can be added at a later date.

2.2 HISTORICAL PERSPECTIVE

Systems were selected to participate in the comparative analysis as of January, 1973. System representatives came together April 23-25, 1973 for the comparative analysis workshop. Throughout the rest of the report the present is assumed to be Apr. 23, 1973, and this will be emphasized in the next two sections. However, the systems came into existence much earlier, and most have continued to exist beyond that date. In this section, each of the selected systems is introduced and presented as it has developed over time.

2.2.1 BASIS was developed at Battelle Memorial Institute's Columbus Laboratories and was first put into service during July, 1970. Initially intended for use by Battelle's information analysis centers and their clients, its use has broadened to cover many applications of information processing and modeling. While Battelle will sell or lease software, and will put up a customer's data base, potential clients are judged in terms of the research content of their proposed effort. BASIS

Since April, 1973 the system has also been installed in [illegible], the largest bibliographic data base has been removed, [illegible] has been developed, functions have been extended for interactive modeling and data analysis, and the system [illegible] used for transportation and energy planning [illegible].

[illegible] International [illegible] [illegible]

[illegible]

2.2.3 DIALOG was developed at Lockheed and was first put into service during 1966. The 1968 version of DIALOG was adapted by Lockheed to NASA's requirements (and was named RECON) and is now in the public domain. Since 1968 the system has been extended in many ways. While it is still possible to buy or lease the software, Lockheed encourages people to subscribe as interactive searchers. DIALOG

Since April, 1973 a great many more data bases have been put online, including the National Agricultural Library's CAIN (Cataloging and Indexing), the American Psychological Association's Psychological Abstracts, INFORM COMPENDEX, INSPEC, Chemical Abstract's CONDENSATES, and others.

2.2.4 INTREX was developed at the Massachusetts Institute of Technology and was first put into service during 1969. It was originally intended as an experimental system (but in an operational environment) for trying out new approaches to in-depth indexing and interaction via terminals which could display both alphanumeric characters and microfiche. INTREX

Since April, 1973 the system has been taken down and is being reprogrammed for use on an IBM 370/165 computer. In a related effort, an interface is being designed that translates what one requests using a single command language into the language of one or more other systems.

2.2.5 NASIS was developed at NASA's Lewis Research Center and was first put into service during [illegible]. The revised version of the system discussed in this report was in operation only a few months during 1972. NASIS was intended for use as a data base management system. Its major use was with descriptors of photographs of the United States taken by the ERTS satellite. The system is no longer being supported by NASA, although the software is still available. NASIS

2.2.6 LEADER (often referred to as LEADERMART when used in conjunction with Lehigh's Mart Library) was developed at Lehigh University and first put into service during March, 1971. It is intended primarily for searching of bibliographic or textual data bases by students at Lehigh, although service was for a time being sold to outside subscribers. LEADER is unique in its emphasis upon natural language processing and de-emphasis upon boolean operators. LEADER

Since April, 1973 the large bibliographic data bases have been removed from the system and reprogramming efforts are underway to make it more usable with textual (non-indexed) material.

2.2.7 DATA CENTRAL was developed at Mead Technology Laboratories and was first put into service during 1968. At various times before April, 1973 it was used for searching Psychological Abstracts, Epilepsy Abstracts, and the case law of the state of Ohio. DATA CENTRAL is intended for use as a data base management system. Either the software can be leased or data bases can be placed on the parent system. DATA CENTRAL

Since April, 1973 the system has been revised and is currently being reprogrammed to run on the DEC PDP-11 series of computers.

2.2.8 RECON was developed by Lockheed for NASA and is maintained by Informatics (in conjunction with their STIMS file maintenance package). Since 1968 the system has been extended in many different ways. RECON is installed not only at NASA, but also at the Department of Justice, the National Oceanographic and Atmospheric Administration, and other government agencies. Informatics supports RECON at these installations as well as selling service and making their computer available for storage of data bases. RECON

Since April, 1973 The Environmental Protection Agency's ENVIRON data bases and George Washington University's POPINFORM data base have put onto the system. The TOXICON data base has been removed and has become the TOXLINE data base on ELHILL.

2.2.9 ORBIT II was developed by System Development Corporation in conjunction with the development of ELHILL for the National Library of Medicine. An early version of ORBIT was first put into service during 1970, and ORBIT II became operational in 1972. SDC both sells or leases software and also puts up data bases for subscription searching on the parent system. ORBIT

Since April, 1973 a new version of the system, ORBIT III, has been developed. In addition, a great many more data bases have been put online, including the National Agricultural Library's CAIN, COMPENDEX, NTIS, INFORM and others. System improvements are being implemented on a continuing basis.

2.2.10 RIQS was developed by Northwestern University and was first put in service in September, 1969 (then called RIMS). Since that time it has undergone extensive changes and an instructional interface RIQSTUTOR has been built for it. RIOS is intended for maintenance and searching of small-to-medium size data bases whether they are bibliographic, textual, or numerical. It can be used in conjunction with graphics plotting and statistical analysis. RIOS

Since April, 1973 the RIOS User's Manual has been completely rewritten, a procedural capability has been developed to enhance both interactive searching and statistical data analysis, and an interactive graphics package has been developed. The system has been installed in Europe and is actively being marketed in the United States.

2.2.11 SPIRES-II was developed by Stanford in conjunction with the development of the BALLOTS library cataloging system. An early version of SPIRES was first put in service in 1969, and SPIRES II became operational during September, 1972. SPIRES II can handle a wide variety of different types of data bases, including those having hierarchies of fields within fields. While it is possible to transfer the software or to put a data base up for searching, SPIRES is intended primarily for searching and file maintenance by people at Stanford. SPIRES

Since April, 1973 the svstem has been revised substantially. It now provides for offline report generation and predefined formats online. As it is being modified to run on the IBM 370/158, it is being augmented to have ' .t language capabilities.

2.3 THE SYSTEM'S ENVIRONMENT

Each of the systems must reside on a computer, under an operating system, (perhaps) connected by a data communications network to a terminal where a user searches in one or more data bases. The data base may belong to the user and is being stored on his computer with software leased from the system owner, the data base may be stored on the system owner's computer, or the user (as is customary) may have no ownership interest in the data base. In this section most of these aspects of the eleven systems are considered. Missing from what follows is a clear picture of the users, how frequently they search, and the problems they run into. System representatives were asked to characterize their users but few could do so.

Outside of the universities, most searchers are intermediaries rather than end users (but not all). An NSF-sponsored study is being conducted by Carlos Cuadra of System Development Corporation to find out the answers to user-related questions.

2.3.1 As of April 23, 1973 slightly modified versions of the system were being operated by others. These versions of the system would have qualified for inclusion in the comparative analysis.

a. ORBIT
b. DIALOG
b. RECON
c. STAIRS
d. DATA CENTRAL

a. National Library of Medicine's
   ELHILL (popularly known as MEDLINE)
b. European Space Research Organization's
   ESRO
   National Aeronautics and Space Administration's
   NASA/RECON
   Atomic Energy Commission's
   AEC/RECON
c. State University of New York's
   BCN (Biomedical Communication Network)
d. Mead Data Central's
   LEXIS

2.3.2 As of April 23, 1973 bibliographic data bases containing more than one hundred thousand records were available on the parent system for commercial searching.

a. ORBIT
b. DIALOG
c. RECON
d. BASIS*
e. LEADER**

a. ERIC (Educational Resources Information
   Center's data bases)
   CHEMON (Chemical Abstract's
   Chemical Condensates)
   MEDLINE (National Library of Medicine's
   MEDLARS)
b. ERIC
   PANDEX (CCM Information Services' data base)
   NTIS (National Technical Information Service's
   U.S. Government Research and Development
   Reports)
c. TOXICON (National Library of Medicine's
   Toxicology data base)
   ENVIRON (Environmental Protection Agency
   data base)

d. NTIS
Chemical Abstract's Condensates
* Science Information Associates [illegible]
marketing of searching service.

e. Chemical Abstract's Condensates
COMPENDEX (IEEE's Engineering Index)
** Metascience handled marketing of searching service.

2.3.3 In addition to the sites mentioned in 2.3.2, system software was also being used at the following locations.

a. ORBIT
b. RECON
c. STAIRS
d. DATA CENTRAL
e. NASIS

a. Department of State
Karolinska Institute (Stockholm, Sweden)
State University of New York (Syracuse)
b. Department of Justice (JURIS)
c. House of Representatives (Bill Status System)
d. Environmental Protection Agency
Wright-Paterson Air Force Base
Union Carbide
e. Carnegie-Mellon Univers.

2.3.4 In addition to the data bases mentioned in 2.3.2, large data bases were available on the parent system for in-house use.

a. STAIRS
b. SPIRES
c. NASIS
d. INTREX

a. COMPENDEX
IBM's Technical Documents
b. MARC (Library of Congress's Machine Readable Cataloging data base)
c. ERTS (NASA's Earth Resources Satellite imagery data base)
d. the INTREX data base

2.3.5 As of April 23, 1973 the system was being used for some data bases that were neither bibliographic nor primarily textual.

RECON
DATA CENTRAL
BASIS
SPIRES
NASIS
RIQS

2.3.6 The system has capabilities equivalent to those of a host language. This means that programs written in COBOL or some other language can call upon the retrieval and display features.

DATA CENTRAL
NASIS

2.3.7 The system is operated in conjunction with a batch retrieval system. Only in the case of NASIS is it possible at the terminal to transfer a request from the online system to the batch system.

STAIRS
DIALOG
BASIS
LEADER
RIQS
NASIS
RECON

2.3.8 The major technique used by the system for translating requests into lists of records is to call upon previously constructed inverted indexes. Items in the index are kept in alphabetical order, and each item is a word, numeric value, or phrase followed by a list of pointers to records in the data base.

ALL
SYSTEMS
EXCEPT
RIQS

2.3.9 Pointers may indicate 1) the data base from which the record comes, 2) the field from which the item comes, 3) the word position within the field, 4) the item's importance to the record, 5) a date for the record, or 6) other information. Those systems that do not store data base and/or field information within the pointer keep separate inverted indexes for each searchable data base and/or field.

a. INTREX
b. DIALOG
c. STAIRS
d. SPIRES
e. LEADER
f. ORBIT
g. RECON
h. DATA CENTRAL
i. BASIS

a. 2,3,4 and 6
b. 1,2,3,4,5,6
c. 2,3, and 6
d. 5 and 6
e. 4 and 6
f. 4
g. 1,2,3, and 5
h. 2 and 3
i. 6 (for links)

2.3.10 An additional technique used by the system for translating requests into lists of records is to call upon every record in the current list (or in the data base if there is not a current list) and scan it to see whether it satisfied the query.

RIQS
INTREX
STAIRS
ORBIT
NASIS

2.3.11 As of April 23, 1973 the system was accessible through the Tymshare network (a commercial, leased telephone line, data communication network).

DIALOG
RECON
ORBIT
BASIS

2.3.12 Whether or not on the Tymshare network, a number of users were accessing the system from remote locations.

a. IBM has its own international data communications network
b. the federal telephone network (WATS)
c. WATS and privately leased links
d. NSF-sponsored leased line networks
e. Lockheed has its own high speed (480 cps) leased line network

a. STAIRS
b. NASIS
c. DATA CENTRAL
d. SPIRES
e. DIALOG

2.3.13 The user can search using 1) any of a wide range of different ASCII video and teletype terminals (like the Hazeltyne 2000 or the Texas Instruments Silent 700), 2) EBCDIC video and typewriter terminals (like the IBM 2260, IBM 3270 or IBM 2741, 3) graphics terminals (like the IMLAC or ARDS terminals).

a. 1 and 2
b. 2 and 3
c. 1,2 and 3
d. 1

a. ORBIT
a. DIALOG
a. RECON
a. STAIRS
a. DATA CENTRAL
a. SPIRES
a. NASIS
b. INTREX
c. BASIS
c. RIQS
d. LEADER

2.3.14 The system runs on a) the medium to large size IBM 360 and 370 series computers (like the IBM 360/40 or 370/145), b) the IBM 360/67, c) the CDC 6400, or d) the UNIVAC 1108. (INTREX ran on an IBM 7094.)

a. DATA CENTRAL
a. RECON
a. DIALOG
a. ORBIT
a. STAIRS
b. NASIS
b. SPIRES
c. RIQS
c. BASIS
c. LEADER
d. BASIS

2.3.15 The software is written primarily in a) FORTRAN, b) PL/1, c) Assembly Language, (SPIRES is written in PL/360 which acts like Assembly language but looks like PL/1. INTREX was written in an extension of Algol.)

a. BASIS
a. RIQS
a. LEADER
b. ORBIT
b. NASIS
c. DIALOG
c. RECON
c. STAIRS
c. DATA CENTRAL

3.0 INSTRUCTIONAL, DIAGNOSTIC, AND CONTROL FEATURES

3.1 INTRODUCTION

Before one can search for information using an interactive retrieval system, one must have access to the system and be able to use it. It is tempting to focus upon command language capabilities, forgetting that people need instruction, make mistakes, and occasionally need retraining. In this section we shall focus upon supportive and controlling features, so that in later chapters we can disregard them. A number of the features are actually not a part of the command language interface, but are necessary for it to function properly. For example, logging onto the computer will be discussed, as well as the training of users by human instructors and trouble-shooting by consultants when command language problems arise.

There is a danger when dealing with computer systems of assuming that everything should be computerized. Many system designers reported that the tutorial material available at the terminal is rarely invoked. There are at least three possible reasons for this. Interactive searching is still quite expensive (15-120 dollars/hr) and users may feel that there are cheaper ways to learn. It could also be that the range of behaviors to be taught or difficulties to be anticipated is so great that enough tutorial material can not be presented within a reasonable amount of time. Third, it is possible that the tutorial material available has not been reworked to the point where it makes sense to users. Whatever the reasons, all of the eleven systems depend heavily upon human assistance when dealing with users.

Few designers still argue that anyone can come in off the street and after fifteen minutes of instruction be searching like a professional. There are too many things that need to be taught. If the user is not familiar with his terminal, he has to be taught how to operate it. This may include plugging telephone receivers into modems, setting duplex switches, transmitting commands, and backspacing. Not only will the user have to learn how the retrieval language works, he probably will also have to learn how to search. Most non-mathematical users initially find boolean operators to be unnatural. It is only after some experience that they realize how important it is to fill out concepts and to progressively combine them. However, the topic usually requiring the most instruction is how to use the data base. Data bases differ markedly in structure, ranging from Medline with its MESH hierarchical thesaurus to ERIC with multiple index terms for the same topic, to Compendex with its index phrases, to still others for which every word in the text is inverted for searching. Some data bases rely heavily upon roles, others upon linkages, and some rely upon relations. For searching to be really effective, the query must be adapted to the peculiarities of the data base. Most searchers today are information specialists who, through constant practice, are comfortable with the terminal, command language, and data bases.

3.2 BECOMING FAMILIAR WITH THE SYSTEM

There are at least three types of people that receive instruction regarding system usage. At the highest level are purchasers of the system software or of storage space for a data base to be accessed via the system. These people generally receive extensive personalized instruction and will not be mentioned again. At the next level are individuals or institutions for whom a search [illegible] is to be [illegible] there are individuals who search because they know someone who has access to a system. Commercial system representatives are most familiar with intermediate level people [illegible] opening an account, they make certain that the account holder or his representative receives instruction in system [illegible] system representatives are generally more familiar with the lowest level users since system representatives and users are in close physical proximity. In both cases, it rarely happens that a person with an information need who has never searched using an interactive retrieval system before, sits down at a terminal [illegible] himself and begins searching. Almost always there is someone he goes to who knows how to use the system, and who will help him as he learns how to search.

[illegible] The system customarily is invoked without going through special procedures known only to those who have been cleared for searching.

a. INTREX
b. SPIRES
b. BASIS

a. people generally go to a termainal that was logged on ahead of time by the INTREX staff.
b. people having account numbers for computer use in general can invoke the system by following normal logon procedures and typing SPIRES or BASIS.

[illegible] ers are cleared for searching, they are customarily required to attend a short course [illegible]

ORBIT
DIALOG
STAIRS
NASIS
RECON
DATA CENTRAL
SPIRES

[illegible]

[illegible]
INTREX
[illegible]
BASIS

3.2.4 As a part of their training, users are strongly encouraged to read the users' guide.

ORBIT
DIALOG
RECON
BASIS
SPIRES
NASIS
RIOS
DATA CENTRAL

3.2.5 Users are provided with pocket-sized folders (3"x8") summarizing the available commands for the system, and, in at least two cases (*), the names of fields in one of the data bases.

STAIRS
DATA CENTRAL
SPIRES
INTREX*
DIALOG*

3.2.6 Even though the user has properly invoked the system, an attempt is made to catch the untrained user and provide him with instruction until he overrides it.

a. After the user types his name and address, the system responds "Welcome to Intrex _____. If you already know how to use Intrex, you may go ahead and type in commands. (Remember, each command ends in a carriage return.) Otherwise, for information on how to make simple searches of the catalog, type
info 2
or, to see the Table of Contents (Part I) of the Intrex Guide which will direct you to other parts of the Guide explaining how to make more detailed searches, type
info 1"

b. After the user types SPIRES, the system responds
"WELCOME TO SPIRES II, _________.
followed by messages of the day, then
IF IN TROUBLE TYPE HELP"

c. After the user is logged in, he is told which data base he is connected to. Then he is asked whether he wishes to have the new or experienced user format. If he responds that he wants the new user format, he is advised
"TYPE ONLY AFTER THE CUE 'USER:', WAIT AFTER THE CUE 'PROG:', ENTER SEARCH STATEMENTS ... DO YOU WISH A FURTHER SUMMARY OF OPERATING FEATURES? TYPE YES OR NO AND CARRIAGE RETURN."

a. INTREX
b. SPIRES
c. ORBIT
d. BASIS
e. DATA CENTRAL
e. LEADER
f. STAIRS
g. RIOS
h. DIALOG

d. After the user logs on he is asked
DO YOU DESIRE OPERATING INSTRUCTIONS?
TYPE YES OR NO
e. From the moment the user logs onto the system, he is advised regarding what to do next. If he has received training, he is able to enter commands not included in the advice.
f. After signing onto a data base, the user is informed that he can invoke the HELP function.
g. If accessed from RIQSTUTOR, the RIQS user is informed
"IF AT ANY TIME WHILE ENTERING SEARCH COMMANDS YOU NEED ASSISTANCE, TYPE HELP..."
h. As a part of the log on greeting, the user is given a telephone number for problems or assistance.

3.2.7 The knowledgeable person can receive enough instruction at the terminal in order to learn how to search. This may include 1) terminal and typing problems, 2) the command repertoire, 3) characteristics of the data base, 4) common pitfalls and their remedies, 5) recent revisions to the system, 6) sample searches to be emulated, or 7) hints for effective searching.

a. DATA CENTRAL
b. SPIRES
c. RIQS
d. INTREX
e. ORBIT
f. STAIRS
g. DIALOG

a. depending upon the data base 1,2,3,4,6 and 7
b. 2,3,5,6
c. 2,3,4,5,6,7
d. 1,2,3,5,6,7
e. 1,2,3,5
f. 2,3,6
g. 2,3,5

3.2.8 The user either can access training material passively via a tutorial or can actively use an index to call up pages of instruction.

a. DATA CENTRAL
b. SPIRES
c. RIQS
d. INTREX
e. ORBIT
f. STAIRS

a. WHAT# allows one to access pages of text. Many of these pages suggest other pages to access. There is a directory for finding out what text is associated with each number.
b. HELP instructs the novice to type EXPLAIN EVERYTHING (a directory of what is available). EXPLAIN EVERYTHING indicates that the user seeking instruction should type EXPLAIN SPIRES for tutorials. SHOW NEWS [illegible] the latest changes in the system.

c. The user must attach RIQSTUTOR, which in turn attaches and communicates with RIQSONLINE. During the tutorial, control can be passed to RIQS via the EXIT command. Control can be returned to the tutor via the HELP command.
d. INFO# allows one to access pages of text. INFO 2 acts as a directory. NEWS allows one to find the latest changes in the system.
e. The new user receives tutorial information. EXPLAIN EXPLAIN acts as a directory.
f. HELP allows one to access pages of tutorial explaining different aspects of the system.

3.3 RECEIVING HELP WHEN IN TROUBLE

It is a rare person who never makes a mistake. The range of problems that can arise when searching is enormous. At best the user realizes he has forgotten how to enter a command. Often he cannot figure out what a message from the computer means, but is satisfied when he receives a more verbose explanation. Problems become more difficult when the user is certain that one thing should have happened but instead another did. For example, the user asks for all records dealing with AMERICAN HISTORY OR UNITED STATES HISTORY and finds that the data base contains nothing. The problem could be that ORs are processed before ANDs, that the data base has been broken into individual words, and that implicit ANDs are inserted between words in a query. Thus the request effectively asks for records containing AMERICAN AND STATES AND HISTORY. Or the problem might be that the user returned the cursor before transmitting the line. Or the problem might be that the data base distinguishes between upper case and lower case letters, and that the upper case query did not retrieve any of the mixed case documents. While the above is hypothetical, it points out the range of causes there may be for a single problem.

Most of the systems do very little on-line to help users with their problems. Almost all of them provide users with telephone numbers to call as a last resort. They rely heavily upon good training in the first place, and interfaces that are so straightforward that users find it difficult to make errors.

3.3.1 The user who has switched to a terse form of interaction with the system, can switch back to a verbose form whenever he feels he needs reminding.

a. INTREX
b. ORBIT
c. DATA CENTRAL

a. LONG to get the verbose mode,
SHORT to get back to the terse mode.
b. VERSION LONG to get the verbose mode,
VERSION SHORT to get the intermediate,
VERSION SYMBOLIC to get the terse mode.
c. When entering the data base, the option
LONG or LKWIC is used to get the verbose
mode, while SHORT or SKWIC is used to get
the terse mode.

3.3.2 The user can ask for an explanation of the system prompt or response he has just received.

a. RECON
b. ORBIT
c. RIQS
d. SPIRES
e. STAIRS
f. DIALOG
g. NASIS
h. DATA CENTRAL

a. HELP error code
b. EXPLAIN will give an explanation of
the latest prompt by the computer.
EXPLAIN followed by the terse form of a
system message will retrieve an
explanation of the message.
HELP will cause a menu of options to be
displayed
(no postings?, see results?, need
to answer a question?, review features?,
review data base?, command names?)
c. If the RIOSTUTOR is attached, then HELP
invokes the tutor and it explains how the
RIQS command component most recently used
works.
d. EXPLAIN error code
HELP will explain to the user the stage
of searching he is currently in,
and will suggest what he might do
to proceed to the next stage.
e. ..HELP error code or ..HELP option
..HELP without a parameter prompts
the user with a list of things
for which he can receive explanations.

f. EXPLAIN error code

g. EXPLAIN error code

EXPLAIN RESPONSE, error code in order to find out what to do to correct an error.

EXPLAIN ORIGIN, error code in order to find out where an error code came from.

h. WHAT#

every prompt and response from the system is numbered so that fuller explanations can be retrieved.

3.3.3 The user can ask for an explanation of a specific command for which he remembers the name but probably not the format.

a. RECON
b. ORBIT
b. DIALOG
b. SPIRES
b. NASIS
c. STAIRS
d. RIQS
e. DATA CENTRAL

a. HELP command name
b. EXPLAIN command name
c. ..HELP command name
d. if the RIQSTUTOR is attached, then the user can ask for explanations of commands at any time during query formulation.
e. DATA CENTRAL has explanations of the specific commands in the WHAT file.

3.3.4 The user can have the processing of his search request traced so that he can get a better feeling for what went wrong.

a. STAIRS
b. INTREX

a. ..SET DETAIL=ON causes the number of records and the number of occurrences for every term in the query to be listed, but does not show how the terms get combined.
b. COUNT causes the number of records remaining to be listed after each new term's stem is ANDed into the result.

3.3.5 A portion of the user's guide is written in such a way that it helps the user with trouble-shooting.

BASIS
STAIRS
RECON
SPIRES
ORBIT
DATA CENTRAL
DIALOG

3.3.6	Possessors of account numbers are provided with a telephone number for calling up a human consultant who can help them with trouble-shooting.	ALL SYSTEMS
3.3.7	The user can interact in real time with the consultant via the command language	a. NASIS b. SPIRES c. DIALOG

a. HELP followed by text prints the message at the consultant's terminal.

b. [illegible]O SPIRES takes the message composed using the text editor and puts it in the consultant's mailbox. Assuming that the user did not turn off monitoring, the consultant can look over the user's recent behavior from his consulting terminal.

c. SEND MESSAGE 1/ followed by text prints a message at the consultant's terminal, who then can send a message back or call the user.

3.[illegible] REGULATING USAGE

There are a few features that users often do not perceive as beneficial, but that ultimately are for their benefit. Some features, like access restrictions, may rarely be brought to the attention of users. Others, like encouraging comments, may be highly visible. The features are included because they make it possible for system representatives to learn from the experiences of users and to regulate who has access to their systems.

3.4.1 When entering the system, the user must pass through the logon procedures of 1) a computer network, 2) an operating system, and 3) the retrieval system itself. He must know 4) an account number, 5) a terminal identifying code, and 6) a password.

a. 3 and 4; sometimes also 1,5, and 6
b. 2,6, and 3; sometimes also 1 and 5
c. 3 and 4; sometimes also 1 and 5
d. 3 and 4
e. 2,4,6 and 3; sometimes also 1,5,and 6
f. 2,4,5,6 and 3
g. 2,6 and 3
h. 2,4,6, and 3
i. 2,3,6, and sometimes 5
j. 3,4, and sometimes 6

a. ORBIT
b. BASIS
c. DIALOG
d. DATA CENTRAL
e. RECON
f. SPIRES
g. LEADER
h. RIQS
i. STAIRS
j. NASIS

3.4.2 By the time the user has logged on, his searching abilities have 1) been restricted to a subset of all possible data bases, 2) a subset of the fields in a data base's records, and 3) a subset of the records in the data base.

a. 1
b. 1 and 2
c. 1, 2, and occasionally 3

a. RIQS
a. BASIS
a. DIALOG
a. ORBIT
a. RECON
a. LEADER
b. DATA CENTRAL
c. NASIS
c. STAIRS
c. SPIRES

3.4.3 By the time the user has logged on, his command behavior is being monitored (disregarding purely charge monitoring) a) only as a contribution to aggregate statistics, b) on a command in core basis rarely used for tracking individual users, c) on an individual user basis so that users can both be studied and helped with problems, or d) on an individual basis but with an option for the user to turn monitoring off.

a. NASIS
b. ORBIT
b. RECON
b. DIALOG
c. LEADER
c. INTREX
d. SPIRES
d. RIQS

3.4.4 The user can enter comments as he searches that appear at the appropriate spot in the monitor log.

a. ORBIT
b. LEADER
c. INTREX
d. RIQS

a. "COMMENT followed by t xt"
b. COMMENT prompts the user for lines of text until the user specifies STOP. Or
   * followed by text for notes to oneself.
c. COMMENT followed by text
d. The user is asked after completing use of RIQS to enter any suggestions or comments he might have.

4.0 QUERY FORMULATION FEATURES

4.1 INTRODUCTION

Perhaps the major value of information retrieval systems is that one can easily gain access to small numbers of potentially relevant records that are buried in immense data bases. By trying out various criteria, by observing how the criteria pare away a data base, and by examining a scattering of retrieved records, the interactive searcher can revise his strategy until he senses that he can do no better. While a good searcher vacillates between expressing criteria and checking to see how well the criteria work, in this chapter we shall only discuss features for formulating queries.

Formulating queries for information retrieval systems is not the same as asking everyday questions. First of all, the context in which most everyday questions are asked greatly limits the possible responses. One does not ask a waitress how much it costs to fly to Tokyo. Secondly, most questions incorporate an understanding of how the respondent structures his knowledge. Thirdly, askers of difficult questions rarely expect thorough or precisely on-point answers. Users of information retrieval systems often do expect to receive thorough and on-point answers. They expect to get these from data bases that contain material contributed by many different people who structure their knowledge in different ways. Computers are not in eye-to-eye contact with humans and generally do not remember people from one session to the next. The information retrieval systems discussed in this report do not answer questions; they provide a "quick and dirty" means for narrowing down the search space so that the odds are improved that the user will find an answer to his question.

Nevertheless, formulating se... ...

... questions, ... and ...

... must be ...

... the

... up of

... inter-

... out the

... explain why

... to respond.

... features ... the reasons for having them will be considered

... alternate ways for implementing them. While an effort is made

... features without tying them to particular storage schemes or data

... the discussion is made concrete. The similarity across

... a common understanding of human expression but from

... storing data and an interest in bibliographic citations.

... stated before, one of the reasons for this analysis is to assess how sen-

... command language features are to current technologies. Where insen-

..., perhaps functional needs can be formally recognized, and treated by

... in a common manner

4.2 DATA BASE SELECTION

... the seven systems, one must select a data base before search-

... The word "data base" is used to refer not only to some ...

... structured records, but also to auxiliary files containing

..., or other material supporting the primary file.

... "data base" selected ...

... the physically-stored version, there are exceptions ... (ENIRAD) where the user-perceived data base may be a ... version. ... of the systems allow the user at his ... together ... data base out of phys... distinct ... interested in a topic spanning ... separately. Only one system (LEADER) ... the search request alone when switching to a new data base. ... allow requests to be stored and ... other data bas... In general, the more controlled ... for a data base, the less likely queries composed ... in it will transfer to other data bases. Even where a single ... against multiple data bases, the user must combine the ... hand. It remains for future studies to show how often ... to access more than one data base, and how bothersome it is that ... results are not merged together automatically.

The features discussed below are those for selecting an initial data base and then for switching to a different one. Tutorial and other help ... features are discussed in the previous chapter. Our conclusion there was that designers assume users do not need help with data base selection ... advice is provided regarding subject coverage, indexing policies, ... search techniques. If systems are intended for use by those with ... needs (e.g., scientists, doctors, lawyers, etc.), perhaps at ... introductions should be provided for each of the available ...

4.2.1 The set of available data bases does not depend upon the user identification. INTREX*

* INTREX makes available only a single data base. For use with other data bases, the system is called other names.

4.2.2 When the user identification happens to limit one to a single data base, that data base is attached automatically. INTREX ORBIT RECON DIALOG

4.2.3 When the user identification does not limit one to a single data base, one data base is still attached automatically. RECON ORBIT DIALOG*

* DIALOG has a user-specified default data base.

4.2.4 The user is automatically presented with a list of the data bases he might attach. NASIS LEADER

4.2.5 The user can ask for a list of the data bases he might attach.

a. BEGIN or EXPLAIN FILES
b. SHOW SUBFILES
c. ..HELP BASES
d. FILES or RETRIEVE (where there is no default file name)
e. "FILES?"
f. BROWSE when using the RIQSTUTOR
g. WHAT 50

a. DIALOG
b. SPIRES
c. STAIRS
d. NASIS
e. ORBIT
f. RIQS
g. DATA CENTRAL

4.2.6 As a part of the available data base display, the user is asked to enter the identifying number of the one he wants to have attached. LEADER DIALOG NASIS RIQS*

* when using the RIQSTUTOR

4.2.7 The user, independently of an available data base display, indicates to the system which data base he wishes to have attached as his first data base.

a. DATA CENTRAL
b. RIQS
c. BASIS
d. STAIRS
e. SPIRES
f. DIALOG
g. NASIS

	SYSTEM PROMPT	RESPONSE
a.	ENTER FILE, MESSAGE OPTION	db, option
b.		ATTACH, logical unit, db
c.	ENTER NAME OF THE DATA BASE TO BE SEARCHED	db
d.	ENTER DATA BASE NAME	db password (password is optional)
e.	-	SELECT db
f.	ENTER:	BEGIN BYPASS (for the user-specified default data base)
g.	-ENTER NASIS COMMAND:	RETRIEVE db

The following four items assume that the user wishes to switch to a new data base.

4.2.8 The user must first exit from the system, although he may still maintain contact with the time-sharing monitor.

INTREX
RIQS

4.2.9 The user indicates that he wants to be returned to the point just before he was asked to select his first data base.

a. STAIRS
b. LEADER
c. BASIS

a. ..CHANGE
b. CHANGE
c. RESTART

4.2.10 The user indicates to the system which data base to attach just as he did initially.

SPIRES
NASIS

4.2.11 A slightly different technique is used for switching to a new data base than was used initially.

a. BEGIN for a list of the attachable data bases and a prompt to select one, or BEGIN db# to bypass the listing and prompting
b. .FILE db#
c. FILE for a reminder of the current data base (and option), with a chance to switch by typing db, option
d. "FILE db"

a. ELHON
b. DIALOG
c. DATA CENTRAL
d. ORBIT

4.3 FORMULATING SIMPLE QUERIES

Many people prefer to learn by doing, and often begin searching before they know how to. All of the eleven systems rely upon the user to formulate the query, but vary in how easy they make matters for the uninitiated. Since people often rely heavily upon first impressions, it is important to consider what each system presumes that the user knows about entering requests, how successfully the system re-orients him when he gets confused, and how rapidly it brings him to the point where he can display retrieved records. Since simplicity is gained at the expense of making presumptions, it is worthwhile to examine what assumptions each system makes, and how systems differ.

It is difficult to anticipate what an untrained user will consider to be a good request. Much depends upon the type of data base selected. In order to remove this variable, and since all of the eleven systems can be used with textual or bibliographic data bases, we shall assume that the user has selected a textual data base. According to designers, the great majority of requests are for content-related records. The untrained user is likely to rely upon prior experience with content-related searching. If he thinks about how he asks people questions, he probably will sketch out most aspects of his query, using good grammar and more words than absolutely necessary.

If he thinks about how he uses indexes or card catalogues, he probably will enter one or two words that capture a single aspect of the query. Instructors of interactive searching typically have to counteract the effects of each of the analogies. The question-asker is encouraged to enter one concept at a time so that he can discover whether clarification is needed [illegible] the concept. The index-thumber is encouraged to enter multiple concepts, connecting them together with boolean operators. Without instruction, the first individual may never retrieve anything, while the other may end up manually searching through voluminous output. The sample queries we will use are "the effects of television violence on children" for question-askers, and "television violence" for index-thumbers.

It should be pointed out that DIALOG and RECON attempt to discourage users from directly entering full queries, and instead encourage them to scan alphabetically adjacent index terms. They do this by having a row of keys on the terminal act like functions keys, and by having the one for EXPAND left of the one for SELECT. The untrained user who begins his request with EXPAND receives back a portion of the index, thereby increasing chances that he will not enter wordy requests in the future.

The reason for not discussing learning by doing in the chapter on instruction is that none of the systems treat simple searching as an opportunity for introducing the user to more sophisticated features. Many users never progress beyond simple searching. Only in the case of the LEADER system is simple searching likely to deal effectively with quite complex requests. LEADER attains this power by using the request as a device for transferring the searcher's attention to index phrase selection.

[illegible] The system assumes the user wants to enter a search request after having selected the data base.

a. ENTER REQUEST
b. SS 1/C? -- SEARCH STATEMENT 1 OR COMMAND?
c. ENTER YOUR SEARCH ONE TERM AT A TIME
   1/
d. PLEASE ENTER YOUR REQUEST. (END IT WITH A PERIOD)
e. TO FIND DOCUMENTS IN THE SYSTEM SPECIFY YOUR
   SEARCH REQUEST BY SUBJECT, AUTHOR, TITLE TERMS,
   OR COMBINATIONS OF THESE AS SHOWN IN THE 3
   EXAMPLES BELOW:
      SUBJECT XENON VISCOSITY (etc.)
f. ENTER SEARCH COMMAND OR TYPE HALT

a. DATA CENTRAL
b. ORBIT
c. BASIS
d. LEADER
e. INTREX
f. RIOS

4.3. The user must indicate to the system that he intends to enter a search request.

a. SELECT
b. ..SEARCH
c. FIND

a. NASIS
a. RECON
a. DIALOG
b. STAIRS
c. SPIRES

4.3.1 The user must specify the field to be searched

a. SUBJECT
b. an index for the selected data base
c. RECORD (to search all fields)

a. INTREX
b. SPIRES
c. RIOS

4.3.2 The user must specify the logical relationship that the [illegible] value has in common with the field in question

a. CONTAINS

a. RI[illegible]

[illegible] The user must specify the relationship holding between multiple words in the query value.

a. 'query words'

* This is true only for data bases not having multi-word index terms

a. RIOS
a. RECON*

4.3.6 The user must specify in the query what is to be done with the results of searching. — a. RIQS

a. DISPLAY field names or numbers or
PLACE RECORDS IN SET set number

4.3.7 The end of the query must be designated in some manner (other than by a carriage return or transmit). — a. RIQS; b. LEADER

a. END
b. .

> *For the next seven items, the words making up the query take a form like THE EFFECTS OF TELEVISION VIOLENCE ON CHILDREN, where no record contains exactly this phrase nor does the user expect one to.*

4.3.8 By expressing a request in this manner, the user is likely to receive worthwhile results. — a. INTREX; b. LEADER

a. While AND is the implied boolean connector, the effects of stemming and full indexing reduce the chances that the user will receive a null result.
b. The words are used to retrieve indexing phrases ranked from those containing all the words to those containing only one. The user then selects some of the phrases and these become his request.

4.3.9 The user will receive a null result with no suggestion for entering worthwhile requests. — DATA CENTRAL; DIALOG; RECON; NASIS; SPIRES*; ORBIT; RIQS

* for data bases where the implied boolean connector is an AND

4.3.10 The user will receive a null request but with a suggestion that improves his chances of entering worthwhile requests. — a. BASIS

a. NO SUCH TERM. WANT ADJACENT TERMS?
   YES:NO/

4.3.11 The user will receive a result, but neither is it very useful nor is he given suggestions for entering worthwhile requests. — STAIRS, SPIRES*

* for data bases where the implied boolean connector is an OR.

4.3.12 Common words like 'the', 'of', and 'in' are deleted from the query. — INTREX, LEADER, DATA CENTRAL*, SPIRES*, STAIRS*

* for those data bases where the data base manager has chosen this option

4.3.13 Words like 'effects' (but not 'children') are depluralized. — LEADER, INTREX, DATA CENTRAL*

* the data base manager specifies which depluralization rules he wants to incorporate

4.3.14 Words like 'television', 'violence', and 'children' have their affixes removed so that they become stems like 'televi', 'viol', and 'childr'. — INTREX

For the next five items, the words making up the query take a form like TELEVISION VIOLENCE where the user does not realize there is a difference between the exact phrase and the intersection of the two terms.

4.3.15 The request is interpreted as a search for the term TELEVISION VIOLENCE in the index.

BASIS
ORBIT*
DIALOG
RECON**
NASIS*

* for data bases not having multiple word index terms, the request will lead to a null result

** true only for data bases having multiple word index terms

4.3.16 The request is interpreted as a search for phrases like TELEVISION AND VIOLENCE, TELEVISION VIOLENCE, TELEVISION OF VIOLENCE in the record.

DATA CENTRAL

4.3.17 The request is interpreted as a search for the exact phrase TELEVISION VIOLENCE in the record

RIOS
RECON*

* true only for data bases with word proximity searching

4.3.18 The request is interpreted as a search for documents indexed under (or containing) both the words TELEVISION and VIOLENCE.

SPIRES*
INTREX

* true only for data bases with AND as the implied boolean connector

4.3.19 The request is interpreted as a search for documents indexed (or containing) either the word TELEVISION or VIOLENCE.

SPIRES*
STAIRS

* true only for data bases with OR as the implied boolean connector

4.3.20 The request is interpreted as a search for documents indexed (or containing) either TELEVISION or VIOLENCE or both (whether or not the terms are contiguous). The results are ranked.

LEADER
STAIRS*

* when using the RANK command

4.4 EXPRESSING A SINGLE CONCEPT

The experienced searcher, unlike the novice, recognizes the importance of breaking queries into distinct concepts, and of polishing each concept until the right balance between precision and ambiguity has been achieved. In natural language communication, it is rarely necessary to focus upon isolated concepts or to worry about how they are expressed. The rich grammatical conventions that we use in natural language for embedding concepts are not available in command languages. By focusing upon concepts, logical bonds linking concepts become clearer, and so can be more easily reconstructed with boolean operators in a query. Concept linking will be discussed in the next section. The rich store of redundant expressions and interconnected ideas that recipients of messages possess is not built into most data bases. By attempting to thoroughly express each concept, new interpretations of the query come to light and shades of meaning stand out. In this section, features for refining concepts are discussed.

The most prevalent features are those for incorporating additional terms into the request. New terms can range from slight variations in spelling to semantic relatives. There are a number of reasons why so many features should be dedicated to incorporating new terms. Users, like most people, find recognition easier than recall. Alphabetical displays and displays of related terms stimulate the user with slight variants for his terms that he might otherwise forget to include. Bad spellers can use alphabetical displays in the same way they use dictionaries. Features that simplify entering large numbers of almost identical terms not only reduce the chances for error but also speed up the query formulation process.

Features for restricting the scope of terms are less prevalent. Most of the systems do not distinguish between upper and lower case letters in a query. Most avoid confusions about whether characters are part of a term or part of a command by eliminating special characters from index terms. (The techniques used by SPIRES and NASIS for handling special characters are not discussed in this report.) Many permit the user to search for exact phrases. A few make it possible for the user to limit how far apart words can appear in a retrieved document. All systems let the users restrict searching to one or more fields. Perhaps the restrictive power of the boolean AND (discussed in the next section) when used to combine different concepts is so great that the need for limiting features at the concept level is reduced.

The following thirteen items facilitate expanding the scope of concepts.

4.4.1 Various formats for expressing certain types of values can be recognized and automatically converted to a standard form before searching.

a. RIQS
a. DATA CENTRAL
b. SPIRES

a. 3 FT. 2 IN. converts into 38 IN.
   JUNE 16, 1973 converts into 06/16/73
b. FRED JONES converts into JONES, F.
   JUNE 1973 converts into 06/--/73.

4.4.2 The user can specify that values falling between a lower and upper bound should be incorporated into the query. An asterisk implies sequential searching. DATA CENTRAL and RIQS achieve this feature using relational operators and the AND logical operator.

a. NASIS
b. RECON
b. DIALOG
c. BASIS
d. ORBIT
e. SPIRES
f. STAIRS

a. TERM1 - TERM2 or TERM1: TERM2
b. TERM1: TERM2
   LIMIT set/## -## years only
c. (TERM1 - TERM2) numeric only
d. 19## THRU 19## years only
e. BETWEEN TERM1 and TERM2 numeric
   FROM TERM1 TO TERM2 only
*f. WL TERM1, TERM2 (within limits)
   OL TERM1, TERM2 (outside limits)

4.4.3 The user can specify criteria for numeric fields via relational operators. For all but RECON, the format is FIELD NAME OPERATOR TERM. For RECON the format is OPERATOR FIELD NAME TERM. STAIRS and RIQS are limited to sequential searching of numeric fields.

a. NASIS
b. RECON
c. SPIRES
d. RIQS
e. STAIRS
f. DATA CENTRAL

	>	≥	=	≠	≤	<
a.	>,GT	>=,GE,¬<	=,EQ,BT,><	¬=,NE	<=,LE,¬>	<,LT
b.	+	+=	=		−=	−
c.	>,AFTER	>=	=	¬=	<=	<,BEFORE
d.	GT	GE	EQ	NE	LE	LT
e.	GT	NL	EQ	NE	NG	LT
f.	>,GTR	NLS,/<	A=,AEQ	A/=,ANEQ	NGT,/>	<,LSS

4.4.4 Common prestored variants for a query term are automatically incorporated into the query (i.e. AEC also retrieves documents indexed under ATOMIC ENERGY COMMISSION).

STAIRS
DATA CENTRAL
BASIS

4.4.5 The user can specify that terms with the same root as the query term should be incorporated into the query (i.e. CHILD# acts like CHILD, CHILD'S, or CHILDREN).

a. DATA CENTRAL
b. STAIRS
c. SPIRES
c. DIALOG
d. LEADER
e. ORBIT
f. RECON
g. INTREX

a. TERM**** incorporates terms having (in this example) four or fewer characters following the root.
b. TERM $. TERM $n, MASK TERM, or MASK TERM n incorporate terms having an unlimited number or no more than n characters following the root.
c. TERM# for data bases defined to permit truncation (a different symbol could be designed as the truncation code in the data base definition).
d. TERM*
e. ALL TERM#
f. TERM:
g. TERM! (since truncation is the normal mode, this specifies an exact match on all characters)

... that terms matching the query ... be incorporated into the query.

a. DATA CENTRAL
b. SPIRES
c. DIALOG
c. ORBIT
d. ORBIT

... ON*LINE retrieves ON-LINE and ON LINE but not ONLINE
... ON LINE retrieves all of the above via something like sequential searching
ON#LINE retrieves ON-LINE and ON LINE but not ONLINE
... RINGSEARCH (FN) ON:LINE retrieves all of the above via sequential searching

The user can ask for a display of terms alphabetically adjacent to the one he designates. Possible features of the display are 1) that two or more of the displayed terms precede the term alphabetically, 2) that the terms are given identification so they can be referred to later, 3) that each term indicates how many records would be retrieved if it were entered in a search request, 4) that each term indicates how many terms are cross indexed with it, and 5) that the term causing the display is highlighted or designated in some manner. (Parenthesis ... FN means that specifying a field name or index ... optional.)

a. STAIRS
b. ORBIT
c. DIALOG
c. RECON
d. NASIS
e. BASIS
f. SPIRES
g. DATA CENTRAL

COMMAND	RESPONSE FEATURES
ROOT TERM	3
NBR TERM (FN	1,3
TERMS	2
EXPAND ...	1,2,3,4,5
EXPAND ...	1,2,3...
TERM*	[illegible]
[illegible]	

[illegible] The user [illegible] a display of alpha- betically adjacent terms can have additional [illegible] subsequent terms displayed.

	PROMPT	FORWARD	BACKWARD	STOP
a.	UP N OR DOWN N	DOWN N	UP N	NONE
b.	CONTINUE PRINTING?	YES		NO
c.		PAGE +	PAGE -	
d.	MORE	YES, or F	B	NO
e.		+KEYWORD	-KEYWORD	

a. ORBIT (NBR)
b. ORBIT (L)
c. DIALOG
c. RECON
c. NAGIS
c. DATA CENTRAL
d. SPIRES
e. STAIRS

4.4.9 The user can ask for a display of terms semantically related (broader, narrower, synonyms, see also, statistically associated) to the one he designates. Possible features of a display are 1) that codes are used to designate the type of relation holding between the display term and the one entered, 2) that the terms are given identifiers so that they can be referred to later, 3) that each term indicates how many records would be retrieved if it were entered in a search request, and 4) that each term indicates how many terms are cross-referenced with it.

a. ORBIT
b. DIALOG
b. RECON
c. SPIRES
d. LEADER

	COMMAND	RESPONSE FEATURES
a.	"TREE TERM"	both one level above and two [illegible]
b.	EXPAND E# or R	1,2,3,4
c.	THESARUS index TERM	all terms one level below [illegible]
	SYNONYM index TERM	all synonyms at this le[illegible]
d.	ASSOCIATION [illegible]	2,3 for phrases statistic[illegible]
	ASSOCIATE [illegible]	[illegible] associated with [illegible]

4.4.10	The user, when incorporating either alphabetically, or semantically related terms into his query, 1) can ask for ALL of them, 2) can string them together with commas, 3) can use a hyphen to specify a range of them, and 4) can incorporate them directly into a boolean expression. a. 2,3,4 b. 2,3 c. 2,3 d. 1,2 e. 1,4	a. NASIS b. RECON c. DIALOG d. LEADER e. ORBIT
4.4.11	The user is automatically told when he enters a search term which other terms are related to it.	BASIS
4.4.12	The user can specify that all terms hierarchically below the designated one are to be incorporated into his response. a. EXPLODE TERM b. available in some data bases by automatic index enhancements	a. ORBIT b. DIALOG
[illegible]	[illegible]	a. LEADER b. STAIRS

4.4.14 The user can specify the case characteristics (all capital letters, little letters) that terms in the records must satisfy.

term .X. where X can be L/C for lower case, U/C for upper case, F/C for first letter capitalized, and M/C for mixed upper and lower case.

STAIRS

4.4.15 The user can specify how central the terms in a query must be to the content of a record.

a. RANGE X where X can be MAJOR, SECONDARY, MINOR, or TOOL.
b. LIMIT X1/X2 where X1 is the query set or sets being qualified and X2 is either MAJ or MIN.
SELECT TERM/FN*

a. INTREX
b. DIALOG

Often data bases star central terms (tungsten*) thus alleviating the need for a special feature at query time, but requiring the user's item-by-item post-selection of results.

4.4.16 The user can specify that words are to appear one before the other in the text of a record. An asterisk implies sequential searching.

*a. STRINGSEARCH (FN) : TERM1 TERM2
b. TERM1 (W) TERM2
c. 'TERM1 TERM2'
d. TERM1 ADJ TERM2
e. TERM1 TERM2
*f. 'TERM1 TERM2'
g. TERM1 - TERM2
*h. RESTR[illegible] FIELD NA[illegible]

a. ORBIT
b. DIALOG
c. RECON
d. STAIRS
e. DATA CENTRAL
f. RIOS
g. INTREX

4.4.17 The user can specify that words are to appear within n words of each other in the text of a record. A - means that the second word must appear before the first, and a + means that the second word must appear after the first

a. DATA CENTRAL
b. DIALOG
c. RECON

	+	-	+ -
a.	TERM1 (WPn) TERM2	(WMn)	(Wn)
b.	TERM1 (nW) TERM2		
c.	TERM1+*n TERM2	-*n	*n

4.4.18 The user can specify that words are to appear within some logical grouping of a record.

a. DIALOG
b. STAIRS
c. INTREX
d. ORBIT

a. TERM1(F)TERM2 in same field.
   TERM1(nS)TERM2 within n sentences of each other or (if n is omitted) in the same sentence.
   TERM1(L)TERM2 in same index linkage
b. TERM1 WITH TERM2 in same sentence
   TERM1 SAME TERM2 in same paragraph
c. WITH TERM2 in same index phrase
d. SENSEARCH (FN) :TERM1 TERM2: in same sentence

4.4.19 The user can specify that the terms are to appear within a particular field (or index). FN means that a field name contains only two characters.

a. INTREX
b. RIQS
c. NASIS
d. SPIRES
e. BASIS
f. DATA CENTRAL
g. STAIRS
h. RECON
i. DIALOG
j. ORBIT

a. SUBJECT TERM
   AUTHOR TERM
   TITLE TERM
b. IF #FIELD1, #FIELD2, #FIELD3 CONTAIN. . .
c. search request, FIELD = FIELD NAME
   FIELD NAME relational operator TERM
d. FIELD NAME TERM
   FIELD NAME relational operator TERM
e. FN TERM FIELD NAME, TERM
f. $FIELD NAME relational operator TERM
g. TERM.FIELD1, FIELD2, FIELD3.
h. FIELD NAME = TERM
i. FN = TERM (indexed fields)
   TERM/FN1,FN2 [illegible]
   [illegible]

a. DATA CENTRAL
b. DIALOG
c. ORBIT
d. INTREX
e. RIQS
f. STAIRS

4.4.. The user can specify that a particular record is to be retrieved, without its being immediately displayed.

a. $NR EQU RECORD#, RECORD#. . .
b. KEEP RECORD#
c. RECORD#
d. DOCUMENT RECORD#
e. BEGIN SEARCH OF RECORD RECORD#
f. ..SELECT ALL RECORD EQ #

[illegible] The OR is [illegible] together, while the AND [illegible] the [illegible] DATA GENERAL [illegible]

[illegible]

results [illegible] ranked beginning with those matching on the greatest number of phrases. We can cut off the display of records as the percentage meeting the relevance criteria gets [illegible]. Unfortunately, there is no method for breaking a request into parts that can be refined separately and later integrated.

Two areas where systems might be improved involve search profiles and sequential searching. Search profiles are often thought of as standing requests for incoming records. It is probably more economical to carry out selective dissemination of information (SDI) via batch rather than interactive searching. However, if profiles are thought of as highly polished request components for use by people searching in inadequately indexed areas, then it makes sense to have them available interactively. In many ways this use parallels publication of scholarly research and demands the same type of peer review. Not only would the task of searching be made easier, but users would effectively be indexing material in their own discipline. Sequential searching also has typically been thought of in terms of batch systems. However, when it is limited to a small set of results, it can be carried out rapidly [illegible] reasonable cost. Many fields in a record are used infrequently and do not warrant the cost of indexing. Yet without a sequential search [illegible] these fields cannot be used during searching at all.

The next four items deal with Boolean operators.

[illegible] The user can [illegible] portions of his request with Boolean operators (connectors).

a. DATA CENTRAL
b. NASIS
c. ORBIT
c. INTREX
d. RECON
e. STAIRS
f. SPIRES
g. RIQS
h. BASIS
i. DIALOG

	AND	OR	AND NOT	OR NOT
a.	AND,&	OR	AND NEG AND /=	OR NEG OR /=
[illegible]	*,&	+,|	¬,¬	
[illegible]	AND	OR	AND NOT	
[illegible]	*,&,AND	+,|,OR	¬,¬,AND NOT,NOT	
[illegible]	AND	OR,XOR	NOT	
[illegible]	AND,&	OR,|	AND NOT,&¬	
[illegible]	AND	OR	AND NOT	OR NOT
[illegible]	AND,.AND.	OR,.OR.	AND NOT .AND. .NOT	
[illegible]	*,AND	+,OR	-,NOT	

[illegible] For systems permitting mixed expressions, connectors with high priority are processed before connectors with low priority, and connectors with equal priority are processed from left to right. The features discussed in the previous section either cannot be [illegible] with Boolean operators or else have higher priority. The one exception is DATA CENTRAL's word [illegible] operator which has lower priority than OR [illegible] has been included below.

a. DATA CENTRAL
b. NASIS
c. ORBIT
d. DIALOG
d. RECON
e. STAIRS
f. SPIRES
g. RIQS
h. BASIS
i. INTREX

	HIGH PRIORITY		LOW PRIORITY
[illegible]	[illegible]	[illegible]	+,WP.WM.AND
	*,&,¬,¬	+,|	
	AN .AND N [illegible]		
[illegible]	*,-,AND,[illegible]	+,[illegible]	
	A[illegible],N[illegible]	[illegible]	
[illegible]	[illegible]		
	[illegible]	[illegible]	
	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	
	[illegible]		

4.5.3	The priority of connectors can be controlled through the use of parentheses.	NASIS DIALOG RECON STAIRS SPIRES RIQS BASIS
4.5.4	A Boolean operator can be specified for insertion between terms in a range of terms. a. set# - set#/operator b. TERM1-TERM2/operator operator/TERM1:TERM2 c. TERM1 operator TERM2 TERM3	a. DIALOG a. RECON b. NASIS c. STAIRS

> *The next eleven items deal with request sets.*

4.5.5	Query input from the user is treated as a complete request. It is possible 1) to store the results of a request and 2) to invoke those results as the domain (or part of the domain) for a later query. a. 1) IF query THEN PLACE RECORD IN SET # 2) BEGIN SEARCH OF SET # BEGIN SEARCH OF UNION (#,#,...,#) BEGIN SEARCH OF INTERSECTION (#,#,...,#) BEGIN SEARCH OF COMPLEMENT (#,#,...,#)	a. RIQS
4.5.6	Query input from the user either begins a new request or else continues the current one. While it is possible to back up and take another tack (for INTREX [illegible]	a. SPIRES b. DATA CENTRAL c. LEADER d. INTREX

	FIRST PART	CONTINUATION	BACKTRACK
a.	FIND query	AND or OR query	BACKUP (one step)
b.	query	AND or OR query	MODIFY # (to step #)
c.	query leads to phrase selection	PHRASES for phrase selection revision	START OVER for query revision while maintaining selected phrases NEWSEARCH for total query revision
d.	FIND NAME query	AND or OR query	SET2 (to bring back SET2)

4.5.7 Query input from the user is treated as a complete request and is assigned a set number (except for INTREX). The results of a request can be incorporated into subsequent requests by mentioning the set number (or set name for INTREX).

ORBIT
DIALOG
RECON
STAIRS
BASIS
NASIS
INTREX*

* result sets, or backup points, are established via the command NAME SET1 where SET1 is to be the name of the set.

4.5.8 The only queries that can contain connectors are those where the only terms are set numbers.

a. DIALOG
b. BASIS

a. COMBINE query
b. (query)

4.5.9 The user can review the text of his query (or queries).

a. "DIAGRAM#" The text of the query for set number # (or for the most current set if # is omitted) is displayed, but with all set numbers replaced by query text (enclosed in parentheses).
b. DISPLAY SET HISTORY For each active set the text of the query (with index phrases replacing EXPAND display line numbers) is listed along with the number of records the query retrieved.
c. [illegible]

a. ORBIT
b. DIALOG
c. RECON
d. STAIRS
e. DATA CENTRAL
f. BASIS
g. LEADER
h. NASIS

d. ..DISPLAY ALL or ..DISPLAY #1,#2 For each set in the range specified, the text of the query is listed along with the number of records the query retrieved.
e. REQUEST or REQUEST #, ANSWER # or ANSWER. For the latest addition to the query or for the designated modification step, either the query text or else the number of retrieved records is displayed.
f. /LIST ALL. Same effect as for STAIRS.
g. PHRASES The index phrases selected so far are listed along with the number of records each retrieves.
h. SETS. Same effect as for DIALOG.

4.5.10 The final set is used for accumulating individual records as they are being displayed.

a. DIALOG
b. RECON

a. KEEP record# or
KEEP set#/item#, item#-item#, ...

4.5.11 Before submitting any queries, the entire data base acts like result set 0, and can be displayed.

STAIRS
NASIS

4.5.12 Result sets can be purged so as to free space for additional sets.

a. STAIRS
b. RIQS
b. RECON
c. ORBIT
d. INTREX

a. PURGE ALL or PURGE #, or PURGE #,#
b. RELEASE SET # or RELEASE SETS #,#,...,#
c. ERASEALL or ERASEBACK # (# and subsequent sets are wiped out) or RESTACK n,...,m TO # (# and subsequent sets are wiped out except for n,...m which are given new set #s beginning at #. If the TO # part is omitted set 1 is assumed.
d. DELETE SET, SET [illegible]

4.5.13 [illegible] session.

a. RIQS
b. INTREX

[illegible]

4.5.14 The total request can be saved for re-use during some other session.

a. STAIRS
b. NASIS
c. SPIRES
d. DIALOG

	SAVE	USE	DELETE
a.	..SAVE NAME	..EXEC NAME	..PURGE NAME
b.	PROFILE	RERUN NAME	
c.		EXECUTE strategy in working memory	
d.	END/SAVE a serial # is assigned	RECALL #	.RELEASE #

4.5.15 Certain commands (usually including the AND NOT connector) can be used only in conjunction with a request set (or ongoing request).

a. STAIRS
b. INTREX
c. ORBIT

a. ..SELECT the system puts out a set number and the user responds with a prior set number (or ALL) followed by query input. Numeric and other non-indexed fields can only be searched in this manner.
b. RESTRICT FIELD NAME TEXT.
c. STRINGSEARCH set# query

5.0 RESULT MANIPULATION FEATURES

5.1 INTRODUCTION

The fact that records in a data base can easily be manipulated by a computer means that records can take on different appearances depending upon the needs of the user. During query formulation, aspects of records can be presented that help the user revise his request and measure his progress. After a query has been formulated, other aspects can be presented that help the user locate answers to his questions. Occasionally, aggregate properties of records are more important than individual records. In this chapter we shall discuss features for interactively displaying records during and after the query negotiation process.

The result manipulation features available in the eleven systems are not as extensive as those needed in management information systems. Ideally, data bases intended for decision making should reflect problem-oriented elements and relationships. It should be possible for the user to organize elementary components so that they reveal strategies for dealing with the problem. Most data bases of the eleven systems reflect a record orientation rather than a problem orientation. Records entering the data base remain identifiable, and generally are the same records that are retrieved later on. While NASIS, DATA CENTRAL, and SPIRES can synthesize retrieved records out of stored record constituents, they are not commonly used in this way.

For data bases to be developed that effectively address problems, both searchers and data base contributors must agree upon problem conceptualizations. There must be some guarantee regarding the quality of queries and retrieved records. While for some RIQS and SPIRES data bases the contri[illegible]

575

ENTRAL

the

enerally, the ... this

searching

user

results, and what should be done

display is much simpler than

user probably does not feel any need

results can be

user

times

a. INTREX
b. DATA CENTRAL
c. LEADER

catalog

documents will

the results of a request are not [illegible] immediately. The user is asked
[illegible] WANT THE DISPLAY REPORTS LISTED

The user will [illegible] a single record [illegible] page of [illegible] multipage record and then will be asked to indicate if he wishes to continue

LEADER
RECON (DISPLAY)
DIALOG
BASIS
DATA CENTRAL*

(YES-NO), NEXT [illegible]
STOP OR TERMS
b. PAGE
* for [illegible] only

The user will receive a display [illegible] and then will be asked if he wishes to continue

CONTINUE PRINTING? (YES/NO)
The user is warned that the output will be lengthy and may select how many documents he wants to see before being interrupted.

The user will receive [illegible] display [illegible] that [illegible] be interrupted [illegible] interrupted automatically [illegible] terminal controller.

SPIRES
INTREX
RIQS
RECON (STIR)
STAIRS
DATA CENTRAL*

[illegible]
[illegible]LATO
[illegible]ATA [illegible]RAL
LEADER
[illegible]
[illegible]AS[illegible]
[illegible]

[illegible]

LEADER

DATA

AIR*

DATA CENTRAL

DISPLAY set , format
DISPLAY set format

DATA CENTRAL
INTREX
DIALOG
RECON
STAIRS

After the prompt
ENTER DESIRED OUTPUT, ... FUL-RET, CONCISE
b. DISPLAY set
OUTPUT ALL
DISPLAY set 5
e. DISPLAY set 2
BROWSE ALL
IF query THEN DISPLAY RECORDS
DISPLAY set causes the prompt
WHAT FIELDS DO YOU WANT TO SEE
The response is ALL
PRINT FULL

DATA CENTRAL*
INTREX
STAIRS
SPIRES
RIQS*

[illegible]

carriage return.

[illegible] LINES [illegible]

MARGIN [illegible] margin size

LENGTH for text width

BLANK for skipping [illegible]

[illegible]

[illegible]

[illegible] STAFF [illegible]
[illegible] LEADER [illegible]
[illegible]
[illegible]
[illegible]

[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
after selecting the [illegible]
[illegible] tasks which [illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]

[illegible]

a. SPIRES
b. RECON
c. DIALOG
d. INTREX

[illegible]

a. DATA CENTRAL
b. NASIS
c. DIALOG
d. RECON
e. BASIS
f. INTREX
g. ORBIT
h. STAIRS
i. SPIRES
j. PIRS
k. LEADER

[illegible] PRINT with the same [illegible] as the [illegible]

[illegible] the [illegible]

[illegible] field name list

[illegible] means print [illegible]

[illegible] the same [illegible]

[illegible] in the [illegible]

[illegible] where PRINT without [illegible]

[illegible] to the [illegible]

[illegible]

	a. BASIS
	b. ORBIT
	c. RECON
	c. DIALOG
	d. NASIS
	e. INTREX

. after requesting printing, the user [illegible] asked

PLEASE TYPE COMPLETE NAME AND
MAILING ADDRESS

. after requesting [illegible] the user [illegible]
is asked
NAME AND [illegible]
ADDRESS
CITY, STATE (PLEASE [illegible] ABBREVIATE [illegible])
[illegible] NAME, IF SAME
TITLE, PHONE
and then is asked if the above is OK

. before ever beginning to search, the user is asked
SEARCH TITLE
NAME OF PERSON CONDUCTING SEARCH
NAME OF PERSON RECEIVING RESULTS
MAIL ADDRESS

. before ever beginning to search
ENTER YOUR NAME:
ENTER YOUR ADDRESS

. the user specifies [illegible]

[illegible]

	a. NASIS
	. INTREX
	. LEADER
	. RIQS
	e. ORBIT
	. DIALOG
	. RECON

[illegible]

The following three items are clusters of display features that have been incorporated into some interactive retrieval systems.

The user [illegible] that [illegible] of the retrieved documents are to be displayed at the terminal. INTREX BASIS DIALOG

The user [illegible] that [illegible] to be generated from the [illegible] and be plotted at the terminal. BASIS RIQS

The user [illegible] that [illegible] are to be passed through statistical analysis and that the results of analysis are to be displayed at the terminal. RIQS DIALOG BASIS

[illegible] CONCLUSIONS

[illegible] INTRODUCTION

[illegible] the report has been descriptive rather than normative. It has [illegible] left to the reader to choose between alternate implementations of equivalent features. A more normative approach will be taken in this chapter. At the April 197[illegible] workshop, designers were asked to specify which features were in some sense "minimal." What this came to mean was that, without a feature, searching would be significantly constrained. The minimal features will be summarized in the next section along with supportive reasoning.

The introduction of supportive reasoning brings up another issue. While students of design can profit from hindsight, they also need principles that lead them into good design. The new capabilities of a rapidly evolving computer technology too often are the only guideposts available. A possible set of user-oriented principles will be presented in the third section along with illustrative examples drawn from the report. The principles will not be [illegible] a well-developed body of man-computer problem-[illegible] existence but not yet been created [illegible] not yet too early to begin deciding what we want to build an information infrastructure and for whom it is intended.

6.1 MINIMAL FEATURES

Twelve of the features listed in the appendix were [illegible] designers at the April 1973 workshop to be of central importance to interactive searching. Two of the features relate to instruction and assistance, seven relate to formulating queries, and three relate to manipulating the retrieved records.

The two features felt essential for instruction and assistance are live help and complete, readable users' guides. That designers consider human and written assistance to be essential is a clear indication that at present these traditional media are more necessary to searching than is assistance via the computerized medium. One should not conclude, however, that on-line documentation and assistance are unimportant. Manuals and [illegible] derive much of their value from the fact that they can be turned to when all else fails. Computers have been used for instruction and [illegible] only a very short time. Authors, instructors, consultants, [illegible] are far more comfortable with traditional media [illegible] the new ones. It is not surprising that mistakes are [illegible] and users of on-line materials, or that people [illegible] traditional channels even though the computerized [illegible] materials are developed for supporting users [illegible] computer literacy, [illegible] essential for effective searching.

[illegible] formulation [illegible]

[illegible] rators, search field control, relational operators, and suffix removal. Of all the features, boolean operators is probably the most essential feature. Without ANDs, ORs, and NOTs it would be difficult to formulate complex requests. Of the eleven systems only LEADER does not provide for them. The indexing phrases [illegible] the LEADER user are assumed to be additive; retrieved documents are ranked in accordance with the additive overlap between request and document. Ranking accomplishes most of what normal users expect from a system, but definitely limits what can be expressed in a query. There is no way to indicate to LEADER that documents containing extraneous concepts should be demoted in the ranking, or that documents containing central concepts should be promoted. There is no way to indicate that two or more concepts are not independent. For example, different phrases may stand for a single concept, but this cannot be specified. The net effect is that without boolean operators the skilled searcher cannot retrieve exactly what he wants.

Control over which fields to examine when matching requests with stored records [illegible] essential. Like people, records tend to have multiple [illegible] can be treated in just one way. By establishing [illegible] that each field contains only a single type of information, it [illegible] possible to compare records along unitary dimensions. [illegible] to take advantage of the various fields, [illegible] in the request. He can not only specify [illegible] to interrogate, but he can usually [illegible] field. All of the systems except [illegible] he wants to interrogate [illegible] index phrases. While th[illegible]

[illegible]

[illegible]

LARGE TEXTUAL DATA BASES
2.3.2, 2.3.4

MANAGEMENT INFORMATION
2.3.5, 2.3.6

FULL TEXT SEARCHING

INDEX SEARCHING

INTERMEDIARY SEARCHER

END USER SEARCHER

VIDEO TERMINAL

DEDICATED TERMINAL

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	INSTRUCTIONAL, DIAGNOSTIC, AND CONTROL FEATURES
			B	B	B	B	C	A	A	A	[illegible] GUIDE (A*)
			V	V		V	V		V	V	SYSTEM 3.2.6
						V	V	V	V	V	CLASS 3.2.2
V	V			V	V	V			V		PERSONAL 3.2.3
	V	V			V	V		V	V	V	READING 3.2.4
V	V					V	V			V	ON-LINE TRAINING 3.2.7, 3.2.8
			V			V	V		V	V	DATA BASE OVERVIEW 3.2.7
						V	V				SAMPLE SEARCHES 3.2.7
	V									V	ON-LINE DOCUMENTATION 3.3.2, 3.3.3
											SEARCH LOGIC TRACING 3.3.4
											[illegible] HELP* 3.3.6
											[illegible]
											[illegible]

MODES OF INSTRUCTION (SYSTEM, CLASS, PERSONAL, READING)

[illegible]

[illegible] in simple terms explain [illegible] and [illegible] the [illegible] are available [illegible] the [illegible] are accessible and how they are structured, [illegible] search strategy, and (f) what to do when [illegible]

[illegible] plete [illegible] readable

-- [illegible] mplete but readable or

[illegible] plete but not readable

[illegible] complete nor readable

[illegible]

[illegible] first-time user.

[illegible] in the

[illegible]

-- The beginning user [illegible] falls into this material.

-- The beginning user easily falls into [illegible] material. When in verbose mode, [illegible] responses cue the user.

-- The beginning user easily falls [illegible] material. He can go through a simple [illegible] [illegible] with explanations.

-- The knowledgeable user can [illegible] to the retrieval system. The [illegible] unique [illegible] command [illegible]

-- When in verbose mode, system [illegible] the user.

-- When in verbose [illegible] responses [illegible] user.

[illegible]

On-line Documentation

One page descriptions of all commands and error messages.

Search Logic Tracing

The searcher can request a detailed description of how his multi-part search request lead to the number of hits reported.

Live Help

Either a telephone number to call if desperate or a message command for requesting help from the on-line human consultant.

Vest Pocket Card

A durable folder containing command names and an explanation of how to get complete command descriptions.

Comments

The user can express his feelings or contribute insights to the system staff. Most useful in conjunction with a system log so that the staff can examine the user's pre-comment behavior.

Monitor Log

Typically a list of the commands entered by the user during his session.

INTREX -- Able to have two terminals see and control a single display (as well as traditional monitoring)

NASIS -- Individual sessions are not logged. Instead, every week command and data base field usage is summarized.

QUERY FORMULATION FEATURES	INTREX	RIQS	NASIS	SPIRES	LEADER	BASIS	DATA C	STAIRS	RECON	DIALOG	ORBIT
SUFFIX REMOVAL* 4.3.13-14, 4.4.5	x			v	v,x		v,x	v	v	v	v
SEARCH FIELD CONTROL*		v	v			v	x	x	v	x	x
DICTIONARY ACCESS* 4.4.7, 4.4.8			v	v		v	v	v	v	v	v
RELATIONAL OPERATORS* 4.4.3		v	v	v			v	v	v		
SPELLING VARIATIONS 4.4.4						v,x	x	x	v	v	v
RELATED TERM CAPABILITY 4.4.9				v	v,x	v			v	v	v
WORD PROXIMITY OPERATORS 4.3.16-17, 4.4.16-18	v	x					v,x	v	v,x	v	v
BOOLEAN OPERATORS* 4.5.1-4	2	1	1	2		3	3	1	1	1	1
REQUEST SETS* 4.5.7, 4.5.13	v	v	x			x		x	x	x	x
PHRASE DECOMPOSITION 4.3.12	x			x	x		x	x			
SEARCH PROFILES 4.5.14			v	v		½		v	½	v	½
SEQUENTIAL SEARCHING 4.5.15	v	v	v			v		v			v

* = an essential feature v = under user control

½ = currently being implemented x = assumed automatically

Suffix Removal

Most systems require the user to indicate suffix removal by entering the root followed by a truncation code. Automatic stemming in INTREX can be overriden by following the term with an exclamation point. DATA CENTRAL automatically deletes final s's.

Search Field Control

If the user does not specify which field he is searching, an X means all fields are assumed while a V means a single field (or a combined field like title and abstract) is assumed. The user can override the default by specifying a field name.

Dictionary Access

Alphabetic portions of the controlled vocabulary, index phrases, or word dictionary are displayed. Only DATA CENTRAL and SPIRES do not specify the number of documents associated with each term.

Relational Operators

The operators GREATER THAN, LESS THAN, and BETWEEN can be used with numeric fields. In STAIRS, the operators can only occur when performing a sequential search.

Spelling Variations

Words like 'avenue' and 'ave.', or 'January' and 'Jan.' are linked so that the use of any one incorporates the others.

Related Term Capability

Where terms in the vocabulary of a data base point to other terms, the user can see the other terms. LEADER has both automatic phrases and optional associated phrases. DIALOG, RECON, and LEADER number the terms so they can be incorporated by number.

Word Proximity Operators

DIALOG, STAIRS, and RIQS allow the user to specify that the terms are to occur in the same field without specifying the exact field. DIALOG, RECON, and DATA CENTRAL allow the user to specify how many words may separate two terms. All systems listed in the word proximity column permit testing for exact phrase matching. ORBIT and SPIRES provide for phrase matching via sequential search.

Boolean Operators

Generally OR is used to combine related terms, AND is used to intersect distinct concepts, and AND NOT is used to exclude previously displayed material. A 1 indicates that AND's are processed before OR's. A 2 indicates that the leftmost operator is processed first. A 3 indicates that CR's are processed before AND's, but, by using variants of the operators, AND's can be processed before OR's.

Request Sets

Each line of search request input is given a set number. Later search requests can incorporate earlier sets by mentioning the set number.

Phrase Decomposition

A natural language phrase is decomposed into significant words. Common words like 'the' and 'about' are deleted. INTREX assumes a logical AND between terms. SPIRES assumes either AND or OR depending upon the file definition. LEADER ranks retrieved phrases avoiding the problem.

Search Profiles

A user can develop a search strategy (or components of a search strategy), store the strategy, and rerun it many days later. DIALOG permits users to develop and store search strategies so that they can automatically be run against file updates.

Sequential Searching

In order to search fields that have not been inverted, each record in the set is examined one after the other. This is the only method for searching in RIQS.

RESULT MANIPULATION FEATURES

	INTREX	RIQS	NASIS	SPIRES	LEADER	BASIS	DATAC	STAIRS	RECON	DIALOG	ORBIT
SEARCH REVIEW* 4.5.9			v		v	v	v	v	v	v	v
PREDEFINED FORMATS* 5.2.6, 5.3.1	v		v	v	v		v		v	v	v
ON-LINE FORMATTING* 5.2.7, 5.3.2-5	1	4	4	1		2	2	1	3	2	1
RAPID SCAN 5.2.5	v	v		v			v	v	v	v	
HIGHLIGHTING 5.2.9, 5.3.9	v				v		v	v	v		
EXPANDING 5.3.10, 5.3.11	v		v		v	v	v	v	v	v	v
SORTING 5.3.6		½	½	½		½	v	v	v	v	½
RANKING 5.3.7					v			v			v
COMPUTING 5.3.8		v	v	v		v	v				
MICROFICHE 5.3.16	v					v				v	
DISPLAY OF GRAPHS 5.3.17		v				v					
STATISTICAL INTERFACE 5.3.18		v				v				v	
OFF-LINE PRINTING* 5.3.13-15	v	v	v	v	v	v	v	v	v	v	v
BATCH RETRIEVAL 2.3.7		v	v		v	v		v	v	v	v
PHOTOCOMPOSITION							v	v		v	
DATA ACCESS PROTECTION 3.4.2			v	v		½		v			½

* = an essential feature v = has been implemented

½ = currently being implemented

Search Review

In order to review what has been done, a short summary is presented of the sets currently active, the number of documents each contains, and the request that caused the set to be created.

Predefined Formats

To allow the user to easily vary the way in which he views retrieved documents. Usually one format displays descriptive fields, another short citations, and a third displays total documents. With management information the predefined formats are likely to specify page layouts.

On-line Formatting

To allow the user to indicate the fields of each document he wishes to see. A 1 indicates that the fields can be put out in one order; a 2 indicates that they can be put out in any order; a 3 indicates that limited page organization is possible; a 4 indicates full scale on-line report formatting.

Rapid Scan

To quickly look over a little information about each document retrieved. The display should continue until the user interrupts it. DIALOG and LEADER simulate rapid scan by stacking commands. RIQS provides for nothing but rapid scan.

Highlighting

To find out what it was in the document that caused it to be retrieved. The INTREX feature does not highlight text but rather puts out only those field values that caused retrieval.

Expanding

To have a complete document displayed after seeing a portion of it. This is usually done either by entering a sequence number or document number.

Sorting

To sort the documents according to the values of a designated field or fields.

Ranking

To rank the documents so that the ones containing the greatest number of search terms are listed first. STAIRS provides for five different ranking options. LEADER automatically ranks output.

Computing

BASIS and RIQS allow the user to define variables as functions of field values. The calculated values may be displayed or operated upon by statistical subroutines. In SPIRES, the average command combines calculating the value and displaying the result.

Microfiche

The terminal controls access and display of the microfiche. In DIALOG the feature is not used.

Display of Graphs

Graphics terminals are used by RIQS and BASIS for on-line plotting of fields from the document set retrieved. In neither case is it possible to superimpose graphs from different sets.

Statistical Interface

The records retrieved by searching can be passed to an on-line statistical package for analysis.

Off-line Printing

The documents are printed and (usually) are ma[illegible] to the user. The user must divert the request to the batch system with both RIQS and SPIRES.

Batch Retrieval

There is communication from the on-line system to the batch system with both NASIS and LEADER. In the other cases, the batch retrieval command language is similar to the on-line language.

Photocomposition

In close association with the retrieval system is an off-line photocomposition system for publishing data base documents.

Data Access Protection

Access to records within the data base or fields within records can be restricted to a subset of all users.