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 preview, pre-defined formats, on-line formatting, and off-line printing. (SK)
a report of the
INSTITUTE FOR COMMUNICATION RESEARCH
STANFORD UNIVERSITY
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 Scientific Information Service
National Science Foundation
Washington, D.C. 20550

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

Prepared by: Paul B. Parker, Ph.D.

The opinions and conclusions are those of the author and do not necessarily represent the views of the National Science Foundation.
The detailed 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, getting help when in trouble, regulating usage, selecting a data base, formulating complex 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 search, statutory operators, dictionary access, request sets, search review, on-line formats, on-line formatting, and off-line printing. It is concluded that a clear distinction exists between management information and bibliographic retrieval. We refer it is intended for use by designers of interactive retrieval systems and by system designers.

# Information retrieval
# Man-machine systems
# Design
# Information systems
# Data storage systems
# Psychometric techniques
# User expectations
TABLE OF CONTENTS

1. Introduction
   1.1 The Intended Audience ..................................... 1
   1.2 The Feature Analysis ....................................... 2
   1.3 How to Read the Report .................................... 5
   1.4 Related Work ............................................... 9

2. The Systems Selected
   2.1 Introduction ................................................ 11
   2.2 Historical Perspective ..................................... 13
   2.3 The System's Environment .................................. 18

3. Instructional, Diagnostic, and Control Features
   3.1 Introduction ................................................ 24
   3.2 Becoming Familiar with the System ......................... 25
   3.3 Receiving Help when in Trouble ............................ 29
   3.4 Regulating Usage ........................................... 32

4. Query Formulation Features
   4.1 Introduction ................................................ 35
   4.2 Data Base Selection ........................................ 36
   4.3 Formulating Simple Queries ................................. 40
   4.4 Expressing a Single Concept ................................ 46
   4.5 Interconnecting Concepts ................................... 56

5. Result Manipulation Features
   5.1 Introduction ................................................ 63
   5.2 Displaying Results Simply .................................. 64
   5.3 Controlling the Display ..................................... 67

6. Conclusions
   6.1 Introduction ................................................ 75
   6.2 Minimal Features ............................................ 76
   6.3 Principles for User-Oriented System Design ............... 81
   6.4 A Future for Interactive Retrieval ......................... 83

7. Appendix ....................................................... 86
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 interacting with different systems. At the very least, it should reduce some of the pain designers feel when building new systems in ignorance of strategies that others have employed.
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. Second, 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.

7.2 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 improvement unless the efforts of designers remain consistent.
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 STC/UOI (the user on-line interaction special interest group of the American Society for Information Science).

In preparation 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 their client base. 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 revisions 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 perspective 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 conjunction 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 sponsored 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 participants 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 languages 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 distinction 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. Raiff 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 Design of Man-Computer Dialogues (8) is one of the only textbooks. The Annual Review of Information Science and Technology 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 for Information Science periodically holds technical sessions dealing with user interface problems.


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, LEAIS, etc.) 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, 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 (CCIC/LAW of Canada, IF 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 Time system would have been included except that it required a terminal that was not available on the West coast, was used with a single data base, and was written at IBM Federal Systems Division so would not have been represented by a system designer. Lawrence Livermore Laboratories' MASTERC
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 PIBETS, Syracuse's SUPERS, 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. It seems 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.
HISTORICAL PERSPECTIVE

Systems were selected to participate in the comparative analysis as of January, 1973. System representatives came together April 24-25, 1973 for the comparative analysis workshops. 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 must 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.

1.1.1 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.

Since April, 1973 the system has also been installed in Iran, the largest bibliographic data base has been revised, a new data base has been developed, functions have been extended for interactive modeling and data analysis, and the system has been used for transportation and energy planning of the...
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 RODON) 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.

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 overlay both alphanumeric characters and microfiche.

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 1972. 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.

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.

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.
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.

Since April, '73 the system has been revised and is currently being reprogrammed to run on the DEC PDP-11 series of computers.

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.

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.
ORBII II was developed by Syster Development Corporation in conjunction with the development of FLMILL 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. CDC both sells or leases software and also puts up data bases for subscription searching on the parent system.

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.

RIOS 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 RIOS TUTOR 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.

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.

Since April, 1973 the system 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 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. 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. ERIC (Educational Resources Information Center's data bases)
      CHEM ON (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)
      ENVIRONMENT (Environmental Protection Agency data base)
2.3.3 In addition to the sites mentioned in 2.3.2, system software was also being used at the following locations.

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

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. 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.
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.

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.

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.

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. 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 database if there is not a current list) and scan it to see whether it satisfied the query.

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

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

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

2.3.13 The user can search using 1) any of a wide range of different ASCII video and teletype terminals (like the Hazeltynne 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 IMIAC or ARDS terminals).

- 1 and 2
- 2 and 3
- 1, 2 and 3
- 1
2.3.14 The system runs on a) the medium to large size IBM 360 and 370 series computer. (like the IBM 360/40 or 370/145), b) the IBM 360/67, c) the CDC 6400, or d) the UNIVAC 1100. (INTREX ran on an IBM 7094.)

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.)
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 troubleshooting 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 FRIC 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.

3.6 BECOMING FAMILIAR WITH THE SYSTEM

There are at least three types of people that receive instruction requiring system design. At the highest level are purchasers of the system, software, or storage space for a database to be incorporated in the system. These people generally receive extensive personal or instruction which will not be continual. At the level of individual or institutions
for whom a search is to be one... if there are individuals who search because they know someone who has access to a system. General system representatives are more familiar with intermediate level people of the system or account, they make certain that the system or the representative involves instruction in system use. General 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 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 start.

3.1.1 The system automatically is invoked without going through special procedures known only to those who have been feared for searching.

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

If users are feared for searching, they are automatically required to attend a short course.
3.2.5 Users are provided with pocket-sized folders (3"x8") summarizing the available commands for the system, in at least two cases (*), the names of fields in one of the data bases.

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 OPERATIONAL FEATURES? TYPE YES OR NO AND CARRIAGE RETURN."
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 revi-
   sions to the system, 6) sample searches to be emulated, or
   7) hints for effective searching.

   a. depending upon the data base 1,2,3,4,6 and
   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
   a. access training material passively
   b. as a tutorial or can actively use an index to recall
   c. types of instruction.

   1. HELP 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.

   2. HELP instructs the user 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 tells
      the latest changes in the system.
The user must attach RIOSTUTOR, which in turn attaches and communicates with RIOSONLINE. During the tutorial, control can be passed to RIOS via the EXIT command. Control can be returned to the tutor via the HELP command.

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.

The new user receives tutorial information. EXPLAIN EXPLAIN acts as a directory. 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 least 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, so that the upper case query did not retrieve any of the mixed case responses. While these examples are hypothetical, it points out the range of causes there can be for a single problem.
Most of the systems provide 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 straightforward that users find it difficult to make errors.

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

- long to get the verbose mode,
- short to get back to the terse mode.
- VERSION LONG to get the verbose mode,
- VERSION SHORT to get the intermediate,
- VERSION SYMBOLIC to get the terse mode.
- When entering the data base, the option long or LONG is used to get the verbose mode, while short or SHORT 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.

- HELP error code
- 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
  (search results?, see results?, need to answer a question?, review features?, review data base?, command names?)
- If the MONITOR is attached, then HELP invokes the tutor and it explains how the user command is executed most recently used.
- EXPLAIN error code
  - will explain why the user thinks he is currently in,
  - will suggest what he might do
  - will proceed to the next stage.
- HELP error code or HELP option
  - will, without a parameter prompts the user with a list of things for which he can receive explanations.
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. 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. ..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.
1.3.6 Possessors of account numbers are provided with a telephone number for calling up a human consultant who can help them with trouble-shooting.

1.4.7 The user can interact in real time with the consultant via the command language:

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

b. TO 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.

7. 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 origination controls, 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

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

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

a. "COMMENT followed by text"

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 databases. By trying out various criteria, by observing how the criteria pare away a database, 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 databases 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.
Coherent or inconsistent? Just what is the

inherent dimension, and what is the

nature of the information of interest? Inherent organizing and hierarchical

interactions of the various features in the data, one cannot 

out the

intellectual work that is involved in making the

fulfillments where there are similar statements. I shall explain why

1. tacit understanding is made.

The same feature, the reason for doing, them will be considered

as important when for implementing them. While an effort is made

by features without tying them to particular storage schemes or data

the discussion is made concrete. The similarity across

systems is not the common understanding of human expression but from.

Data Base Selection

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

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

structured records, but also to auxiliary files containing

data, another material supporting the primary file.

1. The "data base" selected is the one

4.2 DATA BASE SELECTION
In contrast with the physically-stored version, there are exceptions
where the user-perceived data base may be a
structurally different data base. Not all of the systems allow the user at his
request to build a multi-data base out of physically distinct
data bases. If one is interested in a topic spanning more than
one data base, he can then separately, only one system (HEADER) allows
the search request alone when switching to a new data base.
SPLICE, LARS, BASIS, TIMES allow requests to a stored and
switched data base. In general, the more controlled
the switching is for a data base, the less likely queries composed
of searches in it will transfer to other data bases. Even where a similar
switching occurs against multiple data bases, the user must combine the
switching requests to hand. It remains for future studies to show how often
users tend 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 as well as for switching to a different one. Tutorial and other help-
features are discussed in the previous chapter. Our conclusion
here is that system users do not need help with data base selection
features.在全国 at least is provided regarding subject coverage, indexing policies,
and effective search techniques. If systems are intended for use by those
with specific needs (e.g., scientists, doctors, lawyers, etc.), perhaps at
least some initial instructions should be provided for each of the available
data bases.
4.2.1 The set of available data bases does not depend upon the user identification.

* 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.

* 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. 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.

* when using the RIQS TUTOR
The user, independently of any available data base display, indicates to the system which data base he wishes to have attached as his first data base.

**SYSTEM PROMPT**
- ENTER FILE, MESSAGE OPTION
- ENTER NAME OF THE DATA BASE TO BE SEARCH
- ENTER DATA BASE NAME
- ENTER NASIS COMMAND

**RESPONSE**
- db, option
- ATTACH, logical unit, db
- db password (password is optional)
- SELECT db
- BEGIN BYPASS (for the user-specified default data base)
- RETRIEVE db

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

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

The user indicates that he wishes to be returned to the point that he was asked to select his first data base.

... CHANGE
... CHANGE
... RESTART

The user will start the system which data base to switch to is indicated.
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"

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 catalogs, he probably will enter one or two words that capture a single aspect of the query. Instructs of interactive searching typically have in counteract effects of the analogies. The question-asker is encouraged to enter the concept at first so that he can discover whether clarification is necessary. He then sees the concept. The index-thumber is encouraged to enter multiple concepts, connecting them together with boolean operators. Direct instructions to first individual may never retrieve anything, while the other may end up manually searching through voluminous output. The sample queries we will see 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 diazo in this chapter is that none of the systems treat simple searching as an opportunit 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 transforming the user's attention to index term selection.
The system assumes the user wants to enter a search request after having selected the data base.

a. ENTER REQUEST
b. SS L/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
   e. ENTER SEARCH COMMAND OR TYPE HALT

f. The user must indicate to the system that he wishes to enter a search request.

   a. SELECT
   b. .SEARCH
   c. FIND

g. 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)

h. The user must specify the logical relations: that

   a. = VALUE
   b. CONTAINS

i. The user must specify the relationship between multiple terms in the query:

   a. 'query words'
   b. is true only for data bases not having multi-word index terms
4.3.6 The user must specify in the query what is to be done with the results of searching.

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. AND
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. 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.

* for data bases where the implied connector is OR

a. INTREX
b. LEADER

DATA CENTERAL
DIALOG
RECON
BASIS
SPIRIS*
ORBIT
PIOS
4.3.10 The user will receive a null request but with a suggestion that improves his chances of entering worthwhile requests.

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.

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

4.3.12 Common words like 'the', 'of', and '.' are deleted from the query.

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

4.3.13 Words like 'effects' (but not 'children') are depluralized.

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

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

For the next five items, the words take up the query take a form like TELEVISION VIOLANCE 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.

* 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.

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

* 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.

* 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.

* 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.

* when using the BASE command
1.4 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 on 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 isolated terms not only reduce the chances of 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 SPIR 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. 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. 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.

<table>
<thead>
<tr>
<th>Operator</th>
<th>Format</th>
<th>Operator</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;</td>
<td>&gt;,GT</td>
<td>=</td>
<td>=</td>
</tr>
<tr>
<td>&gt;=</td>
<td>&gt;=,GE,\textless</td>
<td>&gt;</td>
<td>&gt;,EQ,\textgreater</td>
</tr>
<tr>
<td>=</td>
<td>=</td>
<td>\textless</td>
<td>\textless,LT</td>
</tr>
<tr>
<td>\textgreater</td>
<td>\textgreater,GE,\textgreater</td>
<td>\textless</td>
<td>\textless,LT</td>
</tr>
<tr>
<td>\textgreater</td>
<td>\textgreater,GE,\textgreater</td>
<td>\textless</td>
<td>\textless,LT</td>
</tr>
</tbody>
</table>

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).

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. \textit{TERM}\texttt{***} incorporates terms having (in this example) four or fewer characters following the root.

b. \textit{TERM}\texttt{\$\$}\textit{TERM}, \textit{MASK}\textit{TERM}, or \textit{MASK}\textit{TERM}\texttt{n} incorporate terms having an unlimited number or no more than \texttt{n} characters following the root.

c. \textit{TERM}'\texttt{n} for databases defined to permit truncation (a different symbol could be designed as the truncation code in the data base definition).

d. \textit{TERM}\texttt{*}

e. \textit{ALL}\textit{TERM}'

f. \textit{TERM}:

g. \textit{TERM}! (since truncation is a normal mode, this specifies an exact match of all characters)
1. ONLINE retrieves ONLINE and ON-LINE but not ONLINE.
2. ALL ONLINE retrieves all of the above in something like sequential searching.
3. ONLINE retrieves all of the above via sequential searching.

You can ask for a display of terms alphabetically relevant to the one he designates. Possible features include:

1. That two or more of the displayed terms precede the term alphabetically.
2. That the term is represented in some manner. (Parenthesis indicates how many terms would be entered in a search request.)
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.
5. That the term causing the display is designated in some manner.

<table>
<thead>
<tr>
<th>COMMAND</th>
<th>RESPONSE FEATURES</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIELD TERM</td>
<td>3</td>
</tr>
<tr>
<td>TERM TERM ON</td>
<td>1, 3</td>
</tr>
<tr>
<td>TERM ON</td>
<td>3</td>
</tr>
<tr>
<td>EXPAND ON</td>
<td>4, 6, 7, 8</td>
</tr>
<tr>
<td>CTRAN ON</td>
<td>1, 2, 3, 4</td>
</tr>
<tr>
<td>TERM</td>
<td>0</td>
</tr>
<tr>
<td>EXPAND ON</td>
<td>4, 6, 7, 8</td>
</tr>
<tr>
<td>CTRAN ON</td>
<td>1, 2, 3, 4</td>
</tr>
<tr>
<td>TERM ON</td>
<td>3</td>
</tr>
</tbody>
</table>
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 nodes 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.

<table>
<thead>
<tr>
<th>COMMAND</th>
<th>RESPONSE FEATURES</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;TREE TERM&quot;</td>
<td>Both one level above and two levels below</td>
</tr>
<tr>
<td>EXPAND E9 or E1</td>
<td>1234</td>
</tr>
<tr>
<td>THESAURUS index term</td>
<td>All terms one level below this</td>
</tr>
<tr>
<td>SYNONYM index Term</td>
<td>All synonyms at this level</td>
</tr>
<tr>
<td>ASSOCIATION</td>
<td>For phrases statistically significant</td>
</tr>
<tr>
<td>ORBIT (SB)</td>
<td>DIALOG</td>
</tr>
<tr>
<td>ORBIT (S)</td>
<td>RECON</td>
</tr>
<tr>
<td>N Base</td>
<td>S PIRES</td>
</tr>
<tr>
<td>N ASIS</td>
<td>LEADER</td>
</tr>
<tr>
<td>STAIRS</td>
<td>DATA CENTR</td>
</tr>
</tbody>
</table>
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, 2

The user is automatically told when he enters a search term which other terms are related to it.

The user can specify that all terms hierarchically below the designated one are to be incorporated into his response.

1. EXPLODE TERM
2. Available in some databases by automatic index enhancements
3. LEADER
4. ORBIT
4.4.14 The user can specify how central the terms in a query must be to the content of a record.

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

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.

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

- TERENCE OR (En) : TERM1 TERM2
- TERM1 AND TERM2
- TERM1 + TERM2
- TERM1 - TERM2
- "TERM1" - "TERM2"
- "TERM1" AND "TERM2"
- "TERM1" OR "TERM2"
- "TERM1" AND "TERM2"
The user can specify that words are to appear within 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.

+ TERM1 (w1) TERM2
- TERM1 (w1) TERM2
+ TERM1 + n TERM2
- TERM1 + n TERM2

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

a. TERM1(F) TERM2 in same field.
b. TERM1(nS) TERM2 within n sentences of each other - (if n is omitted) in the same sentence.
c. TERM1(L) TERM2 in same index linkage

b. TERM1 WITH TERM2 in same sentence
b. TERM1 SAME TERM2 in same paragraph
c. WITH TERM2 in same index phrase
d. SENSEARCH (FN) : TERM1 TERM2 in same sentence

d. ORBIT

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

a. SUBJECT TERM
   AUTHOR TERM
   TITLE TERM
b. IF #FIELD1, #FIELD2, #FIELD3 CONTAIN
   search request, FIELD = FIELD NAME
   FIELD NAME relational operator TERM
c. FIELD NAME TERM
   FIELD NAME relational operator TERM
   FN TERM FIELD NAME, TERM
f. IF FIELD NAME relational operator TERM
   search request, FIELD = FIELD NAME
   FIELD NAME, FIELD2, FIELD3.
   FIELD NAME = TERM
   FN = TERM index n
   TERM/FIELD1, FIELD2, FIELD3.
The user can specify that a particular record is to be retrieved, without its being immediately displayed.

1. $NR\ EQU\ RECORD\#,\ \text{RECORD\#}$
2. KEEP\ RECORD\#
3. RECORD#
4. DOCUMENT\ RECORD#
5. BEGIN\ SEARCH\ OF\ RECORD\ RECURD\#
6. .SELECT\ ALL\ RECORD\ EQ\ #
results is based entirely on those matching on the greatest number of criteria. He suggested that the display of records as the percentage meeting the relevant criteria next. Unfortunately, there is no method of creating a request without parts that can be refined separately and later interacted.

To areas where systems might be improved involve search profiles and search hints. Search profiles are often thought of as standing requests for existing 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 statements 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 could effectively be indexing material in their own discipline. Sequential searches has also 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 and at the users' request. Many fields in a record are used infrequently and not warrant the SDI procedure. Yet without a search tool to make them easier, users face having to work at all.
The next few texts deal with boolean operators.

### Boolean Operators

<table>
<thead>
<tr>
<th>AND</th>
<th>OR</th>
<th>AND NOT</th>
<th>OR NOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>*</td>
<td>/</td>
<td>*</td>
<td>/</td>
</tr>
<tr>
<td>COMP</td>
<td></td>
<td>NOT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AND NOT</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td>/</td>
<td>*</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AND NOT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AND NOT</td>
<td>OR NOT</td>
</tr>
<tr>
<td>*</td>
<td>/</td>
<td>*</td>
<td>OR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AND</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td>/</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AND</td>
<td></td>
</tr>
</tbody>
</table>

### Processing Order

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 described in the previous section either cannot be used with certain operators or else have higher priority. The exception is DATA CENTRAL's word search operator with has lower priority than *.

<table>
<thead>
<tr>
<th>DATA CENTRAL</th>
<th>NASIS</th>
<th>ORBIT</th>
<th>INTREX</th>
<th>RECON</th>
<th>STAIRS</th>
<th>SPIRES</th>
<th>RIOS</th>
<th>BASIS</th>
<th>DIALOG</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
<td>e</td>
<td>f</td>
<td>g</td>
<td>h</td>
<td>i</td>
<td></td>
</tr>
</tbody>
</table>

### Data Central System

- DATA CENTRAL
- NASIS
- ORBIT
- INTREX
- RECON
- STAIRS
- SPIRES
- RIOS
- BASIS
- DIALOG

### ORBIT System

- DATA CENTRAL (a)
- NASIS (b)
- ORBIT (c)
- INTREX (d)
- RECON (e)
- STAIRS (f)
- SPIRES (g)
- RIOS (h)
- BASIS (i)
- DIALOG (j)
Deviations of connectors can be controlled through the use of parentheses.

### 4.5.4
A clean operator can be specified for insertion between terms in a range of terms.

- a. set# - set#/operator
- b. TERM1-TERM2/operator
- c. TERM1 operator TERM2 TERM3

### The next eleven items deal with request sets.

```plaintext
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.

1. 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 (#, #, ..., #)
```

### 4.5.6
Input from the user either begins a new request of else continues the current one. While it is possible to make use take in these tasks for INTERPRETER

- a. SPIRES
- b. DATA CENTRAL
- c. LEADER
- d. ENTPFX
**FIRST_PART**
- FIND query
- query leads to phrase selection
- AND or OR query
- query leads to phrase selection

**CONTINUATION**
- AND or OR query
- PHRASES for phrase selection
- AND or OR query
- PHRASES for phrase selection

**BACKTRACK**
- BACKUP (one step)
- MODIFY # (to step #)
- START OVER for query revision while maintaining selected phrases
- NEWSEARCH for total query revision
- SET2 (to bring back SET2)

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

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

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

- a. COMBINE query
- b. AND query

**.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. LEADER
- d. BASIS
- e. DATA CENTRAL
- f. NASIS
- g. LEADER
- h. BASIS
- i. DIALOG
- j. BASIS
- k. RECON
- l. ORBIT
- m. BASIS
- n. DATA CENTRAL
- o. LEADER
- p. BASIS
- q. DIALOG
- r. BASIS
- s. RECON
- t. ORBIT
- u. BASIS
- v. RECON
- w. ORBIT
- x. BASIS
- y. RECON
- z. ORBIT
For each set in the range specified, the text of the query is listed along with the number of records the query retrieved.

- 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.

- LIST ALL. Same effect as for STAIRS.
- PHRASES. The index phrases selected so far are listed along with the number of records each retrieves.
- SETS. Same effect as for DIALOG.

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

- KEEP record#, 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.

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

- PURGE ALL or PURGE #, or PURGE #,
- RELEASE SET # or RELEASE SETS #, #,...,
- 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.
- KEEP SET
4.5.14 The total request can be saved for re-use during same other session.

SAVE
a. ..SAVE NAME
b. PROFILE

d. FN+/SAVE

EXECUTE strategy in working memory.
a serial is assigned

USE
a. ..EXEC NAME
b. RELYM NAME

RECALL "

DELETE
a. ..PURGE NAME
b. 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. ..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.

. 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 managers and database contributors must agree upon problem conceptualization. There must be some guarantee regarding the quality of queries and retrieved results. While for some PINS and SPIRES data bases the conceptual
...tip. The user may ask the system for relevant results, and what should be done next. The display is much simpler than what the user probably does not feel any need for. The results can be sorted by...
A user will see a single record at a time of multiple records and then will be asked to indicate if he wishes to continue.

- TYPE & NEXT (For NEXT 1, 2, 3, 4, 5, or 64 ONE PAGE)
- PAGE
- EXIT ONLY

The user will see a display of up to five records 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 to see before being interrupted.

The user will receive a printed list of the output that can be interrupted or is interrupted automatically.

It may be read in or the terminal output.
44. IS_A7 set_
ALL sPt'
IrSPLI-Il set.
query',I-SPLAY set
F ELD'_-
fi

after the get
IFQR DESKED OUT: E, "I-SPLAY set"
FIL-RET, COM-IF
3. DISPLAY set; -
- NOT ALL
DISPLAY set; E
4. DISPLAY set; 2 TIMES...CAUSE THE SET FOR
- NOT ALL
5. IF query THEN DISPLAY REPORT, F.S.
DISPLAY set; CAUSE THE PROM;
THAT FIELDS BY THE WANT TO
THE RESPONSE TO ALL.
PRINT PLY.
NAMH

1.

TRE.

After receiving printing, 25 lire will be charged for the first three pages printed. If more are required, each additional page costs 10 lire. This will cover the cost of reproduction and mailing.

NAME AND ADDRESS

If the name is less than 14 characters, please write the person's name on the form. If not, please provide a complete mailing address:

NAME, STATE, ZIP

NAME OF PERSON ORGANIZING SHARE

NAME OF PERSON RECEIVING RESULTS

MAIL ADDRESS

If the name is less than 14 characters, please write the person's name on the form. If not, please provide a complete mailing address:

NAME, STATE, ZIP

NAME OF PERSON ORGANIZING SHARE

NAME OF PERSON RECEIVING RESULTS

MAIL ADDRESS

If the name is less than 14 characters, please write the person's name on the form. If not, please provide a complete mailing address:

NAME, STATE, ZIP

NAME OF PERSON ORGANIZING SHARE

NAME OF PERSON RECEIVING RESULTS

MAIL ADDRESS
The following two additional features of display features have been incorporated into this system:

1. **Dialogue Basis**: 
   - *Re. A. I.*
   - *Re. 1.*

2. **RDF (Revised Dialogue Basis)**: 
   - *Re. RDF* 
   - *DIALOG.*

These features are intended to enhance the interaction between the user and the system, providing a more intuitive and user-friendly interface.
CONCLUSION

INTRODUCTION

...the use has been descriptive rather than normative. It has left it to the reader to choose between alternate implementations of equivalent features. A more normative approach will be taken in this chapter. At the April 1977 workshop, designers were asked to specify which features were in some sense "minimal." What this came to mean was that, with a certain 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 real design. The new capabilities of a rapidly evolving computer technology can often be the only guideposts available. A possible set of computer-oriented principles will be presented in the third section along with illustrative examples from the report. The principles will not be used literally but will instead be used as guidelines to a computer-related problem-solving activity. The results obtained will not have been created by someone in the report but will have been created by students. Further utilization of these principles will likely follow.
not get too early to begin deciding what we want to build in the first infrastructure and how when it is intended.

G. Wesley Pratt

Of the features listed in the appendix were considered essential to 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 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 online documentation and assistance are unimportant. Manuals and help 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 reference only very short time. Authors, instructors, consultants, and users are no more comfortable with traditional media than they are with the new ones. It is not surprising that mistakes are made in preparing and users of on-line materials, or that people do not use traditional manuals, even though the computerized ones are.

Help materials are developed for supporting users. Help materials are not general computer literature, or even a part of general reading material. They are essential for effective searching.

In summary, the relation between users' search patterns and their ability to locate information is critical. This text discusses some of the factors that affect searching behavior and suggests ways to improve searching effectiveness.
Of all the features, boolean and is probably the most essential feature. Without ANEs, ORs, and 
& would be difficult to formulate complex requests. Of the eleven 
operators, LEADER does not provide for them. The indexing phrases 
and that a given user are assumed to be additive; retrieved document-
sets are ordered 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 located 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 with a boolean operators the skilled searcher cannot retrieve 
even what he wants.

A good way when fields to examine when matching requests with 
extraneous concepts is essential. Like people, records tend to have 
many aspects of which can be treated in just one way. By establishing 
fields, we can ensure that each field contains only a single type of 
information. It is also possible to compare records along unitary dimen-
sions. This is useful if the searcher is not able to take advantage of the various fields, 
may be that he wants that is the request. He cannot only specify 
which fields he wants to interrogate, but he can usually 
the set of relevant fields. All of the systems except for 
retina, he wants to interrogate the 
fields that he has added Index phrases. While this
INSTRUCTIONAL, DIAGNOSTIC, AND CONTROL FEATURES

SYSTEM 3.2.1

CLASS MODES OF

PERSONAL INSTRUCTION 3.2.3

READING 3.2.4

ON-LINE TRAINING 3.2.7, 3.2.8

DATA BASE OVERVIEW 3.2.7

SAMPLE SEARCHES 3.2.7

ON-LINE DOCUMENTATION 3.3.2, 3.3.3

SEARCH LOGIC TRACING 3.2.4

LIVE HELP* 3.3.4

* LIVING

MODEL
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

SUFFIX REMOVAL*
4.3.13-14, 4.4.5

SEARCH FIELD CONTROL*

DICTIONARY ACCESS*
4.4.7, 4.4.8

RELATIONAL OPERATORS*
4.4.3

SPELLING VARIATIONS
4.4.4

RELATED TERM CAPABILITY
4.4.9

WORD PROXIMITY OPERATORS
4.3.16-17, 4.4.16-19

BOOLEAN OPERATORS*
4.5.1-4

REQUEST SETS*
4.5.7, 4.5.13

PHRASE DECOMPOSITION
4.3.12

SEARCH PROFILES
4.5.14

SEQUENTIAL SEARCHING
4.5.15

* = an essential feature      v = under user control
x = 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 overridden 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 RIGS 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 OR'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.
<table>
<thead>
<tr>
<th>Feature</th>
<th>Implemented Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESULT MANIPULATION FEATURES</td>
<td></td>
</tr>
<tr>
<td>SEARCH REVIEW*</td>
<td>4.5.9</td>
</tr>
<tr>
<td>PREDEFINED FORMATS*</td>
<td>5.2.6, 5.3.1</td>
</tr>
<tr>
<td>ON-LINE FORMATTING*</td>
<td>5.2.7, 5.3.2-5</td>
</tr>
<tr>
<td>RAPID SCAN</td>
<td>5.2.5</td>
</tr>
<tr>
<td>HIGHLIGHTING</td>
<td>5.2.9, 5.3.9</td>
</tr>
<tr>
<td>EXPANDING</td>
<td>5.3.10, 5.3.11</td>
</tr>
<tr>
<td>SORTING</td>
<td>5.3.6</td>
</tr>
<tr>
<td>RANKING</td>
<td>5.3.7</td>
</tr>
<tr>
<td>COMPUTING</td>
<td>5.3.8</td>
</tr>
<tr>
<td>MICROPICHE</td>
<td>5.3.16</td>
</tr>
<tr>
<td>DISPLAY OF GRAPHS</td>
<td>5.3.17</td>
</tr>
<tr>
<td>STATISTICAL INTERFACE</td>
<td>5.3.18</td>
</tr>
<tr>
<td>OFF-LINE PRINTING*</td>
<td>5.3.13-15</td>
</tr>
<tr>
<td>BATCH RETRIEVAL</td>
<td>2.3.7</td>
</tr>
<tr>
<td>PHOTOCOMPOSITION</td>
<td></td>
</tr>
<tr>
<td>DATA ACCESS PROTECTION</td>
<td>3.4.2</td>
</tr>
</tbody>
</table>

* = an essential feature
v = has been implemented
1/2 = currently being implemented

10.2
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 mailed 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.