CENTER FOR THE INFORMATION SCIENCES
LEHIGH UNIVERSITY BETHLEHEM, PENNSYLVANIA

EXPERIMENTAL RETRIEVAL SYSTEMS STUDIES

Report No. 2

Systems Manual for the Experimental Literature Collection and Reference Retrieval System of the Center for the Information Sciences

The work reported here was partially supported by NSF Grant GE-2569, Air Force Grant AF-AFOSR-724-65, and ONR Contract Nonr-610(08).

15 April 1967

59

Center for the Information Sciences
Lehigh University
Bethlehem, Pennsylvania

Experimental Literature Collection and Reference Retrieval System

SYSTEM MANUAL

Introduction

The retrieval system described in this manual includes, at present, some 2,500 document references in the information sciences. The system is undergoing alterations - transfer from tape to disc, with consequent internal changes in record structure. We feel, however, that it is important and necessary to have a bench mark, which collects and organizes the system description and documents the system processes as of April 1967.

Project Director is Robert S. Taylor, Director of the Center for the Information Sciences. Ronald R. Anderson, Anthony F. Amico, and James S. Green have done most of the work on this manual. Other Research Assistants who have helped develop the systems and design the indexing procedures are Robert M. Curtice, A. M. Isaac, Michael A. Jennings, Andrew J. Kasarda, Ruey-Shuang Lee, Wayne B. Powell, and Victor Rosenberg. Work on this reference retrieval system has been supported by NSF Grant GE-2569, Office of Naval Research Contract Nonr-610(08) and Air Force Grant AF-AFOSR-724-65.

Note: In the text, the system is referred to as the CIS System.

Table of Contents

Experimental Literature Collection and Reference Retrieval System of the Center for the Information Sciences (CIS)

A. General Systems Description

1. There are two major purposes in maintaining this collection:

   a. To provide a collection for experimentation, research, and instruction;

   b. To provide a basic reference source for selected papers, books, reports, and data in the field of the information sciences.

2. Acquisition and Selection

In view of the purposes of this collection, selection of material for input is severely restricted. The following notes provide a framework for selection.

a. General papers on information retrieval, documentation, and computer appreciation are excluded, unless there is a specific reason for acceptance, such as basic policy statement (e.g. Weinberg Report) or a particularly cogent description of the field, or a general paper in which specific important data is included.

b. The following areas are included in the collection, with particular emphasis on research, experimentation, and systems analysis.

(1) Automatic Indexing and Abstracting

(2) Syntactic Analysis (but not when the orientation is exclusively mechanical translation).

(3) Logical and mathematical studies of retrieval, relevance, indexing, etc.

(4) Basic systems studies, including costs, major system studies, parallel system problems, compatability, library automation.

(5) Behavioral studies of users, questions, effect of information on management decision, the research process, and on engineering processes.

(6) Programming languages, particularly symbol manipulation languages.

(7) Automata studies, self-organizing systems, cognitive processes, neurophysiology, linguistics, artificial intelligence, computational linguistics.

(8) Education of personnel in information handling and the information sciences.

(9) Pertinent reviews and bibliographies

3. Physical Files Maintained

a. Numeric File: indexing cards and sheets filed by accession number of document.

b. Author File

Computer print-out. Alphabetical by author.

c. Inverted File for Manual Use (Computer-Generated)

Descriptors, alphabetically listed, with document numbers sorted by final digit.

d. List of Descriptors

Alphabetical listing of authorized terms with scope notes and descriptor numbers. (Descriptor lists in ten categories are available for assistance in indexing and searching.)

e. Computer Programs

Documentation of available programs including description, flow charts, program instructions, and keypunching instructions.

4. Indexing

Documents are issued when available to indexers (graduate students), and an attempt is made to distribute according to each individual's interest. Indexing sheets, with any special instructions, are issued with each batch of materials. Indexing sheets are available which require the following information: (See Figure 1)

a. Author(s) and Title

(1) Authors are recorded, last name first, followed by initial of first name. Up to three authors may be recorded. In the case of a collection, symposium, conference, etc., the editor(s) name(s) should be used with "ed." When no author is indicated, the issuing organization is used.

(2) Title is recorded as it appears on the title page or with standard abbreviations if it, together with the title, exceeds 102 characters and spaces.

b. Code Items

(1) Location Code (column 74)

F = item in CIS file
B = item shelved in CIS collection
Blank = item in Lehigh University Library

CENTER FOR THE INFORMATION SCIENCES
LEHIGH UNIVERSITY

experimental collection
coding sheet

INDEXER | LOCATION | CONTROL CODE | DOCUMENT No.

author and title (max. 102 characters)

descptr. no.		descriptor	descptr. no.		descriptor

doc. no.	ret code	date (yr.)		bibliographic data (max. 192 characters)

Figure 1: Coding Sheet (Reduced Size)

(2) Control Codes (columns 75-76)

Number of cards in title-author set

(3) Document Number (columns 77-80)

Assigned after indexing

(4) Indexer

Initials of indexer

c. Descriptors (see latest list of descriptors). Indexer will decide which term or set of terms best describe or give clues to the document content. Record descriptor number (see list) for each term. If a new term is recommended, it should be recorded with a blank for the number. The maximum number of descriptors for any single document is 26.

d. Bibliographic Reference (maximum 192 characters)

(1) Document Number (columns 1-4, card 1)

Assigned

(2) Retention Code (columns 5-7, card 1)

Not used at present

(3) Date (columns 8-11, card 1)

If date not available, estimate as 1961, 1965, etc.

(4) Citation

Book: place and publisher; if portion of book, add inclusive pages. For Symposiums, Conferences, etc., record name of meeting, sponsoring organization, place and date held, publisher, and inclusive pages of item indexed.

Report: place and name of issuing agency, including report number, project name and number, AD or PB numbers if available, and month of issue.

Journal Article: title of periodical, volume number, inclusive pages, and date.

See American Standard for Periodical Title Abbreviations, ASA Z39.5-1963 for abbreviations.

5. The indexing is edited. Decision is made on suggested new terms. Document numbers are recorded in upper right corner of sheet and on document if in CIS collection. Documents are filed.

(2) Control Codes (columns 75-76)

Number of cards in title-author set

(3) Document Number (columns 77-80)

Assigned after indexing

(4) Indexer

Initials of indexer

c. Descriptors (see latest list of descriptors). Indexer will decide which term or set of terms best describe or give clues to the document content. Record descriptor number (see list) for each term. If a new term is recommended, it should be recorded with a blank for the number. The maximum number of descriptors for any single document is 26.

d. Bibliographic Reference (maximum 192 characters)

(1) Document Number (columns 1-4, card 1)

Assigned

(2) Retention Code (columns 5-7, card 1)

Not used at present

(3) Date (columns 8-11, card 1)

If date not available, estimate as 1961, 1965, etc.

(4) Citation

Book: place and publisher; if portion of book, add inclusive pages. For Symposiums, Conferences, etc., record name of meeting, sponsoring organization, place and date held, publisher, and inclusive pages of item indexed.

Report: place and name of issuing agency, including report number, project name and number, AD or PB numbers if available, and month of issue.

Journal Article: title of periodical, volume number, inclusive pages, and date.

See American Standard for Periodical Title Abbreviations, ASA Z39.5-1963 for abbreviations.

5. The indexing is edited. Decision is made on suggested new terms. Document numbers are recorded in upper right corner of sheet and on document if in CIS collection. Documents are filed.

In accepting a new term, a term number is assigned and scope notes are added if necessary. A class code or codes is assigned each new term to permit division in ten general categories.

Theory (Mathematics, Logic, Communication)

Subject Analysis

System Design

Language (Natural, Artificial, Linguistics, Coding)

Behavioral

Equipment

Format

Applications

General

Identifiers

6. Keypunching is done in batches of approximately 100 items.

a. Author cards are keypunched with the first author's last name, plus first initial, then initials and last names of remaining authors.

b. All keypunching is verified, and alphabetized for merging with the existing serial file.

7. Periodic Processes

a. New terms are recorded and entered in the index term list of each indexer or new term lists are generated with the "Thesaurus Listing Program." A separate file of the index terms is also maintained, by class codes.

b. An alphabetic hard copy listing of the author file is issued with the "List Serial File" program.

c. Index Term and Use Review

With the aid of the output of the "Posting" program, each index term is reviewed as to its use. If it is used widely, a scope note may be in order. If it is used rarely, it may be eliminated and replaced with a "use" reference. Terms may be combined or divided. Any decisions are implemented by the "Delete Document", "Delete Terms", "New Terms" programs.

B. Description of Tape Files

1. Summary of CIS Program Library Tape

Link No.	Program
01	Master Chain
02	Off-line Search
03	New Documents
04	New Terms
05	Delete Documents
06	Delete Terms
07	List Serial File
08	List Serial File, Numeric
09	List Inverted File
10	Posting
11	Thesaurus Listing

The magnetic tape files are written in GE-225 machine language (VFAP), in the binary mode. This permits longer records to be written than can be written in Fortran, although the files are Fortran compatible as only every other physical computer word is utilized.

To create the program library tape, we use a deck which is composed of the object decks of the eleven chain programs. To alter a program on the library tape, the old object deck is exchanged for the new one, and the entire library tape is recreated. The tape on handler no. 5 should have a write ring, and will become the new library tape.

2. Serial File

The serial file is blocked with 4 logical records per physical record. Each logical record is 128 Fortran words long. Thus each physical record is 1,024 words. Each logical record contains:

Words	Data
1-34	Author and Title (word 25 is zero if only words 1-24 are used)
35	Date Code
36	Location Code
37	Document Number
38-64	Term Numbers
65	Document Number
66	Retention Code (not used at present)
67	Date Code
68-128	Bibliographic Data

The last term number for each document must be zero; thus a maximum of 26 index terms per document is permitted. This entire file ends with a document number which is equal to -9999. The file is maintained alphabetically by the last name of the author.

3. Inverted File

The inverted file contains unblocked records of 128 Fortran words. Each primary record contains:

Words	Data
1	Term Code Number
2	Number of Documents Listed in this Record
3-15	The Index Term
16-36	Scope Note
37-38	Classification Codes
39-127	Document Numbers
128	0 or -5

The last document entry must be zero, except when the record is filled. In this case, the last location (word 128) contains the code -5, and the very next record is a continuation of this one. The continuation record contains:

Words	Data
1	Term Code Number
2	Number of Documents Listed in this Record
3-127	Document Numbers

There is no limit to the number of continuation records. Each primary record may contain up to 89 document entries, each continuation record may contain up to 125 entries. This file ends with a term number which is negative.

C. Description of Programs on Library Tape

All major programs are on the CIS program library tape. This tape is a chained Fortran tape. Each program is referred to by its chain number.

a. The programs were written for the GE-225 with the following system configuration: 8K memory, 4 magnetic tape handlers, card reader, auxiliary arithmetic unit, and high speed printer.

b. Since all programs on the CIS library tape are called by first calling Chain 1, the user should consult the program write-ups for both Chain 1 and the program he wants to run before assembling his deck.

Chain 1

Program: Master Chain

Purpose: To allow programs on the CIS systems tape to be called into core and executed

Language: Fortran II with VFAP Input/Output routines

Program I/O:

(A) Card Input:

(i) Card 1:

Columns	Data
1-2	Number of the chain program to be called

(ii) Last Card:

Columns	Data
1-2	00 or blank

Example:

Card 1: $\frac{1}{02}$

Cards 2-N: Data cards used by Chain 2

Last Card: Blank

Note: When a program ends, it calls Chain 1 again, thus permitting another chain program number to be read and called. Calling Chain 0 (or the blank card) causes "Call Exit" to be executed.

(B) Tape Input:

(i) CIS library tape

(C) Printer Output:

(i) The number of the chain which has been called

Operator Instructions:

(A) Tape Set-up:

IR-__(CIS library tape) on handler no. 5 (no ring)

Comments: All programs on the CIS library tape are called by first calling Chain 1. This is accomplished with the aid of a two card routine "Call First Link" (L.U. 225 J4.513). Chain 1 also resets much of upper memory to zero.

Chain 2

Program: Off-line Search

Purpose: To perform either associative or non-associative searches

Language: Fortran II with VFAP Input/Output routines

Program I/O:

(A) Card Input:

(i) Card 1:

Columns	Data
1-2	Number of Non-associative Searches
10	Number of Associative Searches

(ii) Card 2:

Columns	Data
1	(
2-6	Term Number
7	Boolean Operator
8-12	Term Number
13	)
14	Boolean Operator
15	(
16-20	Term Number
21	Boolean Operator
22-26	Term Number
27	)
30-32	Search Identification Number
40	1- if associative search; blank otherwise
50-52	Maximum number of documents to be listed - if zero or blank, all documents will be listed

<u>Note</u>: The non-associative searches must appear together followed by the associative searches. A maximum of 10 non-associative and 4 associative searches is allowed. The Boolean operators may be "+", "-", or "V". The operator in column 14 is the major one. All terms must appear from left to right. If less than a four-term search is used, it is not necessary to complete the search statement.

<u>Examples</u>:

```
         1   10
Card 1:  03  1

         1 2       8               16        22              30    40  50
Card 2:  ( 50100 + 63500) V (50100 + 70500)  001       10

Card 3:  ( 50100 + 63500) V (70500           002       10

Card 4:  ( 50100                             003       20

Card 5:  ( 50100 + 63500)                    004   1   10
```

(B) Tape Input:

(i) Serial File

(ii) Inverted File (if associative search)

(C) Printer Output:

(i) A list of search statements

(ii) A list of documents retrieved

(iii) An associativity table (if associative search)

<u>Operator Instructions</u>:

(A) Tape Set-up:

IR-__(Inverted File) on handler number 2; no ring (if associative search)

IR-__(Serial File) on handler number 3; no ring

Program Operation:

1. An image of the specification is set up with all 4 possible term locations set to zeros:

$$(\underline{0} \underset{v}{\pm} \underline{0}) \underset{v}{\pm} (\underline{0} \underset{v}{\pm} \underline{0})$$

A search is made for the occurrence of specified terms in the list of terms which index each document. If the term does actually occur a +1 is entered in the appropriate place in the image, otherwise a -1 is entered. For example, suppose the search specification was (51000 + 72000) - (63000 _____) and the document being searched was indexed by the terms:

51000 81500 72000 63200

Then the image would be:

(+1 + +1) - (-1 ___)

Blank terms are given a value of +1 and blank operations are taken to be +:

(+1 + +1) - (-1 + +1)

A 1 B 2 C 3 D

Image locations have been labelled for reference.

2. Next another image is set up as follows:

$$(\underline{\quad} \underset{v}{\pm} \underline{\quad})$$

E 4 F

The operator no. 1 is then looked at:

a. If it is +, then both locations A and B must be +1 in order to set E to +1, otherwise set E to -1.

b. If it is V, then either A or B can be +1 to set E to +1; otherwise set E to -1.

c. If it is -, then A must be +1 and B must be -1 to set E to +1; otherwise set E to -1.

3. Operator no. 4 is the same as no. 2.

4. Do step 2 with C, D, operator no. 3 and F.

5. Do step 2 with E, F, operator no. 4 and the result.

6. If the result is +1 the document satisfies the request. If it is -1 it does not.

In our example the second image would be (+1 - -1) which reduces to +1 according to Rule 2b so the document satisfies the search.

Each time a document satisfies a search, the author, title, and bibliographic data are printed along with the search identification. A record is kept of how many documents were printed out and if the specification indicated it, printout suppression may occur. Each search specification is printed for reference and upon completion of the search the number of documents satisfying each request is printed.

All terms which index a document that satisfied an associative request are kept along with the number of times they occurred. (Called the co-occurrence value.) Two hundred and fifty different associative terms is the maximum.

If an associative search is entered, the program computes the association values for the associative terms. Each record of the inverted file is read. If the term is one of the associative terms, the following data is printed:

N = The term number

T = The term

F_p = Its total occurrence

R_p = Its co-occurrence with the initial request

F_s = The number of documents resulting from the initial request

A = The association value:

$$A = \frac{R_p^2}{F_p \cdot F_s}$$

See Appendix 1 for discussion of associativity.

Chain 3

Program: New Documents

Purpose: To add documents to the CIS Serial and Inverted Files

Language: Fortran II with VFAP Input/Output routines

Program 1/O:

(A) Card Input (see Figure 1):

(i) Card 1:

Columns	Data
1-72	Author/s (Last name, initials, two blanks) and title
73	Blank
74	Location Code (F - CIS file; B - CIS book; blank - in Lehigh University Library)
75	Blank
76	Control Code (blank or 1- all author and title information is on this card; 2- two cards required for author and title information)
77-80	Document Number

(ii) Card 2 (if needed):

Columns	Data
1-30	Continuation of author and title information
31-76	Blank
77-80	Document Number

(iii) Card 3:

Columns	Data
1-5	Term Number
6-8	Blank
9-13	Term Number
14-16	Blank
Etc. (as needed)	Etc. (as needed)

Note: A maximum of ten terms can be added on this card.

(iv) Card 4 (if needed):

Note: Format for this card is the same as for Card 3. A maximum of ten terms can be added on this card.

(v) Card 5 (if needed):

Note: Format for this card is the same as for Cards 3 and 4. A maximum of six terms can be added on this card.

Note on Cards 3, 4, and 5: The last term number must be 00000.

(vi) Card 6:

Columns	Data
1-4	Document Number
5-7	Blank
8-11	Year of Publication
12	Blank
13	Control Code (blank or 1- all bibliographic information is on this card; 2- two bibliographic cards; 3- three bibliographic cards)
14-80	Bibliographic Data

(vii) Card 7 (if needed):

Columns	Data
1-80	Bibliographic Data

(viii) Card 8 (if needed):

Columns	Data
1-30	Bibliographic Data

(ix) Last Card:

Columns	Data
77-80	-999

Example:

Card 1: 1 ACKOFF, b R.b HALBERT, b M.bb AN OPERATIONS RESEARCH
47 STUDY OF THE DISS OF SCIEN 74 B 76 2 77 0966

Card 2: 1 TIFIC INFORMATION 77 0966

Card 3: 1 50100 bbb 9 60200 bbb 17 59500 bbb 25 89105 bbb 33 66600 bbb
41 00000

Cards 4 & 5: Not needed

Card 6: 1 0966 bbb 8 1958 b 1314 1 INT CONF ON SCIEN INFORM, PROC.
45 WASHINGTON, 1958, V.1, 97-130.

(B) Tape Input:

(i) Old Serial File (first phase of program)

(ii) Scratch Tape (second phase of program)

(iii) Old Inverted File (third phase of program)

(C) Tape Output:

(i) Scratch Tape (first phase of program; same tape as (ii) above)

(ii) New Serial File (second phase of program)

(iii) New Inverted File (third phase of program)

(D) Printer Output:

(i) A listing of the documents added to the Serial File (author, title, document number, and bibliographic data).

(ii) An error message if the number of terms indexing the added documents does not coincide with the number of terms updated on the Inverted File.

Operator Instructions:

(A) Tape Set-up:

IR-__(Old Serial File) on handler number 3; no ring
Scratch tape on handler number 4; with ring
Programmed PAUSE
IR-__(New Serial File) on handler number 3; with ring
Scratch tape on handler number 4; with ring
Programmed PAUSE
IR-__(Old Inverted File) on handler number 3; no ring
IR-__(New Inverted File) on handler number 4; with ring

(B) Further Instructions:

After first "PAUSE", tapes 3 and 4 will rewind. Remove IR-__(Old Serial File) from handler number 3 and mount IR-__(New Serial File). Toggle zero when ready.

After second "PAUSE", tapes 3 and 4 will rewind. Remove IR-__(New Serial File) from handler number 3 and mount IR-__(Old Inverted File). Remove Scratch tape from handler number 4 and mount IR-__(New Inverted File). Toggle zero when ready.

Comments: The documents must be sorted in alphabetical order by author before being added. The number of documents that can be added in one run is dependent on the number of terms indexing those documents. The following table may be used as a guide.

Index terms/document (average)	Documents added
4	166
5	140
6	125
7	111
8	100
9	90
10	83

Chain 4

Program: New Terms

Purpose: To add new index terms to the Inverted File

Language: Fortran II with VFAP Input/Output routines

Program I/O:

(A) Card Input:

(i) Card 1:

Columns	Data
1-39	New Term (alpha)
40	Number for terms with the same spelling but different meaning
45-74	Scope Note
75	Blank
76-80	Term Number

(ii) Card 2 (if needed):

Columns	Data
1-4	Document Number
5-7	Blank
8-11	Document Number
12-14	Blank
Etc. (as needed)	Etc. (as needed)

Note: Document numbers (up to 12) can be added with the new term. If no document numbers are to be added, this card is left blank or with zeros in Columns 1-4.

Example:

```
Card 1:  1
         ACRONYMS                                                    75
                                                                     50430
Card 2:  1
         0000
Card 3:  1
         ATTRIBUTES                                                  76
                                                                     52620
Card 4:  1       3       15
         1244    0333    1000
Card 5:  1                                          45              76
         LIGHT PEN                                  CRT             71750
Card 6:  1
         0000
```

(B) Tape Input:

(i) Old Inverted File

(C) Tape Output:

(i) New Inverted File

Operator Instructions:

(A) Tape Set-up:

IR-__(Old Inverted File) on handler number 3; no ring
IR-__(New Inverted File) on handler number 4; with ring

Comments: In the data deck set-up, the terms must be ordered sequentially by term number. There is no limit to the number of terms which can be added in one run.

Chain 5

Program: Delete Documents

Purpose: To delete documents from the Serial and Inverted Files

Language: Fortran II with VFAP Input/Output routines

Program I/O:

(A) Card Input:

(i) Card 1:

Columns	Data
1-3	Number of documents being deleted (up to 55)
4-6	Blank
7-9	Number of the IR tape reel used for the new Serial File
10-12	Blank
13-15	Number of the IR tape reel used for the new Inverted File

(ii) Card 2:

Columns	Data
1-3	Blank
4-7	Document being deleted
8-10	Blank
11-14	Document being deleted
Etc. (as needed)	Etc. (as needed)

Note: The order of the document numbers is arbitrary.

(iii) Cards 3-6 (if needed):

Note: Same format as Card 2.

(iv) Last Card:

Columns	Data
1-11	END OF DATA

Example:

```
          1       7       13
Card 1:   b12 bbb bb5 bbb b10

          4     11    18    25    32    39    46    53
Card 2:   I608  I696  0386  I006  0555  I231  I232  I233

          60    67    74
          0987  I688  I817

          4
Card 3:   0898

          1
Card 4:   END OF DATA
```

(B) Tape Input:

(i) Old Serial File (first phase of program)
(ii) Scratch Tape (second phase of program)
(iii) Old Inverted File (third phase of program)

(C) Tape Output:

(i) Scratch Tape (first phase of program; same tape as (ii) above)
(ii) New Serial File (second phase of program)
(iii) New Inverted File (third phase of program)

Note: The IR tape specified by the second number on the first data card (columns 7-9) must be used for the new Serial File. The IR tape specified by the third number on the first data card (columns 13-15) must be used for the new Inverted File.

(D) Printer Output:

(i) A listing of the document numbers being deleted from the Serial File.
(ii) A listing of the documents (author, title, and document number) being deleted.
(iii) The IR tape number on which the new Serial File is generated.
(iv) A listing of the document numbers being deleted from the Inverted File.
(v) The IR tape number on which the new Inverted File is generated.

Operator Instructions:

(A) Tape Set-up:

IR-__(Old Serial File) on handler number 3; no ring
Scratch tape on handler number 4; with ring
Programmed PAUSE
IR-__(New Serial File) on handler number 3; with ring
Scratch tape on handler number 4; with ring
Programmed PAUSE
IR-__(Old Inverted File) on handler number 3; no ring
IR-__(New Inverted File) on handler number 4; with ring

(B) Further Instructions:

After first "PAUSE", remove IR-__(old Serial File) from handler number 3; replace it with IR-__(new Serial File), with write ring. Then toggle zero.

After second "PAUSE", remove IR-__(new Serial File) from handler number 3; replace it with IR-__(old Inverted File). Replace scratch tape on handler number 4 with IR-__(new Inverted File), with write ring. Then toggle zero.

Comments: At present, only 55 documents can be deleted at one time. This can be changed to handle a larger number of documents by increasing the DIMENSION value of the LDOC and KDOC arrays to the desired number. This modification is limited only by available memory space.

If repeated document numbers occur on card input, they will be deleted as many times as they occur.

Chain 6

Program: Delete Terms

Purpose: To remove one or more terms from the Inverted and Serial Files

Language: Fortran II with VFAP Input/Output routines

Program I/O:

(A) Card Input:

(i) Card 1:

Columns	Data
1-3	Number of terms being deleted (up to 55)
4-6	Blank
7-9	Number of the IR tape used for the new Inverted File
10-12	Blank
13-15	Number of the IR tape used for the new Serial File

(ii) Card 2:

Columns	Data
1-2	Blank
3-7	Term being deleted
8-9	Blank
10-14	Term being deleted
Etc. (as needed)	Etc. (as needed)

Note: The order of the term numbers is arbitrary.

(iii) Cards 3-6 (if needed)

Note: Same format as Card 2.

Example:

```
          1       7       13
Card 1:   bb5 bbb b16 bbb bb9

             3        10       17       24       31
Card 2:   bb b0500 bb b1300 bb b3400 bb b0600 bb b6600
```

(B) Tape Input:

(i) Old Inverted File (first phase of program)
(ii) Old Serial File (second phase of program)

(C) Tape Output:

(i) New Inverted File (first phase of program)
(ii) New Serial File (second phase of program)

Note: The IR tape specified by the second number on the first data card (columns 7-9) must be used for the new Inverted File. The IR tape specified by the third number on the first data card (columns 13-15) must be used for the new Serial File.

(D) Printer Output:

(i) A listing of the term numbers being deleted from the Inverted File.
(ii) The IR tape number of the new Inverted File.
(iii) A listing of the term numbers being deleted from the Serial File.
(iv) The IR tape number of the new Serial File.

Operator Instructions:

(A) Tape Set-up:

IR-__(Old Inverted File) on handler number 3; no ring
IR-__(New Inverted File) on handler number 4; with ring
Programmed PAUSE
IR-__(Old Serial File) on handler number 3; no ring
IR-__(New Serial File) on handler number 4; with ring

(B) Further Instructions:

When programmed "PAUSE" occurs, remove IR-__(Old Inverted File) from handler number 3 and IR-__(New Inverted File) from handler number 4. Replace them with IR-__(Old Serial File) on handler number 3 and IR-__(New Serial File) on handler number 4. Then toggle zero.

Comments: If a larger number of terms is to be deleted, change the DIMENSION value of the LDES and KDES arrays to the desired number. This modification is limited only by available memory space.

Chain 7

Program: List Serial File

Purpose: To list the author, title, and bibliographic data of each document in the Serial File

Language: Fortran II with VFAP Input/Output routines

Program I/O:

(A) Card Input:

(i) Card 1:

Columns	Data
1-21	Date of the run

(B) Tape Input:

(i) Serial File

(C) Printer Output:

(i) A listing of the entire Serial File (author, title, date published, and bibliographic data). The pages are numbered and contain an appropriate heading and the date.

Operator Instructions:

(A) Tape Set-up:

IR-__(Serial File) on handler number 4; no ring

Chain 8

Program: List Serial File, Numeric

Purpose: To list the document number and the term numbers indexing each document in the Serial File

Language: Fortran II with VFAP Input/Output routines

Program I/O:

(A) Tape Input:

(i) Serial File

(B) Printer Output:

(i) A listing of the document number followed by the terms that index the document, for each document in the Serial File.

Operator Instructions:

(A) Tape Set-up:

IR-__(Serial File) on handler number 3; no ring.

Chain 9

Program: List Inverted File

Purpose: To list the contents of the Inverted File

Language: Fortran II with VFAP Input/Output routines

Program I/O:

(A) Tape Input:

(i) Inverted File

(B) Printer Output:

(i) A listing of the entire Inverted File (term number, term, scope notes, number of documents indexed by term, and documents indexed by term), in term number sequence.

Operator Instructions:

(A) Tape Set-up:

IR-__(Inverted File) on handler number 3; no ring.

Chain 10

Program: Posting

Purpose: To list the contents of the Inverted File

Language: Fortran II with VFAP Input/Output routines

Program I/O:

(A) Tape Input:

(i) Inverted File

(B) Printer Output:

(i) A listing of the entire Inverted File (term number, term, scope notes, count of documents indexed by term, and documents indexed by term), in term number sequence. The document numbers posted on each index term are sorted into columns based on the final digit.

Operator Instructions:

(C) Tape Set-up:

IR-__(Inverted File) on handler number 3; no ring.

Chain 11

Program: Thesaurus Listing

Purpose: To list the index terms in the Inverted File

Language: Fortran II with VFAP Input/Output routines

Program I/O:

(A) Tape Input:

(i) Inverted File

(B) Printer Output:

(i) A listing of the terms, term numbers, and scope notes on the Inverted File.

Operator Instructions:

(A) Tape Set-up:

IR-__(Inverted File) on handler number 4; no ring

D. Description of On-line Search Program (See Figures 2a, 2b, 3a & 3b)

Through the use of the GE Datanet-15 and the Model 33 Teletype, the capability of on-line access to the CIS Serial and Inverted Files is achieved. This capability is utilized by the On-line Search program.

a. The program was written for the GE-225 with the following system configuration: 8K memory, 4 magnetic tape handlers, card reader, auxiliary arithmetic unit, and high speed printer.

b. By writing a subroutine to modify the Fortran READ and PRINT routines, the teletype console assumes the function normally performed by the card reader and high speed printer in the above configuration. This subroutine together with the object deck from the search program combine to form the object deck for the On-line Search program.

c. The On-line Search program is not on the CIS library tape. Instead, the object deck for the program is submitted as a normal run. The machine operator is notified by the Datanet-15 of a call from the teletype terminal and must then mount the appropriate tapes and feed the object deck into the computer through the card reader. The On-line Search program identifies itself to the user by typing a message on the teletype console when it is ready to begin processing a search request.

d. A conversation routine is included in the logic of the program and provides the basis for man/machine interaction by requiring the user to make decisions which guide him through the program.

e. The regular CIS Serial and Inverted Files on tape are used by the program. They were described earlier.

Program: On-line Search

Purpose: To perform either non-associative or associative searches by means of direct interaction with the computer

Language: Fortran II with VFAP tape input routines and modified Fortran II read and print routines

Program I/O:

(A) Tape Input:

(i) Serial File
(ii) Inverted File (if associative search)

(B) Teletype Input:

(i) Search Statement

Columns	Data
1	(
2-6	Term Number
7	Boolean Operator
8-12	Term Number
13	)
14	Boolean Operator
15	(
16-20	Term Number
21	Boolean Operator
22-26	Term Number
27	)

Note: The search statement is typed in by the user in response to the statement: "I AM READY FOR YOUR SEARCH STATEMENT." The Boolean operators may be "+", "-", or "V". The operator in column 14 is the major one. In other words, a pre-structured input format from which search operations are to be performed is used. All terms must appear from left to right. If less than a four-term search is used, the input may be terminated after the last term. It is not necessary to complete the search statement.

```
        THIS IS THE CIS DOCUMENT REFERENCE SEARCH SYSTEM


  DO YOU WISH AN ASSOCIATIVE SEARCH
?YES
  I AM READY FOR YOUR SEARCH STATEMENT
?(60000

YOUR SEARCH STATEMENTS ARE

     1  60000      0      0      0


   28 DOCUMENTS SATISFIED SEARCH LABELED  1

                T E R M   A S S O C I A T I O N S

TERM NO    TOT. OCC.   CO-OCC.   TERM                              ASSOCIATIVITY

 51600          2          1     AMER. INST. OF CHEMICAL ENGINEERS  0.0179
 59400          9          3     DDC (ASTIA)                        0.0357
 60000         28         28     DESCRIPTORS                        1.0000
 66200         14          3     HEADINGS                           0.0230
 66500         33          5     HIERARCHY                          0.0271
 68700        250         10     INDEXING                           0.0143
 88900         21          4     TERM                               0.0272
 89500         62          9     THESAURUS                          0.0467
 91700         28          6     VOCABULARY                         0.0459
 91900          2          1     WATER                              0.0179


  BASED ON THE ASSOCIATIVITY TABLE
  DO YOU WANT TO MODIFY YOUR INITIAL SEARCH STATEMENT
?YES
  DO YOU WISH AN ASSOCIATIVITY TABLE
  WITH RESPECT TO YOUR NEW STATEMENT
?YES
  I AM READY FOR YOUR SEARCH STATEMENT
?(60000+66500

YOUR SEARCH STATEMENTS ARE

     1  60000+66500      0      0


    5 DOCUMENTS SATISFIED SEARCH LABELED  1
```

Figure 2a: On-line associative search (continued on 2b)

```
                    T E R M   A S S O C I A T I O N S

TERM NO    TOT. OCC.   CO-OCC.   TERM                          ASSOCIATIVITY

 54900         16          1      CATALOG %ING*                    0.0125
 55700        134          3      CLASSIFICATION                   0.0134
 60000         28          5      DESCRIPTORS                      0.1786
 66100         13          1      GROUPING %DOCUMENTS*             0.0154
 66500         33          5      HIERARCHY                        0.1515
 87600          6          1      SYMBOL %IC.S*                    0.0333
 89500         62          2      THESAURUS                        0.0129


  BASED ON THE ASSOCIATIVITY TABLE
  DO YOU WANT TO MODIFY YOUR INITIAL SEARCH STATEMENT
?NO



  1  CIS DOC. NO.  F   865
DATATROL CORP.  COMMON VOCABULARY APPROACHES FOR GOVT SCIENTIFIC +
 TECHNICAL INFO SYSTEMS
 1963 DATATROL CORP, NSF C-342, AD430000, DEC 1963


  1  CIS DOC. NO.  F  1254
HILLMAN, D.  STUDY OF THEORIES + MODELS,  REPT 1  PROBLEMS, SYSTEMS,
 AND METHODS
 1962 LEHIGH UNIV, CENTER FOR INFORM SCIEN, REPT 1, NSF G-24070, AUG 196


  1  CIS DOC. NO.  F   859
LEFKOVITZ, D.  AUTOMATIC STRATIFICATION OF DESCRIPTORS
 1963 UNIV OF PENNA, MOORE SCH OF ELEC ENG, NONR 551(40), SEPT 1963


  1  CIS DOC. NO.  B  1207
MOOERS, C.  A MATHEMATICAL THEORY OF LANGUAGE SYMBOLS IN RET.
 1959 INT CONF ON SCIEN INFORM, PROC. WASHINGTON, 1958, V.2, 1327-1364


  1  CIS DOC. NO.      1515
NEEDHAM, R.  KEYWORDS AND CLUMPS
 1964 J OF DOC 20, 5-15, MAR 1964


WELL, DO YOU WANT TO SEARCH SOME MORE
?YES
```

Figure 2b: On-line associative search (continued from 2a)

```
          THIS IS THE CIS DOCUMENT REFERENCE SEARCH SYSTEM


  DO YOU WISH AN ASSOCIATIVE SEARCH
?NO
  I AM READY FOR YOUR SEARCH STATEMENT
?(71600+53000
  CAN YOU THINK OF SOME MORE SEARCH STATEMENTS
?NO

YOUR SEARCH STATEMENTS ARE

     1  71600+53000      0      0



  1  CIS DOC. NO.  B  1265
ALEXANDER, S.  THE CURRENT STATUS OF GRAPHIC STORAGE TECHNIQUES
 1964 IN CONF ON LIBRARIES AND AUTOMATION, WARRENTON, VA. 1963, PROC.
  WASHINGTON, D.C., 111-140


  1  CIS DOC. NO.  F   813
BOLT BERANEK AND NEWMAN INC  TOWARD THE LIBRARY OF THE 21ST CENTURY
 1964 A REPT ON PROGR MADE IN A PROGRAM OF RES, MAR 1964


  1  CIS DOC. NO.      1489
DENNIS, B.  FIVE OPERATIONAL YEARS OF INVERTED MANIPULATION AND ABSTRACT
 RETR BY COMPUTER
 1962 J OF CHEM DOC 2, 234-242, OCT 1962


  1  CIS DOC. NO.      1513
FAIRTHORNE, R.  AUTOMATA AND INFORMATION
 1961 IN HIS TOWARDS INFORM RETRIEVAL, BUTTERWORTHS, 11-21


  1  CIS DOC. NO.  B  1308
GOLDHOR, H.  ED.  PROCEEDINGS OF THE 1963 CLINIC ON LIBRARY              0
 APPLICATIONS OF DATA PROCESSIN
 1964 UNIV OF ILLINOIS GRAD SCH OF LIBR SCIEN, PROC. APR-MAY 1963,
  URBANA, ILL


  1  CIS DOC. NO.      1331
HOWERTON, P.  INFORMATION HANDLING, FIRST PRINCIPLES
 1963 SPARTAN BOOKS, WASHINGTON
```

Figure 3a: On-line non-associative search (continued on 3b)

```
  1  CIS DOC. NO.  F  1383
INTERNATIONAL BUS. MACH.  MECHANIZED LIBRARY PROCEDURES
   62 IBM, WHITE PLAINS, NY, E20-8094-1


  1  CIS DOC. NO.  B  1269
KING, G.  THE AUTOMATION OF LIBRARY SYSTEMS
 1964 IN CONF ON LIBRARIES AND AUTOMATION, WARRENTON, VA, 1963, PROC.
  WASHINGTON, D.C., 233-242


  1  CIS DOC. NO.  B  1109
LAMKIN, B.  INTEGRATED LIBRARY MANAGEMENT SYSTEMS CONCEPT
 1964 AMER DOC INST, ANNUAL MEETING, 1964, PROC., 1, 141-147


  1  CIS DOC. NO.  B  1264
LIBBY, R.  AUTOMATED STORAGE AND ACCESS OF BIBLIOGRAPHIC INFO FOR      O
 LIBRARIES
 1964 IN CONF ON LIBRARIES AND AUTOMATION, WARRENTON, VA, 1963, PROC.
  WASHINGTON, D.C., 67-107


DO YOU WANT MORE DOCS.
?YES


  1  CIS DOC. NO.  B  1263
PATRICK, R. AND D. BLACK  INDEX FILES  THEIR LOADING AND ORGANIZATION  O
 FOR USE
 1964 IN CONF ON LIBRARIES AND AUTOMATION, WARRENTON, VA, 1963, PROC.
  WASHINGTON, D.C., 29-64


  1  CIS DOC. NO.      1062
QUENZEL, C.  SOME PROPOSALS FOR HANDLING THE INFO. PROBLEM  A BRIEF    O
 BIBLIOGRAPHICAL ESSAY
 1963 AMER DOC 14, 145-148, APR 1963


  1  CIS DOC. NO.  B  1262
SWANSON, D.  DESIGN REQUIREMENTS FOR A FUTURE LIBRARY
 1964 IN CONF ON LIBRARIES AND AUTOMATION, WARRENTON, VA, 1963, PROC.
  WASHINGTON, D.C., 11-25


  1  CIS DOC. NO.  F   789
SWANSON, D.  DIALOGUES WITH A CATALOG
 1964 LIBR QUART, 113-125, JAN 1964

   14 DOCUMENTS SATISFIED SEARCH LABELED  1


WELL, DO YOU WANT TO SEARCH SOME MORE
?YES
```

Figure 3b: On-line non-associative search (continued from 3a)

Examples:

```
                     1 2       7 8       13 14 15 16      21 22      27
Four-term search:    ( 50100   + 63400   )  V  ( 50100    +  71500   )
```

Three-term search: (50100 + 63400) V (71500

Two-term search: (50100 + 63400

One-term search: (50100

(ii) Conversational Responses:

Note: "YES or "NO" answers are required to several questions which are designed to guide the user through the program. The following questions may be asked during the search process:

1) "DO YOU WISH AN ASSOCIATIVE SEARCH?"

2) "CAN YOU THINK OF SOME MORE SEARCH STATEMENTS?"

3) "DO YOU WANT MORE DOCUMENTS?"

4) "DO YOU WISH A COMPLETED SEARCH WITH THE NUMBER OF HITS?"

5) "WELL, DO YOU WANT TO SEARCH SOME MORE?"

6) "BASED ON THE ASSOCIATIVITY TABLE, DO YOU WANT TO MODIFY YOUR INITIAL SEARCH STATEMENT?"

7) "DO YOU WISH AN ASSOCIATIVITY TABLE WITH RESPECT TO YOUR NEW STATEMENT?"

(C) Teletype Output: (See Figures 2a, 2b, 3a & 3b)

(i) Informative Statements:

Note: These are statements which do not require a response from the user. The following statements may be typed:

1) "THIS IS THE CIS DOCUMENT REFERENCE SEARCH SYSTEM."

2) "YOUR SEARCH STATEMENTS ARE" (followed by a list of the search statements just received).

3) "x DOCUMENTS SATISFIED SEARCH LABELED y." (x is the number of documents and y is the number of a particular search statement.)

4) "YOU HAVE STRUCK OUT ON SEARCH LABELED y." (y is the number of a particular search statement.)

5) "GOOD-BYE...COME BACK AGAIN."

(ii) Questions requiring answers from the user:

Note: A list of possible questions appears above under Teletype Input: Conversational Responses.

(iii) A list of documents retrieved.

(iv) An associativity table (if associative search).

Operator Instructions:

(A) Tape Set-up:

IR-__(Inverted File) on handler number 2; no ring (if associative search)
IR-__(Serial File) on handler number 3; no ring

Program Operation: The search logic used is the same as explained earlier for the Off-line Search program callable from the CIS library tape.

A maximum of five non-associative searches may be processed at one time. Each time a document satisfies a search, the author, title, and bibliographic data are printed along with the search identification number. When ten documents have been printed, the user is asked if he wants to see any more. He is asked the same question after each subsequent group of ten documents. If he stops the document printout before the entire tape has been searched, he is given the option of completing the search and printing out the number of documents that would have been retrieved, but not printing out complete bibliographic data.

Only one associative search may be processed at a time. The user is first presented with the total number of documents satisfying his search. Based on this search, he is given an associativity table composed of the index terms associated with the documents retrieved. This table contains the term number, term, total occurrence of the term in the entire document collection, co-occurrence of the term in the

documents retrieved by his search, and an associativity value. Only terms with associativity values greater than 0.0125 are given in the table. The formula for calculating associativity is the same one used in the Off-line Search program callable from the CIS library tape. The user is given the option of modifying his search statement based on the information provided by the associativity table or printing out the documents already retrieved.

Sample printouts from both associative and non-associative searches are shown in Figures 2a, 2b, 3a & 3b.

Comments: The user notifies the computer that he has completed his response by simultaneously depressing the control key and typing the EOM key. The program signals it is ready to accept input from the user by executing a carriage return and line feed followed by a question mark.

E. Description of Disk Files

1. Serial File

The first disk location frame of each serial record is developed from the document accession number by means of the following formula:

DISK FRAME = 2 x (DOCUMENT NUMBER)

Each serial record is 128 FORTRAN-II words in length. However, each serial file record will occupy 256 actual machine words when it is being utilized by the program. This is because only every other word in a non-floating point FORTRAN-II array contains addressable, significant information. In order to save space on disk, special subroutines are used to expand each record read in from disk and to contract each record to be written on disk.

Serial records are written on disk beginning at frame 2 and ending at frame 8001. This will allow expansion up to a total of 4000 document records. Each record occupies 2 disk frames.

Serial record format:

Words	Contents
1-34	Author and Title (word 25 is zero if words 1-24 only are used)
35	Date Code
36	Location Code
37	Document Accession Number
38-64	Term Numbers
65	Document Number
66	Author Order Chain Address
67	Date
68-128	Bibliographic Data

2. Inverted File

The primary records of the inverted file are written sequentially on disk beginning at frame 8002. Each record occupies 2 disk frames. Continuation records are written on disk beginning at frame 9026. If

a record has a continuation record, the last location (word 128) of the record to be continued contains the first frame number of the continuation record.

A disk address table fo the inverted file begins at frame 12280. This table occupies 8 sequential frames (i.e. through frame 12287). The first word of the table contains the number of the first frame of the first inverted record. Starting with the second word of the table, all of the term numbers are written in ascending sequence. Thus, the first disk frame for any primary inverted record can be calculated using this formula:

FIRST FRAME = NO. OF 1ST FRAME + 2 (POSITION IN TABLE -1)

There is no limit to the number of continuation records. Each primary record may contain up to 89 document numbers and each continuation record may contain up to 125 document numbers. The last word in any record (word 128) must be zero except when a continuation is to follow. In this case, word 128 will contain the number of the first frame of that continuation record.

The disk file structure is summarized in the following table:

File	Disk Frame Address
SERIAL	0 - 8001
INVERTED	8002 - 9025
INVERTED CONTINUATIONS	9026 - 12279
INVERTED TABLE	12280 - 12287

F. Description of Disk Programs

At the present time, the entire CIS Document Retrieval System is being converted to operate with the files on disk storage.

a. The programs are being written for the GE-225 with the following system configuration: 8K memory, disk storage unit, card reader, auxiliary arithmetic unit, and high speed printer.

b. As in the On-line Search program written for tape files, the disk programs may be run on-line using the Model 33 Teletype and the GE Datanet-15. This is done by combining the subroutine for modifying the Fortran II READ and PRINT routines with the regular object deck for the disk program.

c. A special set of CIS Fortran II Disk routines has been written to facilitate the usage of reserved disk storage. These subroutines are kept in the computer room and must be inserted into the object deck of the disk program at run time by the operator.

Program: Disk Search

Purpose: To perform both associative and non-associative searches using the Serial and Inverted Files on disk

Language: Fortran II with special Fortran II disk input/output subroutines

Program I/O:

(A) Disk Input

(i) Serial File
(ii) Inverted File
(iii) Inverted File Address Table

(B) Teletype Input:

(i) Search Statement:

Columns	Data
1	(
2-6	Term Number
7	Boolean Operator - if more than one term in the search statement; "blank or)" - if only one term in the search statement
8-12	Term Number
13	)
14	Boolean Operator
15	(
16-20	Term Number
21	Boolean Operator - if four terms in the search statement; "blank or)" - if three terms in the search statement
22-26	Term Number
27	)

Note: The Boolean operators can be either "+", "-", or "V". If less than a four-term search is used, the final parenthesis is optional.

Examples:

Four-term search: $(\overset{2}{50100} + \overset{8}{61000}) \text{ V } (\overset{16}{50100} + \overset{22}{79500})$

Three-term search: (50100 + 61000) V (56500

Two-term search: (50100 + 61000

One-term search: (50100

(ii) Conversational Responses:

Note: "YES" or "NO" answers will be required to questions intended to guide the user through the search program.

(C) Teletype Output:

(i) Interrogative Statements:

Note: These are questions in the program that may be asked to the user. The program will branch according to his response.

(ii) Informative Statements:

Note: These are identifying statements designed to inform the user, but not requiring a response on his part.

(iii) A listing of the documents satisfying the search statement:

Note: Three optional printout formats are available to the user:

1) The document accession number
2) The document accession number, location code, author, and title
3) The document accession number, location code, author, title, and bibliographic data

(iv) A listing of the associativity table (if an associative search is entered)

Note: The information presented in the associativity table is the same as was explained earlier for the On-line Search program.

Operator Instructions:

(A) Disk Set-up:

CIS Serial and Inverted Files and Inverted File Address Table are located permanently on frames 0-12287 of disks 4-5.

(B) Further Instructions:

Include binary deck LU225 E8.508 with the binary deck for this program.

Program Operation:

The validity of each of the terms and Boolean operators is checked by the program, and the user is informed of any error detected. He is also given a list of the actual terms corresponding to the term numbers in his search statement. If, for any reason, the user does not want a search performed on the terms so listed, or if any error has been detected, he may either enter another statement or terminate the program.

For each term represented in an acceptable search statement, an Inverted File record is read into computer memory from the Mass Random Access Disc Storage (MRADS). A list of the accession numbers of all documents indexed by any one term is contained in that term's Inverted File record. According to the logical operations designated between the search terms, the program makes comparisons between the list of documents. On the basis of this comparison procedure, a list of the accession numbers for all documents matching the search specifications is developed in the computer memory.

Following the search and comparison operation, the on-line user is informed of the number of documents yielded by his search statement. In the event that his search yielded no documents, he is given the

choice of either typing in another search or terminating the program. Similarly, if the user decides that the number of documents is either too many or too few, he is given the opportunity to enter a modified version of his original search or terminate entirely.

If the resulting number of documents seems satisfactory to the user, he is asked to select one of the three possible formats in which document information may be typed out. These options are explained under Teletype Output: (iii) above. In the case of output options (2) and (3), a record for each document is read in from the serial file, which is also contained on MRADS. With option (1), the document accession numbers are typed out directly from the list developed in computer memory by the search and comparison procedure. After all the document information has been typed out, the user may enter a new search statement or terminate the program.

APPENDIX 1

Computing Associativity within the Descriptor Set

We wish to find the degree of associativity between a term, i.e. Linguistics in the first example below, and its profile terms. Profile terms are those descriptors us d to index the same document as the original term. To do this, we must know:

F - the absolute frequency of all terms in the profile, i.e. the total number of occurrences in the whole collection.

R - the relative frequency of each profile term, i.e. the number of times it co-occurs with the original term. (i.e. in this case Linguistics)

F_p = absolute frequency of specific profile term

F_s = absolute frequency of original term

R_p = relative frequency of specific profile term

Note $R_s = F_s$

Associativity: $A = \frac{R_p^2}{F_p \cdot F_s}$

Sample for Linguistics

Term	F	R	A
Automatic	65	5	0.0048
Computation(al)	7	5	0.0446
Computer	140	14	0.0175
Indexing	224	1	0.0001
Language, Natural	62	28	0.1581
Linguistics	80	80	1.0000
Parsing	9	6	0.0500
Phoneme	2	2	0.0250
Semantic(s)	34	13	0.0621
Styl(e,istic)	3	2	0.0167
Syntactics	12	4	0.0167
Translators (Computers)	2	1	0.0062

Major Profile Terms for LINGUISTICS (110 Profile Terms)

Cut Off Point A = 0.0125

A	Term
.1581	Language, Natural
.0962	Transformations
.0840	Grammar
.0800	Phrase
.0723	Sentence
.0621	Semantics
.0598	Syntax
.0521	Dependence
.0500	Parsing
.0462	Translation
.0446	Computation(al)
.0360	Structure
.0284	Answers
.0250	Psycholinguistics
	Discourse
	Phoneme
.0192	Analysis
.0175	Computer
.0167	Style
	Syntactics
	Words
.0145	Question
.0128	Logic
.0125	Heuristic
	German
	Sememics
	Problem Solving
	List

Major Profile Terms for CODING (144 Profile Terms)

Cut Off Point A = 0.0125

A	Term
.0944	Punched Cards
.0826	Chemistry
.0680	Superimposed
.0405	Steroids
.0296	Patents
.0291	Ruly English
.0259	Compounds
.0259	Notation
.0175	Symbol(ic,s)
.0148	Structure
.0146	Word Pairs, Groups
.0135	Search(ing)
.0129	Filmorex
.0129	Edge

Major Profile Terms for INDEXING (188 Profile Terms)

Cut Off Point A = 0.0125

A	Term
.2548	Coordinate
.0397	Automatic
.0344	Subject
.0339	Cranfield Project
.0294	Roles
.0282	Classification
.0280	Uniterm
.0259	Link(s,ages)
.0254	Evaluation
.0208	Systems
.0202	Abstracting
.0201	Efficiency
.0187	Retrieval
.0187	Machine
.0181	Comparison
.0175	Chemistry
.0161	Headings
.0137	Term
.0135	Frequency
.0132	Thesaurus

Major Profile Terms for LOGIC (96 Profile Terms)

Cut Off Point A = 0.0125

A	Term
.1250	Propositional
.1134	Lattice
.0744	Grouping (Documents)
.0731	Theory
.0698	Models
.0682	Algorithm
.0577	Topology
.0567	Graph Theory
.0563	Mathematic(s,al)
.0556	Algebra
.0473	Boolean Algebra
.0460	Retrieval
.0375	Calculus
.0208	Tabular
.0185	Matrix
.0170	Network
.0167	Sets
.0160	Syntax
.0128	Linguistics

Major Profile Terms for SYSTEMS (180 Profile Terms)

Cut Off Point A = 0.0125

A	Term
.2204	Information
.1056	Design
.0576	Retrieval
.0465	Costs
.0399	Librar(ies,y)
.0374	Management
.0362	Organization
.0291	Evaluation
.0289	Computer
.0278	Requirements
.0257	Analysis
.0254	Military
.0232	Search(ing)
.0207	Indexing
.0203	Equipment
.0203	Models
.0181	Program(med,ming)
.0174	Automation
.0169	Storage
.0144	Engineering
.0143	Review
.0143	User
.0139	Comparison
.0138	Theory
.0129	Data Processing

Major Profile Terms for COMPUTER (184 Profile Terms)

Cut Off Point A = 0.0125

A	Term
.1623	Program(med,ming)
.0878	Time Sharing
.0684	Memory
.0563	Language, Artificial
.0530	Man
.0498	Project MAC
.0391	Artificial Intelligence
.0360	Compilers
.0293	Machine
.0279	Systems
.0270	Indexing
.0270	Publication
.0266	Associat(ion,ive)
.0264	Language, Natural
.0254	Learning
.0216	Recognition
.0216	Simulation
.0169	Data Processing
.0166	Format
.0166	Retrieval
.0161	Files
.0154	Heuristic
.0154	Graphics
.0154	Pattern
.0152	Type Setting
.0149	Text(ual)
.0135	LISP
.0135	Addresses
.0135	Interface
.0135	Threaded List
.0134	Linguistics

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Center for the Information Sciences
Lehigh University
Bethlehem, Pennsylvania 18015

2a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b. GROUP

3. REPORT TITLE
SYSTEMS MANUAL FOR THE EXPERIMENTAL LITERATURE COLLECTION AND REFERENCE RETRIEVAL SYSTEM OF THE CENTER FOR THE INFORMATION SCIENCES

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(First name, middle initial, last name)*
Robert S. Taylor, Ronald R. Anderson, Anthony F. Amico, and James S. Green

6. REPORT DATE
15 April 1967

7a. TOTAL NO. OF PAGES
56 pages

7b. NO. OF REFS

8a. CONTRACT OR GRANT NO. AF-AFOSR-724-65; Nonr-610(08); NSF-GE-2569

b. PROJECT NO.

c. 61445014

d. 681304

9a. ORIGINATOR'S REPORT NUMBER(S)
Experimental Retrieval Systems Studies, Report No. 2

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT
Distribution of this Document is Unlimited

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Air Force Office of Scientific Research
Directorate of Information Sciences
Office of Naval Research
Information Systems Branch

13. ABSTRACT

This manual describes and documents the retrieval system used by the Center for the Information Sciences for selected current literature of the information sciences, about 2,500 document references. The system is presently on-line via teletype and conversion is in process from tape to disk. Both associative and non-associative search systems are in operation.

DD FORM 1 NOV 65 1473

Security Classification

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Information Retrieval						
Information Sciences						
Search						
On-line						
GE-225						
Associativity						

Security Classification