NASA CONTRACTOR REPORT



NASA CR-1318



# REMOTE INFORMATION RETRIEVAL FACILITY

*by Roger K. Summit*

*Prepared by*
LOCKHEED AIRCRAFT CORPORATION
Palo Alto, Calif.
*for*

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • APRIL 1969

NASA CR-1318

# REMOTE INFORMATION RETRIEVAL FACILITY

By Roger K. Summit



Issued by Originator as Report No. N-07-68-1

Prepared under Contract No. NASw-1454 by
LOCKHEED AIRCRAFT CORPORATION
Palo Alto, Calif.

for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – CFSTI price $3.00

# FOREWORD

This report summarizes work performed under NASA Contract NASw 1454. The first phase consisted of installing a remote terminal base at Ames Research Center, Moffett Field, California, for the use of engineers and scientists in conducting on-line searches on the NASA collection of machine-readable document citations. This data base was stored in a mass storage device at the Lockheed Palo Alto Research Laboratory. Communication between the user and the data base was conducted over a telephone line using an interactive information retrieval language, DIALOG, developed by Lockheed. During the second phase of the contract, the terminal was relocated at NASA headquarters in Washington, D.C. where additional searches were conducted.

Special acknowledgment is due the Scientific and Technical Information Division of NASA for its pioneering effort in information retrieval, and its continuing developmental support in this important discipline.

CONTENTS

## ILLUSTRATIONS

Section 1

SUMMARY OF WORK PERFORMED
Contract NASw-1454

## 1.1 NASA FILE CONVERSION

The NASA linear file, inverted file, and cross-reference file, which were furnished by NASA, were converted to a standard format compatible with the IBM 360 system. Conversion of the linear file resulted in an average compression of 28%. Much of the greater storage efficiency resulted from elimination of coded terms and blank spaces from the records. All files were loaded on the IBM 2321 data cell. Indices to the files were generated using the IBM index sequential file management system (ISFMS). These indices were stored in a special format on the IBM 2311 disk.

A detailed description of the file conversion process can be found in Appendix A.

## 1.2 SYSTEM OPERATION, AMES RESEARCH CENTER

On completion of file conversion and after local checkout of the files with the DIALOG information retrieval system (described in Section 2), a remote terminal consisting of an IBM 2260 keyboard/display device and an IBM 1053 character printer were installed at the Ames Research Center in Mountain View, California. Remote operation was conducted via a 1200 bit/second, full duplex, schedule 4A telephone line using Model 202 data sets.

For a period of 2 months, Ames scientific and library personnel operated the system on the average of 2 hours per day performing literature searches on the computer-based collection of 300,000 report and journal citations. (See Appendix B for search summaries.) Following an announcement of the availability of the service, user's manuals were distributed to interested parties and a signup sheet was maintained. After the first month, signups were backlogged an average of 1 to 2 weeks.

The general reaction of the users, based on their comments and repeated searches, was favorable. Ames is currently contemplating an extention of the service. Excepting substantial trouble with the telephone line, and occasional trouble with the IBM equipment and the application programs, the system operated reliably.

Several minor modifications were made to the DIALOG search system during the course of performance, based on user reaction and observations of user needs.

## 1.3 SYSTEM OPERATION, WASHINGTON, D. C.

The terminal was shipped and installed in Washington, D. C., in July, 1967. During the ensuing 12-months, searches were conducted during a 3-hour period each morning. Whereas searches were primarily conducted by the final customers during the first phase, most searches were conducted by the NASA Headquarters Librarian, Mrs. Karen Milligan, during the second phase.

During this period, approximately 300 searches were conducted. Also, files were enlarged to a total of 400,000 citations, abstracts for the 1967 items were included, and the W series (describing research projects) was included. Service on the whole was more reliable than during the first phase with only negligible telephone line failure. Inclusion of a "soft restart" capability largely eliminated catastrophic software failures. Continued difficulty was experienced with the IBM 2321 Data Cell, however, which accounted for most of the nonrecoverable failures.

## Section 2
## DESCRIPTION OF THE DIALOG SYSTEM

This section reviews the procedural techniques and computer facilities used in conducting on-line searches – specifically, the DIALOG information retrieval language and the computer environment within which this set of programs operates.

### 2.1 DESCRIPTION OF THE DIALOG LANGUAGE

DIALOG is an interactive information retrieval language which allows the user to formulate simple or complex search requests via a video/keyboard display terminal (Fig. 2-1) which is coupled to a computer containing the material to be searched. To use DIALOG, the engineer or librarian enters desired commands such as "begin search," "display," "select," or "combine" by depressing keys on the keyboard (Fig. 2-2). The computer responds by displaying various data on the display screen and/or on the console printer as appropriate.

Four important phases in the search are:

- Select index terms or subject phrases
- Combine individual terms or phrases
- Display output
- Modify search expression

In the first step, the user identifies and selects index terms (i.e., subject phrases) of interest which relate to his search topic. The second step allows the user to build his own specialized subject heading during the search. At each step the user is informed of the number of documents which fulfill his specified search criteria. The third step, display of the output, allows the user to review references for relevancy,

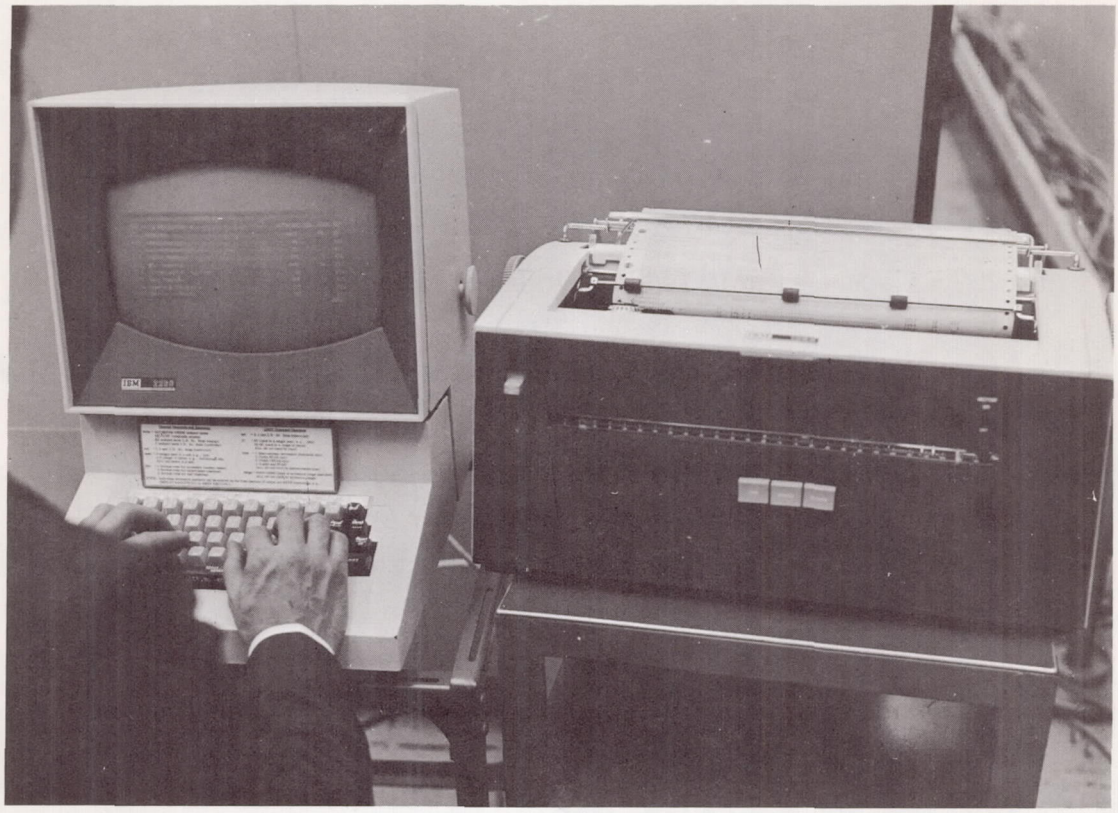


Fig. 2-1 DIALOG Video/Keyboard Display Terminal


Fig. 2-2 Display Keyboard Showing DIALOG Commands

and to select additional terms to better express his search requirement. Once the desired literature references are displayed, the user can press the printout key for a permanent record of the retrieved references, or can modify his search expression based on index terms used in displayed references.

The NASA linear file, inverted file, and related term files are located on the IBM 2321 data cell. The descriptor and accession number indices are located on one of the two 2311 disk drives. The system reaction to several of the commands is described below.

To assist the user in identifying and selecting index terms, a command labeled "EXPAND" is provided which causes a display of the alphabetically near index terms for any term entered. For each term displayed, the number of items indexed by that term as well as the number of thesaurus terms is shown. This function provides the user with a continually updated catalog of the index and cross-reference terms and their usage. Figure 2-3 shows the display resulting from EXPAND TECHNOLOGY. Figure 2-4 shows the display of cross-reference terms resulting from EXPAND E5 (the reference number for TECHNOLOGY on Fig. 2-3). Desired terms can be selected using a command, "SELECT." Each selected term defines a subset of the total collection; namely, that subset which contains that term as an index term. Each subset is numbered and described on the remote character printer providing easy reference for the user.

The subsets corresponding to selected terms (or combinations of terms) can be combined with a powerful command called "COMBINE" to further partition the collection. COMBINE 1*2*3 (where 1, 2, and 3 are subsets of citations corresponding to the first three terms selected) results in a subset 4 which includes only those items containing all three of the specified terms. COMBINE operates on specified subsets with operators "and," "or," and "not," allowing the flexible and recursive construction of any arbitrary Boolean set. The power of this command is great. It allows the user to successively cut-and-try various search expressions. The system

```
                    EXPAND-TECHNOLOGY
REF       DESCRIPTOR                          CITATIONS REL. TERMS REF
E1    TECHNICAL                                     531           E1
E2    TECHNICAL DRAWING                                   1       E2
E3    TECHNICAL WRITING                               9           E3
E4    TECHNIQUE                                    6343       16  E4
E5   *TECHNOLOGY                                   4462        4  E5
E6    TECTONIC MOVEMENT                              50           E6
E7    TECTONICS                                      97           E7
E8    TEE                                            22           E8
E9    TEETERING                                       2           E9
ENTER NEXT COMMAND ►
```

Fig. 2-3 Display Resulting From EXPAND TECHNOLOGY

```
                    EXPAND-E5
REF       DESCRIPTOR                          CITATIONS REL. TERMS REF
E5   *TECHNOLOGY                                   4462        4  E5
E10   AEROSPACE TECHNOLOGY                         657           E10
E11   BIOTECHNOLOGY                                131           E11
E12   MILITARY TECHNOLOGY                          352        1  E12
E13   REACTOR TECHNOLOGY                           484           E13


ENTER NEXT COMMAND ►
```

Fig. 2-4 Display of Cross-Reference Terms Resulting From EXPAND E5

tells the user at each step how many items meet the specified criteria, and based on this feedback the searcher can narrow or broaden any expression to suit his requirement.

To assist the user in further evaluating the result of any search expression, several output commands are defined. DISPLAY causes a display of the specified items (citations or accession numbers) on the cathode ray tube display device; PRINT causes specified items to be printed on a high-speed printer; KEEP causes specified items to be saved in a special subset; and TYPE causes specified items (or their access numbers) to be typed on the IBM 1053 character printer at the terminal. Figure 2-5 shows displays of citations retrieved from the search expression:

(TECHNOLOGY or AEROSPACE TECHNOLOGY) and TRANSFER and INDUSTRY

Examination of displayed items (and the terms used to index these items) frequently suggests additional terms which should be included in the search expression. The step-wise approach to searching used in DIALOG allows easy modification of search strategy without repeating or regenerating the search.

## 2.2 COMMAND LANGUAGE MODIFICATION

Because each of the commands within DIALOG is an integral unit, modifications to individual command definitions were facilitated. It was also possible to introduce additional commands when the need for them was indicated.

Distributions were developed from actual search history execution times which indicated that with four terminals, mean waiting time per command would be between 1 and 2 seconds. Elimination of the TYPE command (use of remote PRINT capability) and a current change which causes output printing to go to tape rather than the on-line printer would reduce the wait time figure.

```
                              DISPLAY   6/2/1
65A31673      00/07/65      UNCLASSIFIED
 SPIN-OFF FROM SPACE.   (N;ASA INFORMATION SYSTEM TO ASSIST TRANSFER OF TECHNOLO
GICAL DATA FROM SPACE PROGRAMS TO POTENTIAL BENEFICIARIES)

   KERR, B. M.     /NASA, SCIENTIFIC AND TECHNICAL    INFORMATION DIV., WASHINGTO
N, D.C./. 203049   SCIENCE JOURNAL, VOL. 1, JUL. 1965, P. 85-90.
 KERR, B. M.
 / AEROSPACE/*AEROSPACE TECHNOLOGY/ DATA/ INDUSTRY/ INFORMATION/*INFORMATION RET
RIEVAL/*NASA PROGRAM/ PROGRAM/ RETRIEVAL/ SPACE/ TECHNOLOGY/ TITANIUM/ TRANSFER

ENTER NEXT COMMAND ▶


                              DISPLAY   6/2/2
65N16989#   NASA-CR-51214    NASR-162   00/06/63     UNCLASSIFIED
 AEROSPACE RESEARCH APPLICATIONS CENTER SUMMARY REPORT, 1 APRIL TO 30 JUNE 1963
  (AEROSPACE RESEARCH APPLICATIONS - CONFERENCE)
 WEIMER, A. M.
          INDIANA UNIV. FOUNDATION, BLOOMINGTON.
 / AEROSPACE/*AEROSPACE TECHNOLOGY/ APPLICATION/ COMMERCIAL/*CONFERENCE/ INDUSTR
Y/ NASA PROGRAM/ RESEARCH/ TRANSFER


ENTER NEXT COMMAND ▶


                              D ISPLAY   6/2/3
66N13375#   NASA-CR-68620 ER-SB-1844    NASW-1139   00/04/65     UNCLASSIFIED
 SPACE  TECHNOLOGY  APPLIED TO  MAN*S  EARTHLY  NEEDS -  A FEASIBILITY STUDY ON
THE TRANSFER OF AEROSPACE TE CHNOLOGY TO INDUSTRY USE   (FEASIBILITY STUDY ON ACC
ELERATING TRANSFER OF AE ROSPACE TECHNOLOGY TO COMMERCIAL INDUSTRY - AEROSPACE LI
TERATURE APPLICABILITY TO INDUSTRY)
 BROCK, A. W. DEMBICZAK, W. J. NAGY, A.
          AMERICAN MACHINE AND FOUNDRY CO., SANTA BARBARA, CALIF.
 / AEROSPACE/*AEROSPACE TECHNOLOGY/ APPLICATION/ COMMERCIAL/ EVALUATION/*INDUSTR
Y/ INFORMATION/*INFORMATION RETRIEVAL/ LITERATURE/ QUALITY/ RETRIEVAL/ SURVEY/ T
ECHNICAL/ TECHNOLOGY/ TRANSFER/ UTILIZATION
ENTER NEXT COMMAND ▶
```

Fig. 2-5 Display Examples

## 2.3 HARDWARE DESCRIPTION

The major hardware components of the IBM 360/30 include:

- Central processor containing 32,000 bytes of core (later upgraded to 64,000 bytes)
- 2 disk drives (IBM 2311) each with a capacity of 7.25 million bytes of storage
- 1 data cell drive (IBM 2321) with a capacity of 400 million bytes of storage
- Data adapter unit (IBM 2701)
- Remote terminal consisting of IBM 2848 control unit, IBM 2260 keyboard display, and IBM 1053 character printer
- IBM 1443 printer

The central processor itself contains 32,000 bytes of high-speed core storage memory, the arithmetic and logic unit (ALU), and the power supplies for the ALU and the channels. [A byte is 8 digital bits plus a parity (check) bit.] Programs and data are stored as magnetic patterns in the memory.

The IBM 2311 disk storage drive is the memory device in which most of the "files" are located and the units in which various sorting and rearrangements of data are accomplished. The 2311 disk storage drive provides random access storage for 7.25 million bytes (or 14.5 million packed decimal digits and signs) on a single disk pack. The data rate of the IBM 2311 is 156 thousand bytes per second. Sequential track-to-track access time is 30 milliseconds. The maximum seek time is 145 milliseconds, and the average seek time is 85 milliseconds. The average rotational delay is 12.5 milliseconds.

The IBM 2321 data cell drive (and data cells) is the bulk storage memory of the system. The IBM 2321 data cell drive extends on-line random access storage capabilities to a volume of data beyond that of other storage devices. Each 2321 offers 400 million bytes (or 800 million packed decimal digits and signs) of on-line data. The data cells

are all removable and interchangeable, permitting an open ended capacity for libraries of data cells. The storage medium is a strip of magnetic tape 2-1/4 inches wide by 13 inches long. Each data cell contains 200 of these strips, divided into 20 subcells of 10 strips each.

The IBM 2701 data adapter unit provides direct connection of a variety of remote and local external devices to an IBM 360. Together, the data adapter unit and a transmission adapter provide a single duplex (one way) or half-duplex (two ways alternately) data path, depending on the particular transmission adapter used, between external device(s) and an IBM 360.

The IBM 2848 display control contains the interface control, a character generator, and buffer storage for the display stations. This buffer storage retains the video data to maintain display regeneration on the CRT.

In addition, one IBM 1053 printer adapter unit may be attached to each display control. This feature allows use of an IBM 1053 printer at a central location to document and record periodic system transactions.

The IBM 1443 printer is the high-speed printer which produces the final output of the system. The IBM 1443 Printer Model N1 prints from 200 to 600 (maximum) lines per minute, depending on the number of characters in the set being used.

## Section 3
## SUMMARY OF SYSTEM USE

Two separate operational phases are described. These two phases differed substantially in terms of the manner in which the system was utilized. During the Ames Research Center Phase, all searching was performed by the primary customer; i.e., scientists and engineers interested in specific, detailed, and often subtle areas of technology. During the NASA Headquarters Phase, the terminal was used primarily by a NASA librarian to perform searches requested by other NASA personnel. The substantial difference between the two phases was that in the former case searches tended to be more specific with more items displayed and more selective printout, whereas in the latter case, the searches tended to be more generally formulated with less reliance on the display to edit results, and more massive printouts.

This section discusses the two phases with regard to search characteristics and user reaction. Summaries of individual searches from Phase 1 operation are included as Appendix B.

### 3.1 AMES RESEARCH CENTER PHASE

During the 2-month period (44 working days or about 80 terminal hours), 75 successful working search sessions were conducted. The actual number of search topics explored was somewhat greater, however, because many sessions included multiple search topics. In addition, there were several demonstrations and instruction searches.

### 3.1.1 User Statistics

Although some service was provided in all but 2 days, there were several interruptions in service during the 2-month period. These interruptions are summarized as follows:

| Type of Failure | Frequency of Occurrence | Approx. Time Lost (hr) | Percent of Total Time |
|---|---|---|---|
| Transmission line | 9 | 8 | 10.0 |
| IBM equipment | 6 | 7 | 8.7 |
| Application software | 5 | 4 | 5.0 |

In cases of failure, the interrupted search was regenerated either at the next scheduled period of operation, or by Lockheed personnel in a local mode of operation.

Time spent per search is distributed as follows:

| Elapsed Search Time (min) | Number of Searches | Percent of Searches |
|---|---|---|
| 0–10 | 11 | 12.4 |
| 11–20 | 16 | 18.0 |
| 21–30 | 10 | 11.2 |
| 31–40 | 19 | 21.3 |
| 41–50 | 12 | 13.5 |
| 51–60 | 11 | 12.4 |
| 61–70 | 10 | 11.2 |

Average minutes per search: 33

Much of this time included learning and practice time. The experienced searcher (after 2 to 3 tries) frequently can complete a relatively complex search in 8 to 10 minutes. In addition, search expressions tended to be both long in terms of the number of index terms included in the search expression, and complex in terms of the number of levels and concept groups in the Boolean expression satisfying the search topic.

| Number of Index Terms per Search Topic | Number of Searches | Percent of Searches |
|---|---|---|
| 1–5 | 33 | 34.4 |
| 6–10 | 39 | 40.6 |
| 11–15 | 15 | 15.6 |
| 16–20 | 4 | 4.2 |
| 21–25 | 5 | 5.2 |

Average index terms per search: 8

| Level of Boolean Expression | Number of Searches | Percent of Searches |
|---|---|---|
| 1 | 7 | 7.3 |
| 2 | 62 | 64.6 |
| 3 | 20 | 20.8 |
| 4 | 7 | 7.3 |

Average level of search: 2.3

| Number of Concept Groups | Number of Searches | Percent of Searches |
|---|---|---|
| 1 | 2 | 2.1 |
| 2 | 40 | 41.7 |
| 3 | 40 | 41.7 |
| 4 | 14 | 14.5 |

Average concept groups per search topics: 2.7

NOTE: The level of a Boolean expression measures the hierarchy of the expression, whereas the number of concept groups refers to the number of different groups of terms connected by "AND." For example, the expression (WELDING) AND (DEFECT OR FLAW) AND (ALUMINUM) is a two level, three expression search.

From these data we see that the average search time was 33 minutes, the average number of terms for search topics was 8, and the average level of search expression

was 2.3 and the average number of concept groups was 2.7. The total number of search topics does not agree in all cases because "elapsed search" time was lost in some instances.

The number of relevant items printed was as follows:

| Number of Items Printed | Number of Searches | Percent of Searches |
|---|---|---|
| 0–10 | 37 | 38.6 |
| 11–20 | 17 | 17.7 |
| 21–30 | 15 | 15.6 |
| 31–40 | 10 | 10.4 |
| 41–50 | 6 | 6.3 |
| 51–60 | 4 | 4.2 |
| 61–70 | 2 | 2.1 |
| 71–80 | 3 | 3.1 |
| 81–90 | 1 | 1.0 |
| 91–100 | 0 | 0.0 |
| >100 | 1 | 1.0 |

Average items printed per search: 22

From the distribution of items printed, it can be seen that most users printed less than 20 items. With on-line retrieval and the display capability, the user frequently prints only items which have been displayed and adjudged relevant. It follows, of course, that he does not print items which, although relevant to the search topic, are already familiar to him. The longer printouts frequently resulted when, after extensive sampling within a set, a user simply decided to print the entire set for review.

### 3.1.2 User Reaction

User reaction can be measured in several ways:

- Relevance versus precision
- Repeat usage
- Testimonial
- System demand

The traditional measures used to evaluate information retrieval systems are recall and precision. Recall refers to the number of relevant items retrieved compared with the total number of relevant items in the file; precision refers to the number of relevant items retrieved compared with the total number of items retrieved. With a recursive, on-line system such as DIALOG, the user successively modifies his search expression until he is satisfied with the relevance and precision of the resulting file. He then prints out selected items from the resulting file. The items printed out represent relevant items, but also represent items with which the user is not already familiar. In other words, they represent net additions to his knowledge in the area. As might be expected, there were large individual differences in the number of items printed as compared with items displayed. In a sample of 10 searches, the percentage of items kept per items displayed varied from 100% in two cases to 40% in the worst case. (The average was 58%.) Realizing that the searchers were professionals in the subjects, these percentages appear to represent substantial contributions.

Repeat usage, in the case of the 2-month Ames operation, is not a valid indicator of user reaction. On the one hand, there were no file updates during this period so that repetitions represented either searches in different subject areas or refinements of previous searches. Experience in this regard, however, is summarized below:

| Number of Search Sessions per User | Number of Occurrences | New Topic | Same Topic (Refinement) |
|---|---|---|---|
| 1 | 37 | 37 | – |
| 2 | 11 | 15 | 7 |
| 3 | 4 | 12 | 0 |
| 4 | 1 | 1 | 3 |

It can be seen that most repeated searches involved new topics. The last case above represented a searcher who was generally interested in all Apollo programs. He eventually printed a long bibliography in this area.

User comments and reactions were solicited as part of the search procedure and are summarized below. (The detailed comments can be found in Appendix B.)

| Comment | Positive | Negative | Not Discussed |
|---|---|---|---|
| 1. Usefulness of particular search | 37 | 7 | 31 |
| 2. Desirability of having DIALOG as a search tool | 36 | 0 | 39 |
| 3. Ease of learning and/or use | 19 | 4 | 52 |
| 4. DIALOG vs. batch or manual search | 20 | 2 | 53 |

Demand for use was substantial once availability was announced. Although only 2 hours per day (8:00−10:00 AM) were available for searching, and an observation session was required before a user could "solo," there was an average backlog of 1−2 weeks toward the end of the 2-month period of operation.

That users would expend their time reading the manual, observing a search session, and would spend 30−60 minutes per search session at a terminal operating the system knowing the service was to be terminated and the file would not be updated, would seem to indicate a positive reaction to the DIALOG system.

### 3.2 NASA HEADQUARTERS PHASE

During the year that the terminal was installed in Washington, D. C., approximately 300 searches were conducted, of which 174 are used in the analysis which follows.*

Service was provided on all but 9 days during the one-year period of operation (approximately 300 hours of available time). Failures are summarized as follows:

*Only those searches conducted by library personnel are included in this analysis.

| Type of Failure | Frequency of Occurrence | Approx. Time Lost (hr) | Percent of Total Time |
|---|---|---|---|
| Transmission line | 7 | 13 | 4.3 |
| IBM equipment | 25 | 23 | 7.7 |
| Application software | 30 | 3 | 1.0 |

During later phases of the operation, failures were nil due to additional error-checking software, and inclusion of the soft restart capability in DIALOG.

The number of terms per search expression was distributed as follows:

| Number of Index Terms per Search | Number of Searches | Percent of Searches |
|---|---|---|
| 1–5 | 104 | 59.8 |
| 6–10 | 40 | 23.0 |
| 11–15 | 14 | 8.0 |
| 16–20 | 9 | 5.2 |
| 21–25 | 7 | 4.0 |

Average index terms per search: 6.5

Another measure of the complexity of the search expression is the level* of the Boolean expression which was distributed as follows:

| Level of Boolean Expression | Number of Searches | Percent of Searches |
|---|---|---|
| 1 | 62 | 35.6 |
| 2 | 78 | 44.8 |
| 3 | 25 | 14.4 |
| 4 | 9 | 5.2 |

Average level of search: 1.9

*The level number is one greater than the largest number of successive parentheses employed.

The number of concept groups per search was distributed as follows:

| Number of Concept Groups | Number of Searches | Percent of Searches |
|---|---|---|
| 1 | 66 | 37.9 |
| 2 | 52 | 29.9 |
| 3 | 31 | 17.8 |
| 4 | 25 | 14.4 |

Average concept groups per search: 2.1

The number of items printed was distributed as follows:

| Number of Items Printed | Number of Searches | Percent of Searches |
|---|---|---|
| 0–10 | 118 | 67.8 |
| 11–20 | 8 | 4.6 |
| 21–30 | 9 | 5.2 |
| 31–40 | 10 | 5.8 |
| 41–50 | 3 | 1.7 |
| 51–60 | 3 | 1.7 |
| 61–70 | 6 | 3.4 |
| 71–80 | 3 | 1.7 |
| 81–90 | 1 | 0.6 |
| 91–100 | 1 | 0.6 |
| >100 | 12 | 6.9 |

Average items printed per search: 22.4

## 3.3 COMPARATIVE ANALYSIS OF AMES AND HEADQUARTERS SEARCH STATISTICS

In the Ames environment most searches were conducted by the engineer or scientist with the information need. In the Headquarters environment, most searches were conducted by library personnel in response to expressed needs. Our hypothesis is

that searches conducted by users of information will differ from those conducted by an intermediary. Specifically, the searches conducted by the information users are apt to be more specific, more complex in structure, and to cause fewer items to be printed out than those conducted by an intermediary. This hypothesis is based on the assumption that communicating a complex notion to another person is more difficult than recognizing descriptive abstracts of documents which closely relate to that complex notion.

Table 3-1 presents a summary of the search statistics from the two environments.

Table 3-1

SUMMARY OF SEARCH CHARACTERISTICS FROM AMES AND HEADQUARTERS SEARCHES

| Characteristic | Ames | Headquarters |
|---|---|---|
| Average Number of Terms | 8.0 | 6.5 |
| Average Level of Search | 2.3 | 1.9 |
| Average Number of Concepts | 2.7 | 2.1 |
| Average Number of Items Printed | 22.0 | 22.4 |

A superficial review of this table seems to bear out the above hypothesis. A more detailed analysis of these four characteristics was carried out as follows.

For each characteristic we are given the distribution of that characteristic for 96 searches at Ames and 174 searches at NASA Headquarters. We are interested in whether the population distributions of these two groups are identical, and if not, which classes differ significantly.

To describe the statistical technique employed, let $n_{11}, n_{21}, \ldots, n_{k1}$ be the allocation of observations among k classes for the Ames group for a particular characteristic. Similarly, let $n_{12}, n_{22}, \ldots, n_{k2}$ be the corresponding allocation for the Headquarters group. Let

$$N_1 = \sum_{i=1}^{k} n_{i1} = 96 \quad \text{and} \quad N_2 = \sum_{i=1}^{k} n_{i2} = 174$$

be the sample sizes for the two groups. Define $n_{i0} = n_{i1} + n_{i2}$ and $En_{i1} = N_1 n_{i0}/(N_1 + N_2)$. When the true probabilities of observations attaining values in the $i^{th}$ class are equal for the two groups, the statistic,

$$Z_i = \frac{n_{i1} - N_1 n_{i0}/(N_1 + N_2)}{\sqrt{N_1 n_{i0}/(N_1 + N_2)}}$$

is approximately normally distributed with mean zero and variance one. The probabilities for the $i^{th}$ class will be said to be significantly different if the absolute value of $Z_i$ is sufficiently large. (Largeness is based on quantiles of the normal distribution.)

When the probabilities of observations attaining values in the $i^{th}$ class are equal for the two groups for every class, then the statistic,

$$\chi^2_{2k-2} = 2 \sum_{i=1}^{k} Z_i^2$$

has approximately a chi-square distribution with $2k - 2$ degrees of freedom. The two distributions are significantly different if $\chi^2_{2k-2}$ is large relative to certain quantiles of the chi-square distribution.

Table 3-2 is an analysis of the data for each characteristic. An asterisk denotes significance at the 5% level.

Table 3-2

ANALYSIS OF AMES AND HEADQUARTERS SEARCH STATISTICS

Index Terms per Search Topic

| Index Terms | $n_{i1}$ | $n_{i2}$ | $En_{i1}$ | $Z_i$ |
|---|---|---|---|---|
| 1–5 | 33 | 104 | 48.7 | -2.25* |
| 6–10 | 39 | 40 | 28.1 | +2.06* |
| 11–15 | 15 | 14 | 10.3 | +1.46 |
| 16–20 | 4 | 9 | 4.6 | -0.28 |
| 21–25 | 5 | 7 | 4.3 | +0.34 |
| | 96 | 174 | 96.0 | $\chi^2_8 = 33.26$* |

Boolean Expression

| Expression | $n_{i1}$ | $n_{i2}$ | $En_{i1}$ | $Z_i$ |
|---|---|---|---|---|
| 1 | 7 | 62 | 24.5 | -3.54* |
| 2 | 62 | 78 | 49.8 | +1.73 |
| 3 | 20 | 25 | 16.0 | +1.00 |
| 4 | 7 | 9 | 5.7 | +0.54 |
| | 96 | 174 | 96.0 | $\chi^2_6 = 33.63$* |

Concept Groups

| Groups | $n_{i1}$ | $n_{i2}$ | $En_{i1}$ | $Z_i$ |
|---|---|---|---|---|
| 1 | 2 | 66 | 24.2 | -8.50* |
| 2 | 40 | 52 | 32.7 | +1.28 |
| 3 | 40 | 31 | 25.2 | +2.95* |
| 4 | 14 | 25 | 13.9 | +0.03 |
| | 96 | 174 | 96.0 | $\chi^2_6 = 165.18$* |

Items Printed

| Items Printed | $n_{i1}$ | $n_{i2}$ | $En_{i1}$ | $Z_i$ |
|---|---|---|---|---|
| 0–10 | 37 | 118 | 55.1 | -2.44* |
| 11–20 | 17 | 8 | 8.9 | +2.72* |
| 21–30 | 15 | 9 | 8.5 | +2.23* |
| 31–40 | 10 | 10 | 7.1 | +1.09 |
| 41–60 | 10 | 6 | 5.7 | +1.80 |
| 61–100 | 6 | 11 | 6.0 | -0.00 |
| 101+ | 1 | 12 | 4.7 | -1.68 |
| | 96 | 174 | 96.0 | $\chi^2_{12} = 51.15$* |

*Significance at 5% level.

The analysis of the data shows an overall statistically significant difference in the distribution corresponding to the two groups for each characteristic.

For the index terms per search topic, the Ames group tends to have smaller probabilities toward the lower tail, higher probabilities in the middle range, and approximately even probabilities in the upper range, in comparison with the Headquarters group. This indicates that for most of the range, the Ames group tends to use larger numbers of index terms than does the Headquarters group.

For the Boolean expressions, the main differences occur with respect to expression one, in which the Ames group exhibits a significantly smaller proportion than at Headquarters.

Similar comments apply to the number of concept groups. Undoubtedly, the expression levels and concept groups are heavily correlated, and the similarity of results is not surprising.

For the Ames group, the items printed tend to be more compactly distributed away from extreme values, than the Headquarters group. This result was unanticipated, and has no ready explanation.

The foregoing analysis indicates that significant differences exist between the two user groups. The implication of this finding to the system designer is that an on-line system, if it is to achieve maximum effectiveness, should be designed to utilize the decision-making and pattern recognition capabilities of the human operator, and thus enable the engineer, scientist, or manager user to employ the computer as an extension of his intellect, not as a substitute for it.

## Appendix A

## FILE CONVERSION REPORT - NASA LINEAR FILE

SUMMARY

A program was written which reads NASA linear file blocks and reformats individual citation records, the formats for which are shown in Figures 1 and 2. The reformatted records are blocked onto successive tracks in the IBM 2321 data cell as indicated in Figure 3. A dictionary of accession numbers vs. data cell addresses is written on the IBM 2311 disk drive using ISFMS (Indexed Sequential File Management System) as illustrated in Figure 4. Checkpoint records are written on tape at the completion of loading of each reel of tape for restart purposes.

The reformatting results in an average compression of 28% if coded terms are eliminated (as they are in the reformatting), or 18% if coded terms are included. There were no particular problems encountered in reformatting the NASA linear file except for accession 63A20145 which was garbled by inclusion of a portion of 63A20141, and two tape redundancies which made blocks containing records 64X12674-6 and 66N22475-6 unreadable. In these three cases, the problem records were ignored. Several equipment and software problems were encountered and surmounted as a result of diligence and persistent effort on the part of the LMSC file conversion programmer and the IBM systems representative.

Data cell load time approximates 90 minutes/reel (15,000 citations/reel). The large amount of record scanning and checking accounts for perhaps 60% of this time. Usefulness of the IBM 2321 Data Cell is improving now that several earlier problems have been eliminated. Many of the difficulties relate to early use of new equipment. Free oxides on new tape strips accumulated on read/write heads, hydraulic pressure failure, and strip initialization problems seem to have decreased significantly. Because we have no experience regarding prolonged use of the device, judgment on this aspect of performance will be deferred.

## Basic NASA Linear File Record

```
6E C$-=14=63N=16459=096111 =283=1222                170563      2  0                    0006      1=     =
  =0045=     =0117=0356=0432=0448=0464=0476=0758$ESTIMATE OF NEUTRON ALBEDO ON MOON SURFACE DUE T=UCO
SMIC RAY BOMBARDMEN=T$40N63-16459  ROCHESTER U., N.Y. ESTIMATE OF NEUTRO=N41ALBEDO ON THE MOON*S SUR
FACE DUE TO COSMIC RA=Y42BOMBARDMENT M. V. KRISHNA APPA RAO  MAY 17, 196=3436P  6 REFS SUBMITTED FOR
 PUBLICATION /CONTRAC=T44AT/30-1/-875/ /NYO-10265=/$=RAO, M. V. K. A.                $=RW983420ROCHEST
ER UNIV., N. Y.                                $=NYO-10265      $=AT/30-1/-875    $=3ALBEDO     =1BASALT
    =3BOMBARDMENT          =1CHARGE     =1CHONDRITE =1COMPOSITION          =1COSMIC     =3COSMIC RADIA
TION      =1ENERGY     =1FLARE      =1HELIUM     =1INTENSITY =1INTERACTION          =3LUNAR SURFACE
   =1MOON       =3NEUTRON    =1NUCLEUS    =1PROTON     =1SPALLATION=1SUN        =1SURFACE    $=1= OSMI=1=$
% )N=1=)NERG=1=)W @C=1=%ELIU=3=%X0*$=1=A6CLE=1=D1E D=1=FLARE=1=JASAL=1=JJRFA=1=LD856=1=MOON =3=SE4Y
=1=SUN   =3=XLBED=1=2ROTO=3=3 UTR=1=4 G   =1=6HARG=3=8S 7U$-
```

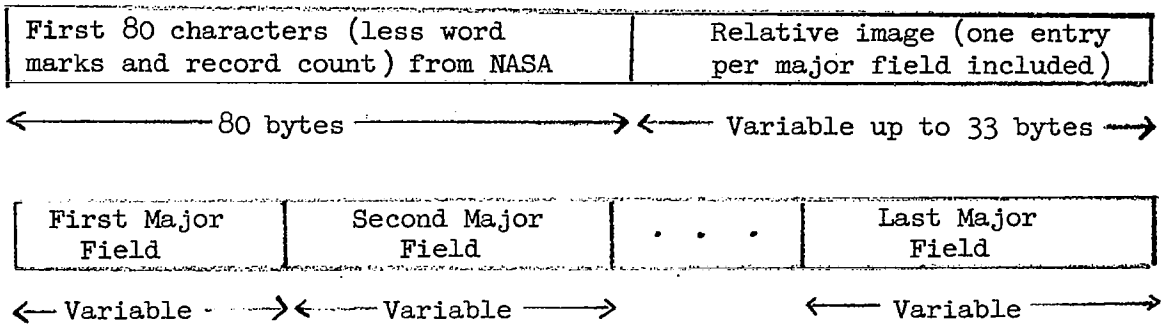
## Reformatted Linear File Record, Character and Hexadecimal Dump

```
1463N16459096111 2831222                170563      2  0                    0006      1C YE 1EAOGAXHAFIA-_A
) 8% [ESTIMATE OF NEUTRON ALBEDO ON MOON SURFACE DUE TO COSMIC RAY BOMBARDMENT][40N63-16459  ROCHESTER
 U., N.Y. ESTIMATE OF NEUTRON 41ALBEDO ON THE MOON*S SURFACE DUE TO COSMIC RAY 42BOMBARDMENT M. V. K
RISHNA APPA RAO  MAY 17, 1963 436P  6 REFS SUBMITTED FOR PUBLICATION /CONTRACT 44AT/30-1/-875/ /NYO-
10265[RAO, M. V. K. A][RW983420ROCHESTER UNIV., N. Y.][NYO-10265][AT/30-1/-875][3ALBEDOH1BASALT(3BOMBA
RDMENTH1CHARGE.1CHONDRITE(1COMPOSITIONH1COSMICK3COSMIC RADIATIONH1ENERGYG1FLAREH1HELIUM.1INTENSITY(1
INTERACTION 3LUNAR SURFACEF1MOONI3NEUTRONI1NUCLEUSH1PROTONH1SPALLATIONE1SUN 1SURFACE
```

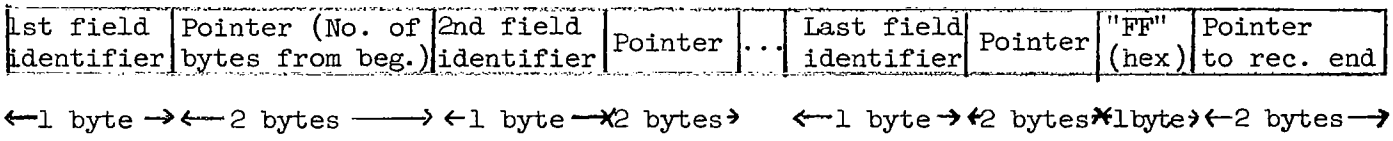
```
F1F4F6F3D5F1F6F4F5F9F0F9F6F1F1F140F2F8F3F1F2F2F2404040404040404040  FIXED HEADER
4040404040F1F7F0F5F6F3404040404040F24040F0404040404040404040404040  (80 BYTES)
4040404040F0F0F0F6404040404040F1030068050008106019607010A708016C609  REL. IMAGE
01D00A01DDEF02AC00C5E2E3C9D4C1E3C540D6C640D5C5E4E3D9D6D540C1D3C2
C5C4D640D6D540D4D6D6D540E2E4D9C6C1C3C540C4E4C540E3D640C3D6E2D4C9  03
C340D9C1E840C2D6D4C2C1D9C4D4C5D5E3F0F4F0D5F6F360F1F6F4F5F94040D9
D6C3C8C5E2E3C5D940E4486B40D548E84B40C5E2E3C9D4C1E3C540D6C640D5C5
E4E3D9D6D540F4F1C1D3C2C5C4D640D6D540E3C8C540D4D6D6D55CE240E2E4D9
C6C1C3C540C4E4C540E3D640C3D6E2D4C9C340D9C1E840F4F2C2D6D4C2C1D9C4  05
D4C5D5E340D44B40E54B40D2D9C9E2C8D5C140C1D7D7C140D9C1D64040D4C1E8
40F1F76B40F1F9F6F340F4F3F6D74040F640D9C5C6E240E2E4C2D4C9E3E3C5C4
40C6D6D940D7E4C2D3C9C3C1E3C9D6D54061C3D6D5E3D9C1C3E340F4F4C1E361
F3F060F16160F8F7F5614061D5E8D660F1F0F2F6F56[00D9C1D66B40D44B40E5  06
4B40D24B40C14[00D9E6F9F8F3F4F2F0D9D6C3C8C5E2E3C5D940E4D5C9E54B6B  07
40D54B40E84[00D5E8D660F1F0F2F6F[00C1E361F3F060F16160F8F7F[08F3C1  08 09
D3C2C5C4D[08F1C2C1E2C1D3E300F3C2D6D4C2C1D9C4D4C5D5E308F1C3C8C1D9
C7C508F1C3C8D6D5C4D9C9E3C500F1C3D6D4D7D6E2C9E3C9D6D508F1C3D6E2D4
C9C312F3C3D6E2D4C9C340D9C1C4C9C1E3C9D6D508F1C5D5C5D9C7E807F1C6D3  0A
C1D9C508F1C8C5D3C9E4D40BF1C9D5E3C5D5E2C9E3E800F1C9D5E3C5D9C1C3E3
C9D6D50FF3D3E4D5C1D940E2E4D9C6C1C3C506F1D4D6D6D509F3D5C5E4E3D9D6
D509F1D5E4C3D3C5E4E208F1D7D9D6E3D6D50CF1E2D7C1D3D3C1E3C9D6D505F1
E2E4D500F1E2E4D9C6C1C3C5
```

| FIELD | DESC. | START BYTE |
|---|---|---|
| 03 | NOC | 104 |
| 05 | DESC. NOTE | 177 |
| 06 | PERS. AUTH. | 406 |
| 07 | CORP. SOURCE | 423 |
| 08 | RPT. NO. | 454 |
| 09 | CONT. NO. | 464 |
| 0A | VOCAB. TERMS | 477 |
| FF | END OF REC. | 684 |

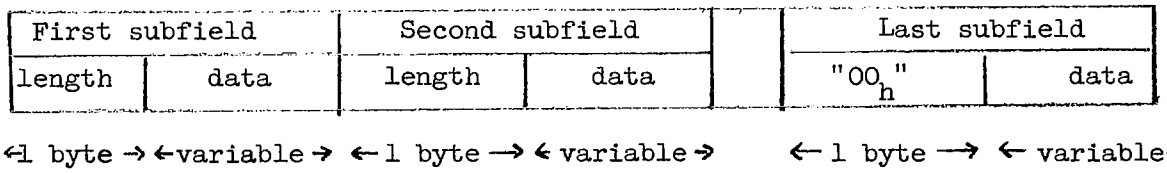
Fig. 1 Example of Basic and Reformatted Linear File Record




Reformatted Record Layout



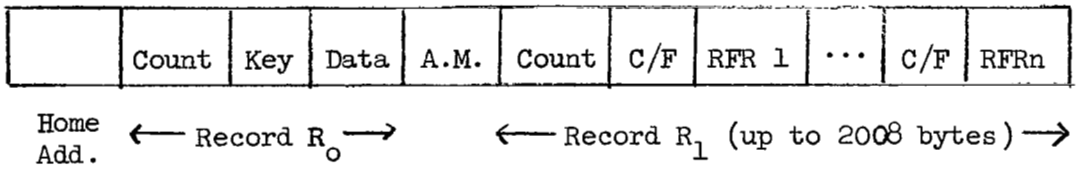

Relative Image Detail



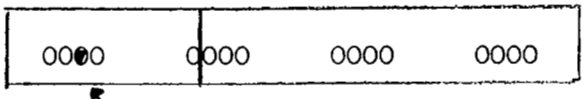

Length of field N = (value of field N+1 pointer) - (value of field N pointer)

Major Field Detail

Fig. 2 Schematic of Reformatted Record




Track Layout - One Physical Record/Track



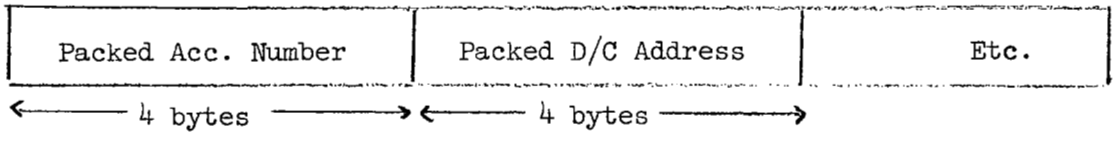

CF Field Detail

RFR - Reformatted logical record

Flag Bit - ON: RFR continued next track

OFF: RFR completed this track

Fig. 3. Record Blocking on the 2321 Data Cell




Track Layout (99 items/record, 4 records/track)

| cell | s/cell | strip | cyl. | head | key lgth. | rec no. | start pos | Item |
|---|---|---|---|---|---|---|---|---|
| 4 | 5 | 4 | 3 | 5 | 0 | 0 | 11 | Bits alloted |

Data Cell Address Packing

Fig. 4 Record Layout - ISFMS Index (2311 disk)

PROGRAM DESCRIPTION

This section contains a point by point description of the appended, annotated program. Figures 5 and 6 respectively show dumps of a 2321 data cell block (reformatted linear file entries) and a 2311 disk index block (accession numbers to data cell addresses).

Orientation of 1401 input data to 360 type output data is accomplished. This includes removal of wordmarks, packing of fields (removal of blanks in several places) and the establishment of binary (count) pointers to distinct fields within the (input) record. Output is written on successive tracks of the 2321 data cell at up to 2000 bytes of data per track. A dictionary (DICT) of accession numbers versus data cell addresses is written on a reserve disk pack on the 2311 disk drive. This is written with ISFMS (Indexed Sequential File Management System) ... a part of the BOS 8K operating system ... and occupies about 75 cylinders. On our IBM 360 model 30, reel run time averages 90 minutes. A succession of checkpoint records is constructed on a one entry per reel basis; said record being written after the previous one on a special tape mounted at the end of processing of each input reel (following its removal). These reflect the machine status at reel-end time and permit restarting the job with any reel. The 2311 disk pack is periodically copied to tape as a protection against loss. The data cell is removed and replaced with a dummy cell when filled.

The area between statements 844 and 948 was conceived, and largely coded, by a former employee and uses a rigorous character by character approach to the problem. Others might argue that the TRT (translate and test) plus multi-character move operations constitute the proper approach.

Statements 750 to 758 initialize basic program parameters. Reel 1 processing starts at START (711); other reels begin at RSTART (779). OPEN DICT (759) and SETFL DICT (768) cause the 75 cylinders of the 2311 pack to be preformatted in a suitable way to receive the KEY/DATA pairs, referenced later in this description. MTSOS (784) advises the operator of the reel # of the tape that must now be mounted in order to continue processing and waits for him to advise that this has been completed.

```
DUMP FROM 2321, UNIT X'193' ON 09/05/66

VOLUME LABEL FOLLOWS

 VOL(C)   0000   VOL1LMSC-10    D

 VOL(H)   0000   E5D6D3F1   D3D4E2C3   60F1F000   00000004   40404040   40404040   40404040   40404040
          0020   40404040   40404040   40404040   40404040   40404040   40404040   40404040   40404040
          0040   40404040   40404040   40404040   40404040

 HA-RO    0000   000C0703   050C0703   05000000   080C0703   050007D0   00

CELL = 001, SBCELL = 012, STRIP = 007, CYL = 003, TRACK = 005, REC = 001, KL = 000, DL = 2000

 COUNT    0000   0C070305   010007D0

 DATA(C)  0000    HME 1WAVER%14[63N16459]096111 2831222          170563    2  0              0006     1C YE IFA
          0100   OGAXHAFIA- A) 8% ESTIMATE OF NEUTRON ALBEDO ON MOON SURFACE DUE TO COSMIC RAY BOMBARDMENT 40N63-1645
          0200   9  ROCHESTER U., N.Y. ESTIMATE OF NEUTRON 41ALBEDO ON THE MOON*S SURFACE DUE TO COSMIC RAY 42BOMBARD
          0300   MENT M. V. KRISHNA APPA RAO  MAY 17, 1963 436P  6 REFS SUBMITTED FOR PUBLICATION /CONTRACT 44AT/30-1
          0400   /-875/ /NYO-10265/ RAO, M. V. K. A. RH98342OROCHESTER UNIV., N. Y. NYO-10265 AT/30-1/-875H3ALBEDOH18
          0500   ASALT(3BOMBARDMENTH1CHARGE.1CHONORITE11COMPOSITIONH1COSMICK3COSMIC RADIATIONH1ENERGYG1FLAREH1HELIUM.
          0600   1INTENSITY(1INTERACTION 3LUNAR SURFACEF1MOONI3NEUTRONI1NUCLEUSH1PROTONG1SPALLATIONE1SUN 1SURFACEC%14
          0700   [63N16460]20211 1531222          000263    1  3          0092  1   1C YE 7FBRGBVHBV1B  B
          0800   C% CALORIMETRIC ENTHALPY PROBE USING EVAPORATIVE FILM COOLING TO MEASURE GAS COMPOSITION AND ENERGY
          0900   BALANCE - TECHNIQUE TO MEASURE TOTAL ENTHALPY 40N63-16460  CORNELL AERONAUTICAL LAB., INC., 41BUFFAL
          1000   O, N.Y. AN EVAPORATING FILM CALORIMETRIC 42ENTHALPY PROBE  0FINAL REPORT, NOV. 1961 - DEC. 431962G F
          1100   . C. HAAS  WRIGHT-PATTERSON AFB, OHIO, 44AERONAUTICAL RESEARCH LABS., FEB. 1963  92P  127 45REFS /CO
          1200   NTRACT AF 33/657/-7774/ /ARL-63-47/ HAAS, F. C.2C5101763CORNELL AERONAUTICAL LAB., INC., BUFFALO, C5
          1300   101763N. Y. ARL-63-47 AF 33/657/-7774E1ARC(3CALORIMETRY(1COMPOSITIONG1CONVECTIONI1COOLINGH1ENERGY 3E
          1400   NTHALPY(1ENVIRONMENT(1EVAPORATIONN3EVAPORATION COOLINGF1FILM+3FILM COOLINGF1FLUXE1GASF1HEATK1HIGH TE
          1500   MPERATUREG1LEVEL(1MEASUREMENT.1OPERATIONG3PROBEH1SAMPLEH1STREAM 1STRUCTURE/Y14[63N16461]090311 1231212
          1600                 220463    1  3              0087     1C YE 1FAZGBLHBH BM B5 MEAN ELECTRON DENSITY
          1700   PROFILES OF IONOSPHERE FROM VERTICAL SOUNDING FLIGHTS 08CENTRAL RADIO PROPAGATION LAB. 40N63-16461
          1800   NATIONAL BUREAU OF STANDARDS.  CENTRAL 41RADIO PROPAGATION LAB., BOULDER, COLO. MEAN 42ELECTRON DENS
          1900   ITY VARIATIONS OF THE QUIET 43IONOSPHERE NO. 9 - NOVEMBER 1959 J. W. WRIGHT, L. 44R. WESCOTT, AND D.

                                                  OF FIELD          ACCESSION NUMBER
 DATA(H)  0000   0008D4C5   00F1E6C1   E5C5[D2AC]   F1F4[F6F3]   D5F1F6F4   F5F[9]F0F9   F6F1F1F1   40F2F8F3
          0020   F1F7F2F2   40404040   40404040   40404040   40F1F7F0   F5F6F340   40404040   40F24040
          0040   F0404040   40404040   40404040   40404040   40F0F0F0   F6404040   404040F1   03006805
          0060   00810601   9607D1A7   D801C6D9   01D00A01   DDFF02AC   00C5E2E3   C9D4C1E3   C540D6C6
          0080   40D5C5E4   E3D9D6D5   40C1D3C2   C5C4D640   D6D540D4   D6D6D540   E2E4D9C6   C1C3C540
          00A0   C4E4C540   E3D640C3   D6E2D4C9   C340D9C1   E840C2D6   D4C2C1D9   C4D4C5D5   E300F4F0
          00C0   D5F6F360   F1F6F4F5   F9404000   D6C3C8C5   E2E3C5D9   40E44B6B   40D54BE8   4B40C5E2
          00E0   E3C9D4C1   E3C540D6   C6400000   E4E3D9D6   D540F4F1   C1D3C2C5   C4D640D6   D540E3C8
          0100   C540D4D6   D6D55CE2   40E2E4D9   C6C1C3C5   40C4E4C5   40E3D640   C3D6E2D4   C9C340D9
          0120   C1E840F4   F2C2D6D4   C2C1D9C4   D4C5D5E3   40D44B40   E54B40D2   D9C9E2C8   D5C140C1
          0140   D7D7C140   D9C1D640   40D4C1E8   40F1F768   40F1F9F6   F340F4F3   F6D740D0   F640D9C5
          0160   C6E240E2   E4C2D4C9   E3E3C5C4   40C6D6D9   40D7E4C2   D3C9C3C1   E3C9D6D5   406IC3D6
          0180   D5E3D9C1   C3E340F4   F4C1E361   F3F060F1   6160F8F7   F5614061   D5E8D660   F1F0F2F6
          01A0   F56100D9   C1D66B40   D44B40E5   4B40D24B   40C14B00   D9E6F9F8   F3F4F2F0   D9D6C3C8
          01C0   C5E2E3C5   D940E4D5   C9E54B6B   40D54B40   E84B00D5   E8D660F1   F0F2F6F5   00C1E361
          01E0   F3F060F1   6160F8F7   F508F3C1   D3C2C5C4   D60AF1C2   C1E2C1D3   E30BF3C2   D6D4C2C1
          0200   D9C4D4C5   D5E308F1   C3C8C1D9   C7C50BF1   C3C8D6D5   C4D9C9E3   C50DF1C3   D6D407D6
          0220   E2C9E3C9   D6D508F1   C3D6E2D4   C9C312F3   C3D6E2D4   C9C340D9   C1C4C9C1   E3C9D6D5
          0240   08F1C5D5   C5D9C7E8   07F1C6D3   C1D9C508   F1C8C5D3   C9E4040B   F1C9D5E3   C5D5E2C9
          0260   E3E80DF1   C9D5E3C5   D9C1C3E3   C9D6D50F   F3D3E405   C1D940E2   E4D9C6C1   C3C506F1
          0280   D4D6D6D5   09F3D5C5   E4E3D9D6   D509F1D5   E4C3D3C5   E4E208F1   D7D9D6E3   D6D50CF1
          02A0   E2D7C1D3   D3C1E3C9   D6D505F1   E2E4D500   F1E2E4D9   C6C1C3C5   D86CF1F4   [F6F3D5F1]
          02C0   [E6E4E6F0]   F1F2F0F2   F1F140F1   F5F3F1F2   F2F24040   40404040   40404040   404040F0
          02E0   F0F0F2F6   F3404040   404040F1   404DF340   40404040   40404040   40404040   404040F0
          0300   F0F9F240   40404040   40F10300   680500F7   06021907   02250802   6509026F   0A027FFF
          0320   036C00C3   C1D3D6D9   C9D4C5E3   D9C9C340   C5D5E3C8   C1D3D7E8   40D7D9D6   C2C540E4
          0340   E2C9D5C7   40C5E5C1   D7D6D9C1   E3C9E9C5   40C6C9D3   D440C3D6   D6D3C9D5   C740E3D6
          0360   40D4C5C1   E2E4D9C5   40C7C1E2   40C3D6D4   D7D6E2C9   E3C9D6D5   40C1D5C4   40C5D5C5
          0380   D9C7E840   C2C1D3C1   D5C3C540   6040E3C5   C3C8D5C9   D8E4C540   E3D640D4   C5C1E2E4
          03A0   D9C540E3   D6E3C1D3   40C5D5E3   C8C1D3D7   E800F4F0   D5F6F360   F1F6F4F6   F04040C3
          03C0   D6D9D5C5   D3D340C1   C5D9D6D5   C1E4E3C9   C3C1D340   D3C1C24B   6B40C9D5   C3486B40
          03E0   F4F1C2E4   C6C6C1D3   D66B40D5   4BEB4840   C1D540C5   E5C1D7D6   D9C1E3C9   D5C740C6
          0400   C9D3D440   C3C1D3D6   D9C9D4C5   E3D9C9C3   40F4F2C5   D5E3C8C1   D3D7E840   D7D9D6C2
          0420   C540404C   C6C905C1   D340D9C5   D7D6D9E3   6B40D5D6   E54B40F1   F9F6F140   6040C4C5
          0440   C34B40F4   F3F1F9F6   F24C40C6   4B40C34B   40C8C1C1   E24040E6   D9C9C7C8   E360D7C1
          0460   E3E3C5D9   E2D6D540   C1C6C26B   40D6C8C9   D66B40F4   F4C1C5D9   D6D5C1E4   E3C9C3C1
          0480   D340D9C5   E2C5C1D9   C3C840D3   C1C2E24B   6B40C6C5   C24B40F1   F9F6F340   40F9F2D7
          04A0   4040F1F2   F740F4F5   D9C5C6E2   4061C3D6   D5E3D9C1   C3E340C1   C640F3F3   61F6F5F7
          04C0   6160F7F7   F7F46140   61C1D9D3   60F6F360   F4F76100   C8C1C1E2   6B40C64B   40C34B32
          04E0   C3F5F1F0   F1F7F6F3   C3D6D9D5   C5D3D340   C1C5D9D6   D5C1E4E3   C9C3C1D3   40D3C1C2
          0500   4B6B40C9   D5C34B6B   40C2E4C6   C6C1D3D6   6B00C3F5   F1F0F1F7   F6F3D54B   40E84B00
          0520   C1D9D360   F6F360F4   F700C1C6   40F3F361   F6F5F761   60F7F7F7   F405F1C1   D9C30DF3
          0540   C3C1D3D6   D9C9D4C5   E3D9E80D   F1C3D6D4   D7D6E2C9   E3C9D6D5   0CF1C3D6   D5E5C5C3
          0560   E3C9D6D5   09F1C3D6   D6D3C9D5   C708F1C5   D5C5D9C7   E80AF3C5   D5E3C8C1   D3D7E80D
          0580   F1C5D5E5   C9D9D6D5   D4C5D5E3   0DF1C5E5   C1D7D6D9   C1E3C9D6   D515F3C5   E5C1D7D6
          05A0   D9C1E3C9   D6D540C3   D6D6D3C9   D5C706F1   C6C9D3D4   0EF3C6C9   D3D440C3   D6D6D3C9
          05C0   D5C706F1   C6D3E4E7   05F1C7C1   E206F1C8   C5C1E312   F1C8C9C7   C840E3C5   D4D7C5D9
          05E0   C1E3E4D9   C507F1D3   C5E5C5D3   0DF1D4C5   C1E2E4D9   C5D4C5D5   E308F1D6   D7C5D9C1
          0600   E3C9D6D5   07F3D7D9   D6C2C508   F1E2C1D4   D7D3C508   F1E2E3D9   C5C1D400   F1E2E3D9
          0620   E4C3E3E4   D9C5[D1AB]   F1F4[F6F3]   [D5F1F6F4]   F6F1F0F9   F0F3F1F1   40F1F2F3   F1F2F1F2
          0640   40404040   40404040   40404040   40F2F2F0   F4F6F340   40404040   40F14040   F3404040
          0660   40404040   40404040   40404040   40F0F0F8   F7404040   404040F1   03006505   00810601
          0680   E9070213   0802480A   0254FF02   F500D4C5   C1D540C5   D3C5C3E3   D9D6D540   C4C5D5E2
          06A0   C9E3E840   D7D9D6C6   C9D3C5E2   40D6C640   C9D6D5D6   E2D7C8C5   D9C540C6   D9D6D440
          06C0   E5C5D9E3   C9C3C1D3   40E2D6E4   D5C4C9D5   C740C6D3   C9C7C8E3   E200F0F8   C3C5D5E3
          06E0   D9C1D340   D9C1C4C9   D640D7D9   D6D7C1C7   C1E3C9D6   D540D3C1   C24B40F4   F0D5F6F3
          0700   60F1F6F4   F6F14040   D5C1E3C9   D6D5C1D3   40C2E4D9   C5C1E440   D6C640E2   E3C1D5C4
          0720   C1D9C4E2   4B4040C3   C5D5E3D9   C1D340F4   F1D9C1C4   C9D64007   D9D6D7C1   C7C1E3C9
          0740   D6D540D3   C1C2486B   40C2D6E4   D3C4C5D9   6B40C3D6   D3D64B40   D4C5C1D5   40F4F2C5
          0760   D3C5C3E3   D9D6D540   C4C5D5E2   C9E3E840   E5C1D9C9   C1E3C9D6   D5E240D6   C640E3C8
          0780   C540D8E4   C9C5E340   F4F3C9D6   D5D6E2D7   C8C5D9C5   40D5D648   40F94060   40D5D6E5
          07A0   C5D4C2C5   D940F1F9   F5F940D1   4B40E64B   40E6D9C9   C7C8E36B   40D34B40   F4F4D94B
          07C0   40E6C5E2   C3D6E3E3   6B40C1D5   C440C44B
```

Fig. 5  Character and Hexadecimal Dump of Reformatted Data Cell Record

DUMP FROM 2311, UNIT X'191' ON 08/15/66

VOLUME LABEL FOLLOWS

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| VOL(C) | 0000 | VOL10947370 | D | | | EBSC E058K | | | |
| VOL(H) | 0000 | E50603F1 | F0F9F4F7 | F3F7F000 | 00000004 | 40404040 | 40404040 | 40404040 | 40404040 |
| | 0020 | 40404040 | 40404040 | 40404040 | C3400206 | 12FF0240 | 40404040 | 40404040 | 40404040 |
| | 0040 | 40404040 | 40404040 | 40404040 | 40404040 | | | | |
| HA-RO | 0000 | 00007700 | 02002700 | 02000000 | 08002700 | 02000E29 | 00 | | |

CYL = 039, TRACK = 002, REC = 001, KL = 004, DL = 0792

ACC. NUMBER: 3516459,63N16459

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| LOCATION | | | | | | | 0001 | 0110 | 0111 | 011 | 00101 |
| | | | | | | | 1 | 12 | 7 | 3 | 5 |
| | | | | | | | CELL | SKELL | STRIP | CYL | TRACK |
| COUNT | 0000 | 00270002 | 01040318 | | | | | | | | |
| KEY(H) | 0000 | 3516506F | | | | | | | | | |
| DATA(H) | 0000 | 3516408F | 163A203D | 3516409F | 163A232A | 3516410F | 163A2606 | 3516411F | 163A2944 |
| | 0020 | 3516412F | 163A28A8 | 3516413F | 163A3000 | 3516414F | 163A330B | 3516415F | 163A363E |
| | 0040 | 3516416F | 163A3910 | 3516417F | 163A3C7C | 3516418F | 163A4000 | 3516419F | 163A43E8 |
| | 0060 | 3516420F | 163A4800 | 3516421F | 163A4BDA | 3516422F | 163A4E74 | 3516423F | 163A52BF |
| | 0080 | 3516424F | 163A565A | 3516425F | 163A5967 | 3516426F | 163A5D37 | 3516427F | 163A613C |
| | 00A0 | 3516428F | 163A652C | 3516429F | 163A6BAD | 3516430F | 163A6B3A | 3516431F | 163A6E9E |
| | 00C0 | 3516432F | 163A719F | 3516433F | 163A74BB | 3516434F | 163A7800 | 3516435F | 163A7B02 |
| | 00E0 | 3516436F | 163A7DFC | 3516437F | 163A819B | 3516438F | 163A842F | 3516439F | 163A8800 |
| | 0100 | 3516440F | 163A8B4D | 3516441F | 163A9000 | 3516442F | 163A92D4 | 3516443F | 163A958F |
| | 0120 | 3516444F | 163A98D6 | 3516445F | 163A9BCF | 3516446F | 163A9FCC | 3516447F | 163B01A0 |
| | 0140 | 3516448F | 163B0443 | 3516449F | 163B0690 | 3516450F | 163B0A2B | 3516451F | 163B0D95 |
| | 0160 | 3516452F | 163B1150 | 3516453F | 163B1457 | 3516454F | 163B1800 | 3516455F | 163B1B4E |
| | 0180 | 3516456F | 163B1E62 | 3516457F | 163B2174 | 3516458F | 163B2468 | 3516459F | 163B280A |
| | 01A0 | 3516460F | 163B2AB8 | 3516461F | 163B2E26 | 3516462F | 163B314F | 3516463F | 163B342A |
| | 01C0 | 3516464F | 163B36B9 | 3516465F | 163B3ACC | 3516466F | 163B3D9D | 3516467F | 163B40C4 |
| | 01E0 | 3516468F | 163B43B1 | 3516469F | 163B460E | 3516470F | 163B496C | 3516471F | 163B4D2E |
| | 0200 | 3516472F | 163B50A4 | 3516473F | 163B545E | 3516474F | 163B56C7 | 3516475F | 163B5A26 |
| | 0220 | 3516476F | 163B5CCE | 3516477F | 163B601E | 3516478F | 163B6095 | 3516479F | 163B63C1 |
| | 0240 | 3516480F | 163B6800 | 3516481F | 163B6B61 | 3516482F | 163B6E44 | 3516483F | 163B70C5 |
| | 0260 | 3516484F | 163B73E9 | 3516485F | 163B7800 | 3516486F | 163B7B14 | 3516487F | 163B7F35 |
| | 0280 | 3516488F | 163B8207 | 3516489F | 163B8557 | 3516490F | 163B8894 | 3516491F | 163B8BA9 |
| | 02A0 | 3516492F | 163B8EFE | 3516493F | 163B9000 | 3516494F | 163B92CA | 3516495F | 163B9505 |
| | 02C0 | 3516496F | 163B98F3 | 3516497F | 163B9CDB | 3516498F | 163C0015 | 3516499F | 163C022C |
| | 02E0 | 3516500F | 163C0521 | 3516501F | 163C0829 | 3516502F | 163C0B19 | 3516503F | 163C0D9C |
| | 0300 | 3516504F | 163C112E | 3516505F | 163C13EF | 3516506F | 163C1800 | | |

Fig. 6 Dump of Disk Index Block

A record is read from tape at EXCP TAPE (822), one or more records are reformatted in the area mentioned above (844-948), and when 2000 bytes of output data are ready (or too little space is left in track image area to accommodate the field pointer area of still another record) program control reaches PACK (951). At this point, the EBCDIC representation of the accession number is altered in form (PACK operation) cutting its byte length in two. This item constitutes our dictionary KEY entry for this pair. There follows the construction (series of shift operations) resulting in the word DATA (986). It consists of 32 bytes divided as follows (left to right):

| | | |
|---|---|---|
| 1st 4 | cell number | |
| next 5 | subcell number | |
| next 4 | strip number | i.e., the 2321 address |
| next 3 | cylinder number | |
| next 5 | track number | |
| next 11 | relative (to start of track) position of start of reformatted record | |

KEY and DATA are the pair of entries presented to ISFMS at line 992. ISFMS has been directed (lines 10-13) to block these two 4 byte fields at 99 pairs per 2311 "record". This allows four records to be accommodated on each 2311 track.

MOVE (line 1029) is the subroutine that transfers all input characters (from IPAREA) to the output (reformatted) area at FRONT.

LIMIT and LIMIT2 (1045, 1046) are used in the process of determining when the field pointer count fields (of a record starting somewhere downstream in the 2321 track image area ... 1st 2000 bytes of FRONT) have been filled ... a process that requires processing the entire logical input record and may cause considerable overflow beyond the first 2000 bytes of the FRONT area. Hence the space reservation of 4000 bytes.

Data Cell records are prefixed with a C/F (count flag) area of two bytes which gives the byte count of this segment of an accession on the 2321 track and further indicates (if bit 3 of first such byte is on ... and only then) that the remainder of this record (i.e., it is not here complete) is to be

found on the next available track (Volume Label and Alternate Track areas are avoided).

The references to KSU (1597) and ESCAPE (1604) are associated with preventing undesired strip restoration to its subcell while reformatting is taking place ... the latter process frequently exceeds 800 ms, the maximum strip holding time of the 2321.

Routine SHIFT (at 1323) moves the data from that portion of a reformatted record that just exceeded the 2000 byte limit of the track image area back to the prime work area (at FRONT) once the data at FRONT has been written out on the data cell. This routine includes some coding (lines 1329 to 1339) of Mr. Keith Eckhardt of IBM ... one of their Systems Engineers ... whose valuable advice on ISFMS and data cell write operations have made this program possible. His assistance has been invaluable.

GERR (gross error, 1512) refers to a ≠ (record mark) being found in an unacceptable spot. 63A20145 is such a record.

Principal programmer for the file conversion and load routines is D. C. Shoultz. His conscientious devotion and thoroughness are responsible for the success of the conversion.

19 January 1967

# FILE CONVERSION REPORT - NASA INVERTED AND RELATED TERMS FILES

Contract NASw 1454


SUMMARY

A program was written to accomplish the following:

1. Load related terms records on data cell (Relative Master file).
2. Load inverted file records on data cell (V-post file).
3. Generate an index of Descriptor vs. Inverted File and Related Terms.

Figure 1 depicts the chronological processing sequence outlined above.

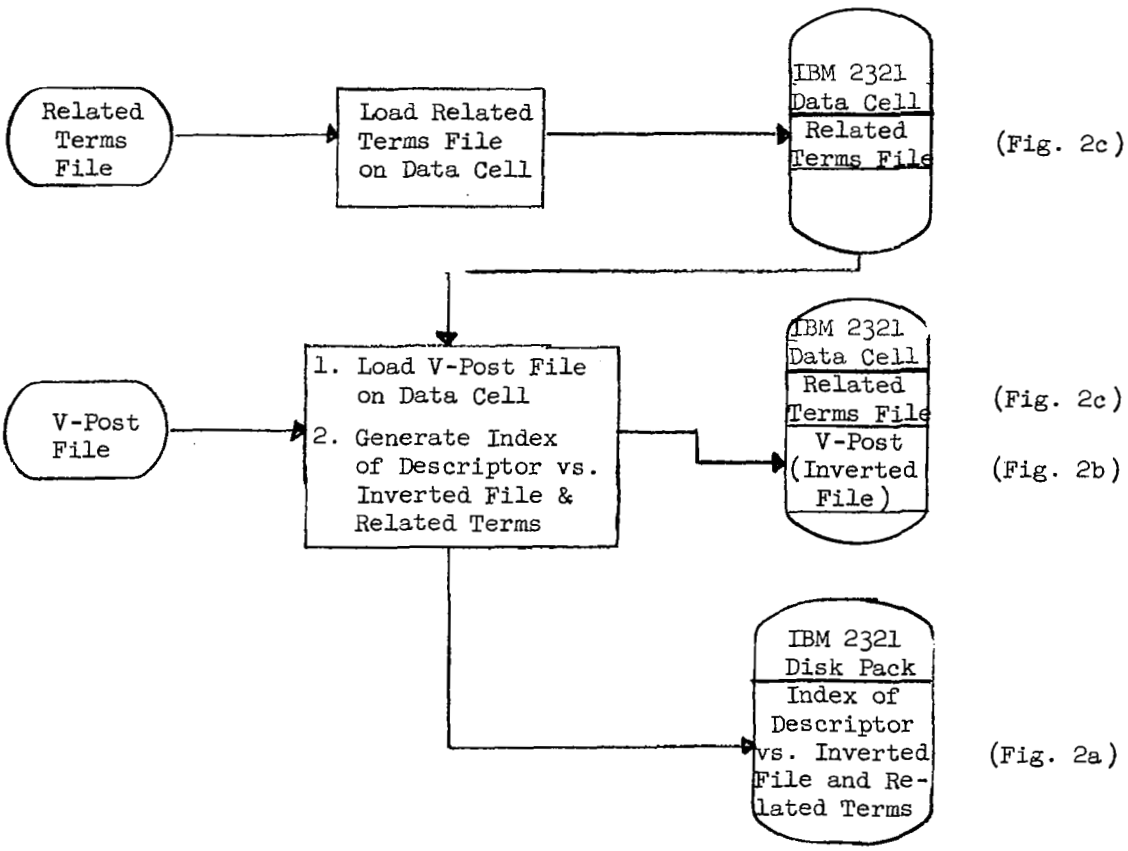




Fig. 1 Chronological Processing Chart

Figure 2 contains schematics of the referenced output records. Figures 3-5 are character and hexadecimal dumps of the referenced output records. An appendix contains an annotated listing of the subject program. Data Cell load time approximates 25 minutes/reel.

## PROGRAM DESCRIPTION

The related terms tape processing program is merely a tape reader, term counter and data cell writer (except that counts for a descriptor are for items on this track only and may necessitate repeating descriptor on upcoming track). As such it deserves little comment. The output format is that of Fig. 2c and is used by the inverted file program as needed. The reason for putting it on the 2321 prior to processing the V-post tapes is that this facility has only one tape unit.

The approach taken to the inverted file (V-post tapes) was that of:

1. extraction - of essential data from V-post series-descriptors and accessing
2. combination - of pointer to data in (1) with a pointer to related terms paired with the descriptor at hand. This is illustrated in Fig. 2a.

## OPERATING PROCEDURE

A. Required Configuration (in addition to CPU and storage controller)

2 each 2311 disk drives (units 190,191)
1 each 2321 data cell drive (unit 193)
1 each 2400 series 7 track tape unit (unit 180)
1 each 1052 typewriter (SYS004)

B. Operating Procedure:

1. Mount BOS pack (with 2321 error recovery procedures) on unit 190.
2. Mount (surface) initialized disk pack on 191.
3. Mount initialized cell 6 on 193.

4. Ready forms in printer, hex deck (supplied) in card reader. Press blue end-of-file button.

5. IPL from 190.

   Hit interrupt. Reply (small) c, E 0 B to SAR on 1052. Job should run to end of file on input tape (Label file is shipped). At EOF time message EOF A appears on 1052. Mount new reel and type (small) c to continue. To terminate job (close files), reply with something else, not supervisor processed, such as f.

a. IBM 2311 (Disk) Vocabulary Term Index

| Item | Descriptor | Data Cell Seek Add. for Inv. File Entry | Number of Accessions | Track Loc. of 1st Acc. | Data Cell Seek Add. for Rel. Terms Entry | No. of Rel. Terms | Track Loc. of 1st Rel. Terms Entry |
|---|---|---|---|---|---|---|---|
| Example | Zone | 0607050211 | 0617 | 0320 | 061204030C | 05 | 0066 |
| Bytes | 50 EBCDIC Char. | 5 hex char. | 2 hex char. | 2 hex char. | 5 hex char. | 1 hex char. | 2 hex char. |

b. IBM 2321 (Data Cell) Inverted File Entry (6-7-5-2-17)

| Item | Descriptor | | Accessions | | |
|---|---|---|---|---|---|
| Example | Zone | 0000 | 2510046F | 2510054F | etc. |
| Bytes | 50 | -2- | 4 | 4 | |

c. IBM 2321 (Data Cell) Related Terms Entry (6-18-4-3-12)

| Item | Count | Descriptor | Related Term 1 | Related Term 2 | etc. |
|---|---|---|---|---|---|
| Example | 0005 | Zone | Auroral Zone | Brillouin Zone | |
| Bytes | 2 | 50 | 50 | 50 | |

Fig. 2 Record Formats

```
DUMP FROM 2311, UNIT X'191' ON 12/31/66
VOLUME LABEL FOLLOWS

  VOL(C)   0000   VOL10947370   D                          LMSC BOS8K

  VOL(H)   0000   E5D6D3F1   F0F9F4F7   F3F7F000   00000004   40404040   40404040   40404040   40404040
           0020   40404040   40404040   40D3D4E2   C340C2D6   E2F8D240   40404040   40404040   40404040
           0040   40404040   40404040   40404040   40404040

  HA-R0    0000   00008B00   08008B00   08000000   08008B00   08000E29   00

CYL = 139, TRACK = 008, REC = 001, KL = 050, DL = 3484

  COUNT    0000   008B0008   01320D9C

  KEY(C)   0000   ZYMOGEN

  DATA(C)  0000   ZONAL HARMONICS                                    FGEBJ 1A-        ZOND  I SPACE PROBE
           0100                    FGEBJ ABQ        ZOND  II SPACE PROBE                              FGEBJ F8&
           0200    ZOND  III SPACE PROBE                    FGEBJ DB*        ZONE
           0300                    FGEBJFPC-FKDC□E WZONE MELTING                                      FGEC   D&
           0400     ZONE REFINING                                  FGEC   D@        ZONING TECHNIQUE
           0500                  FGEC   E%        ZOOLOGY                                            FGEC  OFHFKDC□
           0600   AAMZOOM                                    FGEC  GFM        ZPR
           0700                  FGEC  BGUFKDC□AA ZTA GRAPHITE                                      FGEC  DG-
           0800      ZUBOV PROPOSITION                              FGEC  DGU        ZUNI MISSILE
           0900                  FGECA I 4        ZYMOGEN                                            FGECA 8 □
           1000

  DATA(H)  0000   E9D6D5C1   D340C8C1   D9D4D6D5   C9C3E240   40404040   40404040   40404040   40404040
           0020   40404040   40404040   40404040   40404040   40400607   05021100   31016000   00000000
           0040   000000E9   D6D5C440   40C940E2   D7C1C3C5   40D7D9D6   C2C54040   40404040   40404040
           0060   40404040   40404040   40404040   40404040   40404040   40060705   02110001   02580000
           0080   00000000   0000E9D6   D5C44040   C9C940E2   D7C1C3C5   40D7D9D6   C2C54040   40404040
           00A0   40404040   40404040   40404040   40404040   40404040   40404040   06070502   11000602
           00C0   90000000   00000000   00E9D6D5   C44040C9   C9C940E2   D7C1C3C5   40D7D9D6   C2C54040
           00E0   40404040   40404040   40404040   40404040   40404040   40404040   40404006   07050211
           0100   000402DC   00000000   00000000   E9D5D5C5   40404040   40404040   40404040   40404040
           0120   40404040   40404040   40404040   40404040   40404040   40404040   40400607   FIG 2a
           0140   05021106   17032006   12040300   05000669   D6D5C540   D4C5D3E3   C9D5C740   40404040
           0160   40404040   40404040   40404040   40404040   40404040   40404040   40404040   40404040
           0180   40060705   0300001E   04500000   00000000   0000E9D6   D5C540D9   C5C6C9D5   C9D5C740
           01A0   40404040   40404040   40404040   40404040   40404040   40404040   40404040   40404040
           01C0   40404040   06070503   00002F04   FC000000   00000000   00E9D6D5   C9D5C740   E3C5C3C8
           01E0   D5C9D8E4   C5404040   40404040   40404040   40404040   40404040   40404040   40404040
           0200   40404040   40404006   07050300   000A05EC   00000000   00000000   E9D6D6D3   D6C7E840
           0220   40404040   40404040   40404040   40404040   40404040   40404040   40404040   40404040
           0240   40404040   40404040   40400607   05030000   16064806   1204030C   010194E9   D6D60440
           0260   40404040   40404040   40404040   40404040   40404040   40404040   40404040   40404040
           0280   40404040   40404040   40404040   40060705   03000007   06D40000   00000000   0000E9D7
           02A0   D9404040   40404040   40404040   40404040   40404040   40404040   40404040   40404040
           02C0   40404040   40404040   40404040   40404040   06070503   00000207   24061204   030C0101
           02E0   FAE9E3C1   40C7D9C1   D7C8C9E3   C5404040   40404040   40404040   40404040   40404040
           0300   40404040   40404040   40404040   40404040   40404006   07050300   00040760   00000000
           0320   00000000   E9E4C2D6   E540D7D9   D6D7D6E2   C9E3C9D6   D5404040   40404040   40404040
           0340   40404040   40404040   40404040   40404040   40404040   40400607   05030000   0407A400
           0360   00000000   000000E9   E4D5C940   D4C9E2E2   C9D3C540   40404040   40404040   40404040
           0380   40404040   40404040   40404040   40404040   40404040   40404040   40060705   03010009
           03A0   00340000   00000000   0000E9E8   D4D6C7C5   D5404040   40404040   40404040   40404040
           03C0   40404040   40404040   40404040   40404040   40404040   40404040   40404040   06070503
```

Fig. 3 Character and Hexadecimal Dump Descriptor Index Entries

```
DUMP FROM 2321, UNIT X'193' ON 12/31/66
VOLUME LABEL FOLLOWS

  VOL(C)    0000    VOL1LMSC-60    D

  VOL(H)    0000    E5D6D3F1  D3D4E2C3  60F6F000  00000004  40404040  40404040  40404040  40404040
            0020    40404040  40404040  40404040  40404040  40404040  40404040  40404040  40404040
            0040    40404040  40404040  40404040  40404040

  HA-R0     0000    00070502  11070502  11000000  08070502  110007D0  00

CELL = 006, SBCELL = 007, STRIP = 005, CYL = 002, TRACK = 017, REC = 001, KL = 000, DL = 2000

  COUNT     0000    07050211  010007D0

  DATA(H)   0000    E9D6C4C9  C1C3C1D3  40D3C9C7  C8E34040  40404040  40404040  40404040  40404040
            0020    40404040  40404040  40404040  40404040  40400000  3115283F  3116975F  3119655F
            0040    3122406F  3123835F  3124484F  3514901F  4110494F  4111792F  4113276F  4116583F
            0060    4117553F  4117767F  4117768F  4125081F  4128547F  4128576F  4521130F  4524776F
            0080    4524951F  4527274F  4527397F  4528766F  4529199F  4533214F  4533224F  4710670F
            00A0    4710675F  4713579F  5112716F  5116482F  5120628F  5120631F  5120654F  5122064F
            00C0    5123746F  5128741F  5130239F  5134197F  5136825F  5511446F  5523399F  5524314F
            00E0    5529493F  5710643F  5713053F  5737065F  6110275F  6112850F  6112852F  6114881F
            0100    6120891F  6121125F  6122430F  612351OF  6510322F  6511605F  6515052F  6516175F
            0120    6516601F  6518007F  6736273F  E9D6D5C1  D340C8C1  D9D4D6D5  C9C3E240  40404040
            0140    40404040  40404040  40404040  40404040  40404040  40404040  40404040  40400000
            0160    3113832F  3124708F  3124733F  3124734F  3124736F  3124740F  3125382F  3514920F
            0180    3519554F  3519566F  3713095F  4112167F  4112661F  4114692F  4118561F  4118931F
            01A0    4119738F  4121201F  4128488F  4128489F  4128491F  4128499F  4512236F  4517612F
            01C0    4518239F  4526248F  4526887F  4527255F  4530947F  4533624F  5123197F  5124047F
            01E0    5124048F  5127896F  513381 6F  5134643F  5135231F  5519856F  5522840F  5717502F
            0200    5719970F  6111106F  6120883F  6123493F  651 1253F  6517259F  6517319F  6519541F
            0220    6710679F  E9D6D5C4  4040C940  E2D7C1C3  C540D7D9  D6C2C540  40404040  40404040
            0240    40404040  40404040  40404040  40404040  40404040  40400000  6121049F  E9D6D5C4
            0260    4040C9C9  40E2D7C1  C3C540D7  D9D6C2C5  40404040  40404040  40404040  40404040
            0280    40404040  40404040  40404040  40400000  5125405F  6111661F  6122302F  6513485F
            02A0    6521566F  6521579F  E9D6D5C4  4040C9C9  C940E2D7  C1C3C540  D7D9D6C2  C5404040
            02C0    40404040  40404040  40404040  40404040  40404040  40404040  40400000  6114073F
            02E0    6119904F  6121755F  6522432F  E9D6D5C5  40404040  40404040  40404040  40404040
            0300    40404040  40404040  40404040  40404040  40404040  40404040  40404040  40400000
            0320    2510046F  2510054F  2510112F  2510183F  2510288F  2510334F  2510378F  2510603F
            0340    2510612F  2510627F  2510657F  2510690F  2510736F  2510749F  2510785F  2510828F
            0360    2510831F  2510877F  2510931F  2510932F  2510937F  2510950F  2510983F  2510987F
            0380    2510988F  2511070F  2511088F  2511177F  2511225F  2511350F  2511392F  2511396F
            03A0    2511430F  2511483F  2511608F  2511628F  2511629F  2511630F  2511631F  2511632F
            03C0    2511633F  2511639F  2511642F  2511653F  2511691F  2511738F  2511739F  2511800F
            03E0    2511815F  2512039F  2512042F  2512069F  2512075F  2512140F  2512156F  2512236F
            0400    2512381F  2512420F  2512426F  2512471F  2512593F  2512598F  2512603F  2512612F
            0420    2512664F  2512682F  2512770F  2512798F  2512802F  2512808F  2512809F  2512863F
            0440    2512880F  2512901F  2512902F  2512925F  2512954F  2513036F  2513093F  2513110F
            0460    2513118F  2513149F  2513156F  2513206F  2513246F  2513248F  2513253F  2513254F
            0480    2513255F  2513314F  2513553F  2513560F  2513677F  2513688F  2513860F  2513864F
            04A0    2513924F  2513927F  2513960F  2514008F  2514016F  2514062F  2514162F  2514176F
            04C0    2514178F  2514209F  2514211F  2514213F  2514256F  2514257F  2514341F  2514553F
            04E0    2514613F  2514616F  2514632F  2514996F  2515031F  2515210F  2515233F  2515336F
            0500    2515339F  2515554F  2515575F  2515693F  2515738F  2515742F  2515822F  2516108F
            0520    2516165F  2516181F  2516187F  2516278F  2516326F  2516332F  2516365F  2516396F
            054Q    2516419F  2516443F  2516493F  2516776F  2516789F  2516870F  2517224F  2517243F
            0560    2517250F  2517342F  2517422F  2517447F  2517562F  2570986F  2571060F  2571646F
            0580    2711069F  2711073F  2711269F  2711454F  2711457F  2711461F  2711462F  3110120F
```

Fig. 4 Hexadecimal Dump of Inverted File Entries

```
DUMP FROM 2321, UNIT X'193' ON 12/31/66

VOLUME LABEL FOLLOWS

  VOL(C)    0000    VOL1LMSC-60     D

  VOL(H)    0000    E5D6D3F1  D3D4E2C3  60F6F000  00000004  40404040  40404040  40404040  40404040
            0020    40404040  40404040  40404040  40404040  40404040  40404040  40404040  40404040
            0040    40404040  40404040  40404040  40404040

  HA-R0     0000    00120403  0C120403  0C000000  08120403  0C0007D0  00

CELL = 006, SBCELL = 018, STRIP = 004, CYL = 003, TRACK = 012, REC = 001, KL = 000, DL = 0711

  COUNT     0000    1204030C  010002C7

  DATA(C)   0000     AZIRCONATE                                          STRONTIUM ZIRCONATE
            0100       EZONE                                             AURORAL ZONE
            0200         BRILLOUIN ZONE                                  NULL ZONE
            0300         RECOVERY ZONE                                   ROYAL ZONE
            0400          AZOOLOGY                                         ICHTHYOLOGY
            0500            AZPR                                             ZERO POWER REACTOR /ZPR/
            0600              A99999999999999999999999999999999999999999999999999999999999999999999999999999999999999999999999999999999
            0700    99999999993

  DATA(H)   0000    0001E9C9  D9C3D6D5  C1E3C540  40404040  40404040  40404040  40404040  40404040
            0020    40404040  40404040  40404040  40404040  40404040  E2E3D9D6  D5E3C9E4  D440E9C9
            0040    D9C3D6D5  C1E3C540  40404040  40404040  40404040  40404040  40404040  40404040
            0060    40404040  40400005  E9D6D5C5  40404040  40404040  40404040  40404040  40404040
            0080    40404040  40404040  40404040  40404040  40404040  40404040  4040C1E4  D9D6D9C1
            00A0    D340E9D6  D5C54040  40404040  40404040  40404040  40404040  40404040  40404040
            00C0    40404040  40404040  40404C40  C2D9C9D3  D3D6E4C9  D540E9D6  D5C54040  40404040
            00E0    40404040  40404040  40404040  40404040  40404040  40404040  40404040  4040D5E4
            0100    D3D34DE9  D6D5C540  40404040  40404040  40404040  40404040  40404040  40404040
            0120    40404040  40404040  40404040  40404040  D9C5C3D6  E5C5D9E8  40E9D6D5  C5404040
            0140    40404040  40404040  40404040  40404040  40404040  40404040  40404040  40404040
            0160    4040D9D6  E8C1D340  E9D6D5C5  40404040  40404040  40404040  40404040  40404040
            0180    40404040  40404040  40404040  40404040  40404040  0001E9D6  D6D3D6C7  E8404040
            01A0    40404040  40404040  40404040  40404040  40404040  40404040  40404040  40404040
            01C0    40404040  40404040  C9C3C8E3  C8E8D6D3  D6C7E840  40404040  40404040  40404040
            01E0    40404040  40404040  40404040  40404040  40404040  40404040  40400001  E9D7D940
            0200    40404040  40404040  40404040  40404040  40404040  40404040  40404040  40404040
            0220    40404040  40404040  40404040  4040E9C5  D9D640D7  D6E6C5D9  40D9C5C1  C3E3D6D9
            0240    4061E9D7  D9614040  40404040  40404040  40404040  40404040  40404040  40404040
            0260    0001F9F9  F9F9F9F9  F9F9F9F9  F9F9F9F9  F9F9F9F9  F9F9F9F9  F9F9F9F9  F9F9F9F9
            0280    F9F9F9F9  F9F9F9F9  F9F9F9F9  F9F9F9F9  F9F9F9F9  F9F9F9F9  F9F9F9F9  F9F9F9F9
            02A0    F9F9F9F9  F9F9F9F9  F9F9F9F9  F9F9F9F9  F9F9F9F9  F9F9F9F9  F9F9F9F9  F9F9F9F9
            02C0    F9F9F9F9  F9F973
```

EZONE

FIG. 2C

Fig. 5 Character and Hexadecimal Dump of Related Term Entries

## Appendix B
## INDIVIDUAL SEARCH SUMMARIES

Appendix A contains summaries of a representative sample of searches performed by NASA Ames Research Center scientists and engineers. In selecting the sample, an effort was made to include examples of the range of complexity of the Boolean search expressions, user's comments, and times spent performing each search. In several instances the user was able to perform several searches in the same console session; the numbered expressions are used to indicate this. In the Boolean search expression the symbols "+," "*," and "−" are used to indicate the Boolean connectives OR, AND, and NOT, respectively.

SEARCH PERFORMED BY: Carr B. Neel, Gasdynamics

SEARCH TITLE: Scattering of Light by MIE and Rayleigh Processes

BOOLEAN SEARCH EXPRESSION: Particle * (Rayleigh Scattering + MIE Scattering Function + MIE Theory) * (Scattering + Scattering Coefficient + Scattering Cross Section + Scattering Function + Atmospheric Scattering + Backscatter + Diffraction + Electromagnetic Scattering + Forward Scatter + Light Scattering + Extinction + Refraction + Cross Section + Absorption Cross Section + Radiation Absorption + Absorption + Refractivity)

SEARCH TIME: 55.61 min. NUMBER OF RELEVANT CITATIONS PRINTED: 18

COMMENTS: This is my first search, and it has been extremely helpful in locating articles very rapidly which otherwise would have been difficult to find. This is a very useful technique, and I highly recommend its adoption throughout NASA.

SEARCH PERFORMED BY: Duane W. Dugan

SEARCH TITLE: Nuclear Propulsion

BOOLEAN SEARCH EXPRESSION: (Nuclear + Nuclear Power + Nuclear Propulsion + Thermonuclear Propulsion) * (Window + (Interplanetary Flight + Orbital Launch) * (Payload + Performance)).

SEARCH TIME: 54.73 min. NUMBER OF RELEVANT CITATIONS PRINTED: 10

COMMENTS: This system appears to be potentially very useful and could save much time in searching the literature for pertinent references.

SEARCH PERFORMED BY: John C. Arvesen

SEARCH TITLE: Cerenkov Radiation Theory and Use as Standard Light Source

BOOLEAN SEARCH EXPRESSION: ((Cerenkov Effect + Cerenkov Radiation) – (Cosmic + Cosmic Radiation + Cosmic Ray Shower + Shower)) * (Spectral Analysis + Spectral + Spectrum)

SEARCH TIME: 64.18 min. NUMBER OF RELEVANT CITATIONS PRINTED: 23

COMMENTS: Very worthwhile took a half-hour to locate more sources than I found in a week. However, the references should go back further than 1962.

SEARCH PERFORMED BY: R. C. Whitten

SEARCH TITLE: Lunar Luminescence

BOOLEAN SEARCH EXPRESSION: (Lunar * Luminescence) + Lunar Luminescence

SEARCH TIME: 51.32 min. NUMBER OF RELEVANT CITATIONS PRINTED: 49

COMMENTS: Very helpful on obtaining information not available in usual places for search.

SEARCH PERFORMED BY: W. L. Crawford

SEARCH TITLE: Time Series

BOOLEAN SEARCH EXPRESSION: ((Time + Series + Time Series) * (Frequency Analysis + Data Analysis + Fourier Analysis + Harmonic Analysis + Numerical Analysis + Spectral Analysis + Statistical Analysis + Spectral + Spectral Analysis))– (Emission + Luminescence + Sun + Light)

SEARCH TIME: 40.37 min. NUMBER OF RELEVANT CITATIONS PRINTED: 79

COMMENTS: Very useful. We should obtain one for permanent use.

SEARCH PERFORMED BY: Philip Wilcox

SEARCH TITLE: (1) Rocket Noises (2) Time Series Programs

BOOLEAN SEARCH EXPRESSION: (1) (Noise + Acoustics) * (Launch + Launching) * (Rocket + Space Vehicle + (Vehicle * Space) + Saturn Launch Vehicle) (2) Time Series * (Computer Program + Computer Programming + Data Processing)

SEARCH TIME: 64.82 min. NUMBER OF RELEVANT CITATIONS PRINTED: (1) 43 (2) 13

COMMENTS: Very valuable for quick search. Easy to use after initial familiarization.

SEARCH PERFORMED BY: J. M. Coogan

SEARCH TITLE: Radar Imaging Interpretation

BOOLEAN SEARCH EXPRESSION: (Interpretation + Analysis + Exploitation) * (Radar Map + Radar Photography + Satellite-Borne Radar + (Radar * (Image + Imagery)))

SEARCH TIME: 51.02 min. NUMBER OF RELEVANT CITATIONS PRINTED: 19

COMMENTS: I made a hand search over this same area, thus far it looks as if the machine approach is at least as thorough and of course much easier. No problems anticipated. Hope the system becomes a permanent tool. Appreciate the patient instruction of the LMSC Rep.

SEARCH PERFORMED BY: Paul Droll

SEARCH TITLE: Magnetic Field Generating Systems

BOOLEAN SEARCH EXPRESSION: Magnetic Field * (Magnetic Field Coil + Coil + Magnetic Coil + Solenoid) * (Square + Uniform + Homogeneity)

SEARCH TIME: 38.40 min. NUMBER OF RELEVANT CITATIONS PRINTED: 9

COMMENTS: Having a good idea of what I was looking for, I have obtained exactly what I wanted. I regret that the printout is not at the location of the searcher. I noticed that only a category of basic references are listed – an expansion of the sources would be a big improvement (naturally).

SEARCH PERFORMED BY: Edgar M. Van Vleck

SEARCH TITLE: Information Transfer Satellites

BOOLEAN SEARCH EXPRESSION: (Information + Data) * (Communication + Communications) * (Human + Computer + Display) * Network

SEARCH TIME: 55.61 min. NUMBER OF RELEVANT CITATIONS PRINTED: 20

COMMENTS: Excellent system, should be used throughout NASA. Nothing equivalent exists Q.E.D.

SEARCH PERFORMED BY: R. J. Debs

SEARCH TITLE: Neutron Activation

BOOLEAN SEARCH EXPRESSION: Calcium * (Bone + Skeleton)

SEARCH TIME: 26.82 min. NUMBER OF RELEVANT CITATIONS PRINTED: 14

COMMENTS: This is my first, short-period attempt to use you, you monster. The results are excellent. One minor criticism some way of eliminating the necessity for shifting to make the commands, might be easier for the new user. This is a great concept, well worked-out.

SEARCH PERFORMED BY: I. G. Poppoff

SEARCH TITLE: Electron Attachment H20

BOOLEAN SEARCH EXPRESSION: ((Electron * Attachment) + (Water + Water Vapor)) * (Drift Rate + Drift)

SEARCH TIME: 32.72 min. NUMBER OF RELEVANT CITATIONS PRINTED: 4

COMMENTS: Found entries that I could not find in a manual search.

SEARCH PERFORMED BY: John Rakich

SEARCH TITLE: Gaussian Quadrature

BOOLEAN SEARCH EXPRESSION: (1) Gauss * (Integration + Quadrature) (2) (Gauss + Gauss Function) * (Integration + Quadrature + Numberical Integration + Quadrature Approximation)

SEARCH TIME: 54.69 min. NUMBER OF RELEVANT CITATIONS PRINTED: (1) 9 (2) 2

COMMENTS: After three tries I feel confident using this program. I think it is a fine tool.

SEARCH PERFORMED BY: William P. Gilbreath

SEARCH TITLE: Surface Properties

BOOLEAN SEARCH EXPRESSION: (1) Adsorption * Surface * Area (2) Nitrogen * ((Heat * Vaporization) + (Enthalpy * Vaporization)) (3) Low Temperature * (Calorimeter + Calorimetry) (4) (Rock + Mineral) * (Fracture + Mechanics + Strength) * (Vacuum + Lunar + Lunar Environment) (5) (Solid * (Adhesion + Cohesion + Metal-Metal Bonding)) – Adhesive

SEARCH TIME: 72.15 min. NUMBER OF RELEVANT CITATIONS PRINTED: (1) 1 (2) 5 (3) 6 (4) 6 (5) 2

COMMENTS: Went OK but would be nice if various authors could be selected or rejected

SEARCH PERFORMED BY: William P. Gilbreath

SEARCH TITLE: Solid Surface Energy

BOOLEAN SEARCH EXPRESSION: (1) (Surface Energy + Surface Tension) * (Solid + Solids) (2) (Surface Energy + Surface Tension) * Metal – (Liquid + Liquid Metal) (3) (Hardness + (Indentation * Creep) * (Environment + Vacuum)

SEARCH TIME: 47.83 min. NUMBER OF RELEVANT CITATIONS PRINTED: (1) 17 (2) 23 (3) 20

COMMENTS: Method OK, very rapid but would nice to have full abstract printed. Search went well probably due to observing a search previously.

SEARCH PERFORMED BY: Lloyd D. Corliss

SEARCH TITLE: Techniques in Model Matching

BOOLEAN SEARCH EXPRESSION: (Model + Aircraft Model + Dynamic Model + Mathematical Model) * (Matching + Following + Adaptive) + (Automatic Control + Optimal Control + Systems Analysis + Adaptive Control System + Closed Loop System + Reference System + Reference)

SEARCH TIME: 56.27 min. NUMBER OF RELEVANT CITATIONS PRINTED: 33

COMMENTS: Most comprehensive search in this field. Would like to see the ENTER command incorporated in with some of the other commands. Thank you.

SEARCH PERFORMED BY: Val Watson

SEARCH TITLE: Numerical Solution of Poisson's Equation

BOOLEAN SEARCH EXPRESSION: (Poisson Equation + Dirichlet Problem) * (Numerical Analysis + Numerical + Iterative Solution + Finite Difference Method + Computer Simulation)

SEARCH TIME: 56.60 min. NUMBER OF RELEVANT CITATIONS PRINTED: 27

COMMENTS: This computer technique allowed me to find relevant articles that were filed under headings that would not be obvious to the engineer without such a computer. For example, finite difference methods is a numerical method that is not also listed under numerical methods. Recommend that author file be added, that continuous backspace be added, and that multilevel arithmetic capability be added to the combine command.

SEARCH PERFORMED BY: J. M. Coogan

SEARCH TITLE: Infrared Imagers

BOOLEAN SEARCH EXPRESSION: Image * (Scanner + Scanning Device + Sensor + Detector) * Ultraviolet

SEARCH TIME: 9.56 min. NUMBER OF RELEVANT CITATIONS PRINTED: 28

COMMENTS: Again I am unable to stress sufficiently the tremendous help this system is both in terms of labor and time. Thank you.

SEARCH PERFORMED BY: Gary J. Griffith

SEARCH TITLE: Spatial Filtering

BOOLEAN SEARCH EXPRESSION: (Optical + Light) * (Spatial Filtering + (Spatial + Filtration + Processing))) – Interferometer

SEARCH TIME: min. NUMBER OF RELEVANT CITATIONS PRINTED: 31

COMMENTS: Techniques easy to learn. Manual search has been conducted for last several months DIALOG appears much faster, but the results from this particular search were impeded by machine trouble. Search is being reproduced to obtain results.

SEARCH PERFORMED BY: P. R. Wilcox

SEARCH TITLE: Rocket Noises

BOOLEAN SEARCH EXPRESSION: (1) Spectral * (Launch + Launching) (2) ((Noise + Acoustics + Spectral + Spectral Analysis + Spectral Noise) * (Takeoff + Launch + Launching) * (Apollo Spacecraft + Rocket + Space Vehicle + Saturn Launch Vehicle + Apollo Project + (Vehicle * Space))) − (Animal + Performance + Man + Human + Personnel + Hazard + Geophysics)

SEARCH TIME: 48.24 min. NUMBER OF RELEVANT CITATIONS PRINTED: (1) 15 (2) 40

COMMENTS: Our division, a headquarters component, often has to respond in a short time to requests for information from our management. Thus far this system for quick access to information is the only one I have seen that can meet our needs.

SEARCH PERFORMED BY: F. G. Casal

SEARCH TITLE: Solar Probes or Heliocentric Missions

BOOLEAN SEARCH EXPRESSION: ((Energy + Energy Requirement) * (Solar Probe + ((Solar + Sun + Heliocentric Orbit) * (Probe + Space Probe + Mission)))) + (Solar Probe * Propulsion)

SEARCH TIME: 55.06 min. NUMBER OF RELEVANT CITATIONS PRINTED: 76

COMMENTS: System very useful in conducting extensive searches.

SEARCH PERFORMED BY: Hornby

SEARCH TITLE: Oceanography-Space Commonality

BOOLEAN SEARCH EXPRESSION: Oceanography * Exploration

SEARCH TIME: 19.09 min. NUMBER OF RELEVANT CITATIONS PRINTED: 7

COMMENTS: OK. Very efficient reduction of data to yield desired information.



NASA-Langley, 1969 —— 25 CR-1318