CONSTRUCTION OF THE NASA THESAURUS:

COMPUTER PROCESSING SUPPORT

by

[1] William Hammond

5 January 1968

FINAL REPORT ON

CONTRACT NAS-1629

[1]NOW WITH:
INFORMATION INDUSTRIES, INC.
6825 REDMOND DRIVE
MCLEAN, VIRGINIA 22101
[703] 356-7372

ARIES Corporation
7678 Old Springhouse Road
Westgate Research Park
McLean, Virginia 22101
(703) - 893-4450

CONTENT:

CONSTRUCTION OF THE NASA THESAURUS:
COMPUTER PROCESSING SUPPORT

INTRODUCTION

National Aeronautics and Space Administration's Contract #NAS-1629, dated 17 May 1967, called for ARIES Corporation to provide the necessary computer processing services to produce a NASA Thesaurus. These services included (1) keypunching the terminology to ARIES specifications from approximately 19,000 Term Review Forms furnished by NASA; (2) modifying a set of ARIES programs to satisfy NASA specifications, principally to accommodate 42-character terms (versus 36 for the Department of Defense Thesaurus of Engineering and Scientific Terms); (3) inclusion of notes to accompany the headings for the subject category display; (4) expanding the indentations between successive generic levels in the hierarchical display; and (5) formating computer generated reproduction copy in the two-column pagination.

During the course of contract, at its own expense, ARIES rewrote its extensive set of thesaurus-related programs to produce an integrated thesaurus software package for the IBM 360 computer models 30 and 40 with OS, DOS and TOS operating systems. Since this software package was introduced about midway of the NASA Thesaurus project, the description of the processing that follows is in terms of the procedures that are compatible with the new set of programs.

The body of this report that deals with the computer application is based on information contained in a preliminary draft of a forthcoming book by William Hammond entitled, Creative Application of Computers in Library Systems.

In order not to interrupt the intellectual processing of the terminology, the main body of the term review forms were released to ARIES in four major increments. Each increment was processed through an identical cycle: The term review forms were received by ARIES on a Friday at the end of the workday and were keypunched at the ARIES Data Center over the weekend; the forms were returned to NASA the following Monday morning; the punched cards were processed by ARIES on a "crash" basis until an edit print was produced.

CREATING THE THESAURUS CORPUS

The punched card format for the initial input and the flow chart for the file creation processing are shown in Appendix A, pages 1-3. In this processing, scope notes, subject category codes and cross references (also referred to as subterms) were keypunched; the computer processing generated the main term record and records for the reciprocal of the cross references. All records generated by the computer were tagged at the time of generation. Where a computer-generated cross-reference record was also keypunched from another term review form, the generated duplication was dropped in a subsequent computer pass. Thus, the edit print for the first increment contained a main term record for each term review form furnished as well as a generated main term record for any cross-referenced term

for which a term review card had not been furnished. The latter, however, carried the generated "QT" identifying tag. The same was true of cross references; where the reciprocal of a cross reference to a term for which a term review form had not been furnished was automatically generated, the identifying "QT" tag was added. Scope notes and subject category codes were present in the edit print only for those terms for which the term review forms were furnished. An edit of the first increment was completed by NASA and processed by ARIES before the second increment was added to the thesaurus corpus.

Succeeding increments were processed by the computer in a like manner but were merged with the preceding increments before the new edit prints were produced. The generated QT tag was eliminated from a given record whenever a duplicate record keypunched from a term review form was encountered in the merge. Examples of the edit prints are contained in Appendix A, page 12.

Some difficulty was encountered in maintaining the generated term tag throughout successive processing stages, utilizing programs originally designed for "one-step" processing, with intervention for intellectual processing between successive steps. This prompted ARIES to expedite its production of a "next generation" set of thesaurus programs that had been in work for some time on a low priority basis. These programs permitted computer controlled thesaurus construction--or maintenance-- from initial input to reproduction copy. Intervention at any point for intellectual review and/or processing was retained as an option.

Throughout the first four increments that collectively constituted the gross thesaurus corpus, corrections were principally in word form and spelling. It was not until the fifth "intermediate thesaurus" print that the full repertoire of maintenance operations were permitted. (See functions of Modifile, Appendix A, page 8) to provide a suitable corpus from which to produce the "Review Copy" of the NASA Thesaurus in two-column format ready for publication.

Subsequent to the publication of the Review Copy of the Thesaurus, ARIES' contract was amended to include the responsibility for providing an editorial "overview" of the Thesaurus and provide recommendations for improvements in content and internal consistency. A copy of the separate report on this task provided by the subcontractor is attached as Appendix C. Corrections, deletions and additions resulting from this and other NASA editorial effort affected about four percent of the thesaurus records. This figure included added records required to expand the generic (BT-NT) structure and to accommodate terms up to 42 characters in length (limited to a maximum of 36 characters in the intermediate editions).

The thesaurus specifications (see Appendix B) call for explicit display of the entire generic set (BT-NT set) along with each main term record in the thesaurus. Also, all cross references (subterm records) are reciprocal; i.e., if term B has a broader term A, then A must display a narrower term B; if term B has a narrower term C, then term C must display a broader term A as well as a broader term B. (See Figure 1 on the opposite page for successive stages of the BT-NT

INITIAL INPUT	EXPAND I	EXPAND II
TERM A	TERM A	TERM A
(R) NT TERM B	(R) NT TERM B	(R) NT TERM B
	(E1) NT TERM C	(E1) NT TERM C
		(E2) NT TERM D
		(E2) NT TERM E
TERM B	TERM B	TERM B
BT TERM A	BT TERM A	BT TERM A
(R) NT TERM C	(R) NT TERM C	(R) NT TERM C
	(E1) NT TERM D	(E1) NT TERM D
		(E2) NT TERM E
TERM C	TERM C	TERM C
BT TERM B	(E1) BT TERM A	(E1) BT TERM A
(R) NT TERM D	BT TERM B	BT TERM B
	(R) NT TERM D	(R) NT TERM D
	(E1) NT TERM E	(E1) NT TERM E
TERM D	TERM D	TERM D
BT TERM C	(E1) BT TERM B	(E2) BT TERM A
(R) NT TERM E	BT TERM C	(E1) BT TERM B
	(R) NT TERM E	BT TERM C
		(R) NT TERM D
TERM E	TERM E	TERM E
BT TERM D	(E1) BT TERM C	(E2) BT TERM A
	BT TERM D	(E2) BT TERM B
		(E1) BT TERM C
		BT TERM D

NOTES:

(R) Reciprocal records generated by computer in converting from punched cards to tape.

(E1) Records generated by computer in First Expand Pass.

(E2) Records generated by computer in Second Expand Pass.

Figure 1

expansion from the minimum input--next broader term only).

From an examination of Figure 1, it can be seen that when processing a linear file, any gaps in the generic hierarchy exceeding three levels cannot be detected in a single pass. In Figure 1, five levels are implicit. In converting the punched card input referred to earlier, the reciprocal records that were generated are denoted by (R). The first expand pass generated the implicit BT-NT cross references, denoted by (E1), over a gap of three generic levels. Thus on the first pass, the complete generic set would be generated only for term C. A second pass would supply the remaining implicits, denoted by (E2). From an examination of the processing logic, it can be seen that if the complete BT-NT set for term C had been included in the original input, the first expand pass would have generated the full BT-NT set for all five terms.

The second expand pass needs only to concern itself with those main term record sets where a generated BT or NT record added on the previous pass will result in the generation of a different set of expand records. This will occur when the main term record set prior to the previous expand pass contains one or more NT records and a BT record was added, or when an NT record was added to the existing set that contained one or more BT records. For the purposes of our thesaurus software discussion, a main-term record set consists of the main term record together with all cross-reference records listed under the main term record. As can be seen from Figure 1, two passes will span a generic gap of five levels, and three passes will span a gap of nine

levels. Each subsequent expand pass will increment the generic span by four levels and on each pass only those main-term record sets that will generate a different set of expand records need to be processed. If the implicits generated on the second or successive passes exceed those of the previous pass, the hierarchical structuring should be validated by intellectual review. For example, in the early processing stage of the NASA Thesaurus, the term MONOPLANE was listed as a narrower term (NT) to the term AERODYNAMIC CONFIGURATIONS, which in turn was listed as an NT to AERODYNAMICS. On the second expand pass, a long list of aircraft designations (since most of the current models are monoplanes) appeared as NT's under AERODYNAMICS and AERODYNAMIC CONFIGURATIONS, and also under any other term carried as a BT to AERODYNAMIC CONFIGURATIONS. This was not the relationship that NASA had in mind and the MONOPLANE-AERODYNAMIC CONFIGURATIONS relationship was changed to an RT, thus eliminating these unwanted accumulations.

The entire thesaurus construction--or maintenance--process must be monitored by someone familiar with the terminology involved and also knowledgeable of the "thesaurus model" and the logic of the computer programs. This becomes obvious when one traces the addition or deletion of a cross-reference.

If an NT cross-reference is deleted, then it is necessary to delete not only its reciprocal BT record (to prevent the deleted record from being regenerated on a subsequent pass) but also other NT records that may be present at different positions in the file on the sole basis that they are also listed as NT's to the NT being deleted or

being changed to an RT. Also where NT's or BT's are added, their reciprocals and their implied generics must be added. Where these additions introduce conflicts with existing RT's, then the conflicting RT and its reciprocal must be deleted.

In the construction of a new thesaurus, or for a significant modification of an existing thesaurus, an intellectual review of at least the BT set for each main term record is essential. The BT set is suggested because it is more conveniently distributed throughout the thesaurus. (Few main term record sets contain more than three BT's in either the DOD or NASA Thesaurus; in some instances the NT sets will exceed a hundred.) For each BT record in a given main term record set, the reviewer should ask himself the question, "Is this main term one of these BT's?" Also, "Do I want to cumulate and display the associated NT sets in the thesaurus whenever this main term is displayed elsewhere in the thesaurus as an NT?" (See earlier reference to AERODYNAMIC CONFIGURATIONS being shown as a BT to MONOPLANE.)

Experience in the construction of several thesauri has demonstrated that the project will be less costly--and less time consuming--if "in processing" intellectual reviews of the generic structure are made after the reciprocals are validated by the computer and again after the first expand pass. Corrections and resolution of discrepancies must be accomplished at the end of each such review. A final review of the thesaurus corpus will be most profitable if a permuted display, a subject category display, and hierarchical displays of

the vocabulary can be provided as reviewing aids.

Where major changes are being made in the constitution of the thesaurus corpus, two points of interruption in MODIFILE processing are available for intellectual review. (Refer to flow chart in Appendix A, page 7.) The run that converts the change data on punched cards to magnetic tape also edits the data fields in the cards. In the IBM 360 TOS and DOS operating systems, the cards containing errors are left off the tape and stacker selected. For the OS operating system, these errors are left off the tape and are printed out for correction. To identify other forms of erroneous change data--or to add or delete or modify the initial change input--a listing of the sorted change file can be printed out for intellectual review.* A subroutine is available to alter and/or correct the change data tape prior to processing it against the thesaurus corpus.

The second interruption can be taken at the point in the processing just preceding the final MODIFILE data merge with the thesaurus corpus. This final MODIFILE data tape can be printed out for review; however, to introduce corrections at this stage in the processing is not practical. The changes of any significance will require going back to the first interrupt point.

It is essential that the reviewer be fully conversant with the functions of MODIFILE that are shown in Appendix A, page 8. In the

---

* Might also be used as input to affect automatic changes in past indexing.

interest of computer economy, each new entry and each main term record for which a change is processed in the first MODIFILE computer pass, are output on a separate tape. Deletions are also output on a separate tape since their reciprocals, which are identified in the first MODIFILE pass and which must also be deleted, appear at other positions in the file.

Duplicate entries generated or introduced in the MODIFILE processing will automatically be deleted. If conflicting functions are introduced and one is a delete, the delete will take precedence. If conflicts in term relationships are introduced, they will be automatically eliminated (if no manual intervention is specified) in subsequent processing with the following retention precedence:

MT (occasionally cross-referenced to itself)
USE
UF
BT
NT
RT

In summary, the Modifile logic is designed to maintain the integrity of the intellectual decisions and reflect them throughout the thesaurus corpus. When a term is to be deleted from the magnetic tape which stores the thesaurus corpus, it is essential that all traces of the term be purged from the file. This is of necessity a multi-pass operation when the file is maintained in linear format on magentic tape since the residual entries--and their influences--are scattered throughout the file. Subtleties are not easily traced, particularly where a term being added or deleted affects the hierarchical structuring of other terms in the file. Also, as in the case

of several thesauri, scope notes may make reference to terms appearing elsewhere in the file.

Where spelling of a term is being altered, processing is also complex; the old spelling must be replaced by the new spelling throughout the thesaurus. Additions and deletions of cross-references also have a residual effect on the remainder of the thesaurus corpus. The Modifile program set is designed to reconcile the residual effect of all alterations to the thesaurus as well as the "bookkeeping" to insure that new entries are not cross-referenced to terms since deleted or modified in spelling -- cumulative actions that are often being executed simultaneously in the same maintenance processing.

Samples of the two-column format produced for the publication of the NASA Thesaurus Working Copy are contained in Appendix A, pages 13 through 19.

HIERARCHICAL DISPLAY GENERATOR

The hierarchical generator utilizes the magnetic tape file of the final thesaurus corpus as input. Displays are automatically generated only for those terms that have no broader terms in the thesaurus and have at least three generic levels listed. However, as an optional feature, the terms to be displayed can be designated by the thesaurus editor. It was suggested to both NASA and DOD that they employ the hierarchical display format throughout the main body of their published thesauri. For the DOD thesaurus this would have meant reducing the four-column format to three columns and adding about a

hundred pages to the printed book. NASA felt that its preliminary edition, or "working copy", should receive operational usage under their current approach. In the NASA Thesaurus the heads of hierarchies for this display carry an identifying tag (#) wherever they appear in the thesaurus main body. In the DOD Thesaurus the intermediate level terms (those terms with narrower term cross-references) carry an identifying tag (-) wherever they appear in the thesaurus main body.

This operation requires the computer to select the hierarchy sets to be displayed, compile the broader term set for each term in the hierarchical display, and then determine the immediate broader term (or terms if more than one tree is involved) to each term in the set. Numeric codes were substituted for redundant occurrences of the terms to form the array shown in Figure 2. Coding was primarily to conserve core storage since several of the hierarchical sets contained so many terms that the program logic would have to be modified or a much larger computer would have been required.

Although the hierarchical display example shown in Figure 3 is one of less than average complexity, a description of its development will serve to illustrate the computer logic in the generation of the hierarchical display. A much more detailed description of the computer programming techniques is contained in a paper by Johnie Sullivan, who wrote the IBM 360-40 programs that were used to produce the hierarchical displays for both the NASA and DOD thesauri.

Term Code			Codes of Broader Terms			Code of Immediate Broader Term
32	Microwave equipment 0903					
	UF	Microwave devices				
21	NT	Backward wave tubes	32	37	41	41
22		Carcinotrons.	32	37	41	41
23		Ferrimagnetic amplifiers. . .	32	30		30
24		Horn antennas	32	31		31
25		.Klystrons	32	37		37
26		.Lens antennas	32	31		31
27		Luneberg lenses	32	26	31	26
28		Magnetrons.	32	37		37
29		Maser amplifiers.	32	30		30
30		.Microwave amplifiers.	32			32
31		.Microwave antennas.	32			32
33		Microwave filters	32			32
34		Microwave oscillators	32			32
35		Microwave reflectometers. . .	32			32
36		Microwave spectrometers . . .	32			32
37		.Microwave tubes	32			32
38		Platinotrons.	32	37		37
39		Reflex klystrons.	32	25	37	25
40		Slot antennas	32	31		31
41		.Traveling waves tubes	32	37		37

Figure 2, Generic Array

Microwave equipment
.Microwave amplifiers
..Ferrimagnetic amplifiers
..Maser amplifiers
.Microwave antennas
..Horn antennas
..Lens antennas
...Luneberg lenses
..Slot antennas
.Microwave filters
.Microwave oscillators
.Microwave reflectometers
.Microwave spectrometers
.Microwave tubes
..Klystrons
...Reflex klystrons
..Magnetrons
..Platinotrons
..Traveling wave tubes
...Backward wave tubes
...Carcinotrons

Figure 3, Hierarchical Display

The main term record set for the term Microwave equipment, shown in Figure 2, is in the same sequence that each appears in the main body of the thesaurus. In the interest of simplicity, the numeric codes for the terms retain this same sequence. In the first processing phase, the computer selects and writes out on the same work tape the hierarchy sets to be displayed together with the entire set of BT records in the thesaurus corpus. The records for display in the hierarchy sets are output just as they appear in the thesaurus file. The BT records are output in reciprocal form so that they can be matched with their corresponding terms in the display set to transfer the accumulation of codes of the broader terms and to form the array shown in Figure 2 in the columns headed by Codes of Broader Terms. The array at this stage in the processing tells us which terms in the set are broader than others in the same set. For instance, it tells us that the term Backward wave tubes has three broader terms in the set; Term 32, Microwave equipment; Term 37, Microwave tubes; and Term 41, Traveling wave tubes. It does not tell us the relative generic order of the three broader terms.

In the next phase of processing, our objective is to determine the immediate broader term for each term in the display set. Starting at the top of the list with Term 21, Backward wave tubes, we must determine whether Term 37 or Term 41 is the next broadest to Term 21. To do this, we must determine which of the two terms--37 or 41--is the narrower--a sort of "lose or win" tournament. We scan down the list of terms to Term 37, Microwave tubes; there we see that Term 37

does not have a broader term in the list other than Term 32 which is the head of the hierarchy. We, therefore, scan down the list to Term 41. There we see that Term 41 has a broader Term 37; therefore, we conclude that Term 41, Traveling wave tubes, is the next broader term to Term 21, and so we place "41" at the head of the last column in the array as the numeric code for the next broader term to Term 21. Processing successive term records in the set in a similar manner will complete the determination of the next broader term to complete the last column in the array shown in Figure 2. In the actual processing, duplicate term records were added to the array whenever one of the terms in a display had more than one broader term in the set.

To formulate the hierarchical display shown in Figure 3 from the fully developed generic array in Figure 2, we print out the display header, Microwave equipment, left justified in the print area. We next scan the column of immediate broader term codes until we come to one whose immediate broader term code matches the display header. In this instance, it is Term 30, Microwave amplifiers; we now print Microwave amplifiers as the second term in the display indented two positions (the dot in the second indented position was used by DOD to assist the user following the structure; NASA used more exaggerated indentation to accomplish the same effect). Since the dot in front of Term 30 indicates that it has narrower terms, we must at this point go back to the head of the column of "Codes of Immediate Broader Terms" and scan it for terms whose immediate broader term is 30. Following this procedure, we would first select Term 23, Ferrimagnetic amplifiers,

print it out, indented one position from Term 30, and since Term 23 has no narrower terms, continue to scan down the column for 30. We next come to Term 29, Maser amplifiers; since it has no narrower terms in the list, it is printed out in the same print position as Term 23. There are no more code 30's in the column; therefore, we have completed the sub-display under Microwave amplifiers and must now resume our scanning of the column "Codes of Immediate Broader Terms" for the next term whose immediate broader term is Microwave equipment; this turns out to be Term 31, Microwave antennas. We then write out Microwave antennas as we did Microwave amplifiers and continue the processing as we have described it to this point, until the full hierarchical display has been printed (or a print tape produced for subsequent print formating).

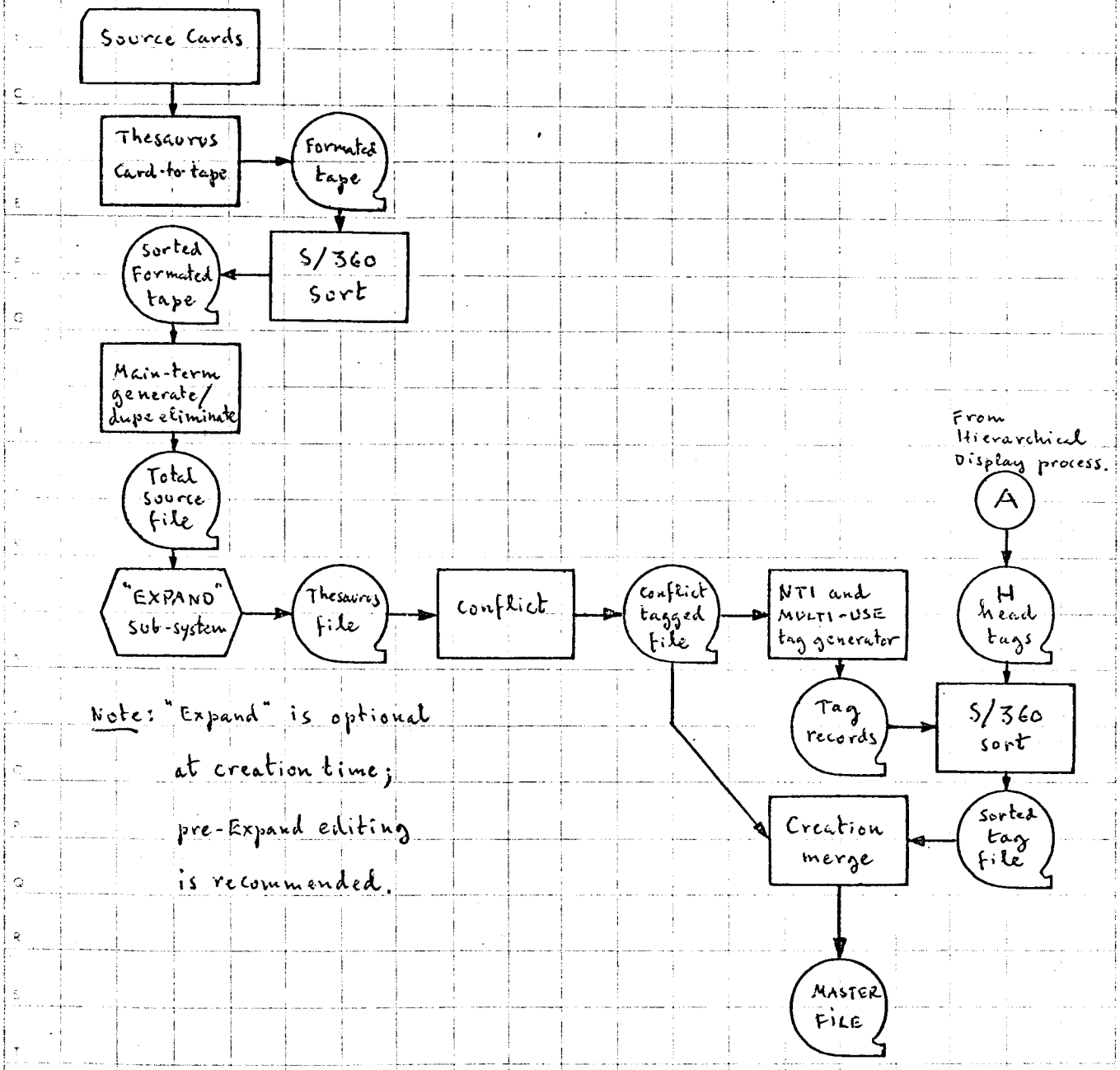
The "bookkeeping" for the processes is much more complex than actually described when large displays contain many more generic levels and when there are several overlapping generic trees; however, the logic remains the same.

APPENDIX A

ARIES CORPORATION DIAGRAMMING AND CHARTING WORKSHEET

Application ARIES Thesaurus Package Date _____ Page 1 of 2

Procedure Creation Sub-System Drawn By Staffan Rosenborg

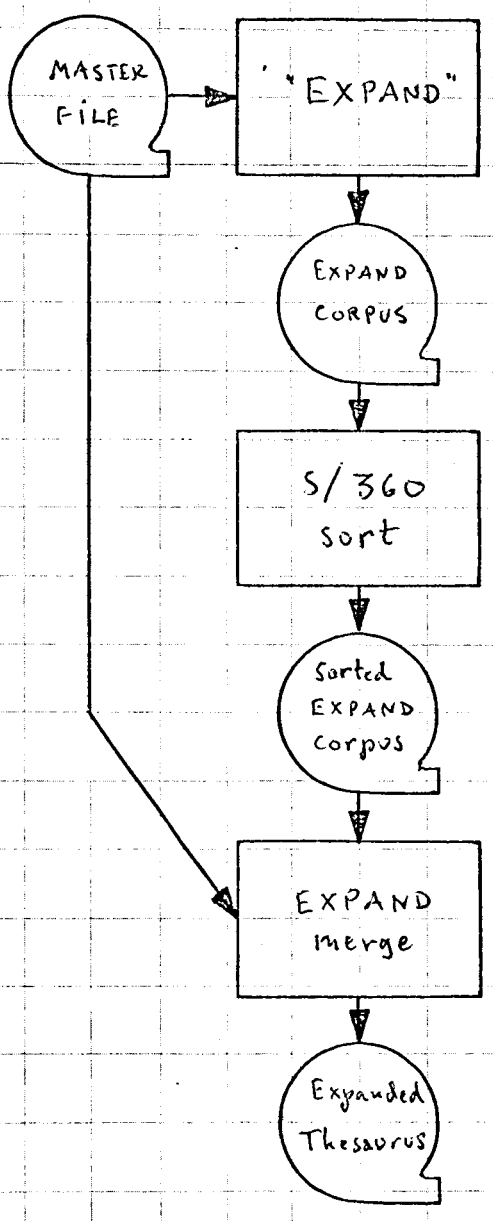


Note: "Expand" is optional at creation time; pre-Expand editing is recommended.

ARIES CORPORATION DIAGRAMMING AND CHARTING WORKSHEET

Application ARIES Thesaurus Package Date _____ Page 2 of 2

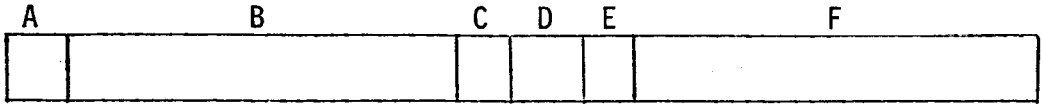
Procedure "EXPAND" Sub-System Drawn By Staffan Rosenborg



Note: The Expand Merge has the capability of successive expansion. When this option is used, the expanded thesaurus is re-introduced to the Expand Merge as the Master File. A second output (New Expand) of the initial merge is sorted and used as the second input.

CARD FORMAT FOR INITIAL INPUT

This card format was adopted by ARIES to facilitate data interchange in the early stages of collaboration between Project LEX and the Engineers Joint Council. It is a modified version of a format already in use by a number of contributors to the original EJC thesaurus updating project.



FIELD	CARD COLUMN(S)	CONTENT
A	1-2	Reserved -- (was originally used to carry a 2-digit subject code for the LEX Data Store of contributed vocabularies).
B	3-38	Main-term field.
C	39-40	Reserved -- (was originally used to carry a 2-character source code for the LEX Data Store of contributed vocabularies).
D	41-43	Relationship or record type designator: BT, NT, RT, USE, UF, or SN.
E	44	Scope-note sequencing code (1-9) or subject-category record designator (C).
F	45-80	Sub-term field.

Main-term card:

Alpha/numeric of Main Term in Field B.

All other fields blank

Note: The punching of Main-term cards is optional -- Main-term records can be automatically generated from sub-term cards.

Sub-term cards:

Alpha/numeric of Main Term in Field B; this Field is then duplicated in the cards for all sub-terms belonging to a given Main-term entry.

Relationship designator: BT, NT, RT, USE, or UF in Field D.

Alpha/numeric of Sub-term in Field F.

Scope-note cards:

Alpha/numeric of Main Term in Field B (same as for sub-term cards).

SN in Field D.

Scope-note line sequence number (1-9) in Field E.

Scope note alpha/numeric in Field F.

Subject-category cards:

Alpha/numeric of Main Term in Field B (same as for sub-term cards).

SN in Field D.

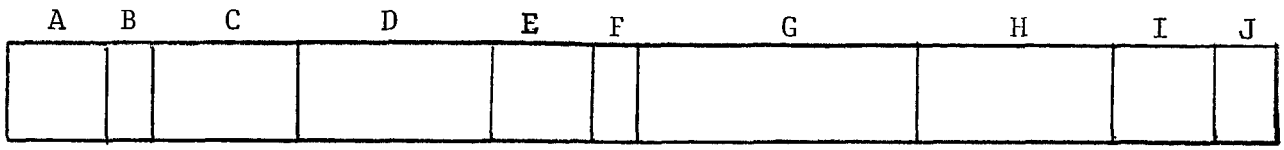
C in Field E.

Category numbers in Field F. These are punched as 4-digit sets, beginning in Col. 45, with a blank column separating the sets. Seven category-number sets can be accommodated on one card. If more are required, Fields A-E are duplicated, and the string of subject-category sets is resumed in Col. 45.

DESCRIPTION

OF

MAGNETIC TAPE LAYOUT FOR NASA THESAURUS



109 CHARACTER RECORD -- 10 FIELDS -- IBM MODE

(Blocking Factor optional depending on capacity of computer; 800 BPI recording density).

FIELD CONTENT

FIELD	POSITION(S)	CONTENT
A	1-7	Reserved for line sequence code employed in maintenance.
B	8	Term relationship code; see code key below.
C	9-14	Reserved for main term entry numeric sequence code employed by system.
D	15	Line sequence code for scope note; type of pseudo scope note; see code key below.
E	16-20	Reserved for subterm entry numeric sequence code employed by system.
F	21-23	Reserved for system use
G	24	Reserved for term tag
H	25-66	42-character term entry, scope note line, or subject category codes; this field is also referred to as the sub-term field; see key below.
I	67-108	42-character term entry; this field also referred to as main-term field.
J	109	Reserved for systems use

"Reserved" fields serve multi-purpose use and their content will vary with the application.

FIGURE 1

RECORD IDENTIFICATION CODE KEY

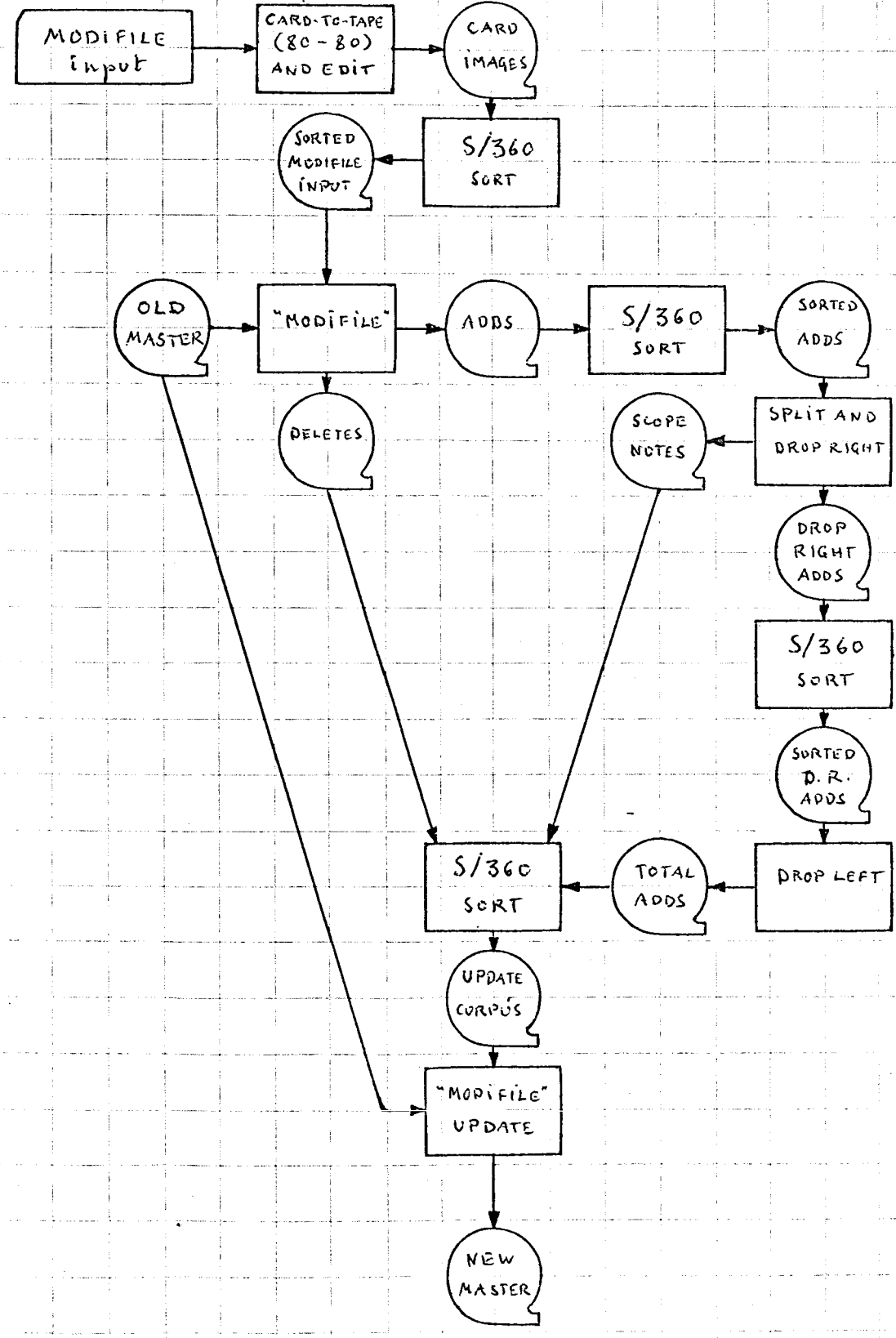
- Main-term record (MT) is identified by a 1 in position 8.
- Scope Note (SN) is identified by a 2 in position 8 and a numeric line sequence code (1 through 9) in position 15.
- Subject Category Code is identified by a 2 in position 8 and a C in position 15. Each subject category is represented by a set of 4 numeric digits, beginning in position 25, with a blank between each 4-digit set. Thus, seven category codes may be carried in a single record. If more than seven codes are required, an additional record is formed.
- A "USE" cross-reference is identified by a 3 in position 8.
- A "Use for" (UF) cross-reference (reciprocal of USE) is identified by a 4 in position 8.
- A "Broader term" (BT) cross-reference is identified by a 5 in position 8.
- A "Narrower term" (NT) cross-reference is identified by a 6 in position 8.
- A "Related term" (RT) cross-reference is identified by a 7 in position 8.

FIGURE 1A

ARIES CORPORATION DIAGRAMMING AND CHARTING WORKSHEET

Application ARIES Thesaurus Package Date _____ Page 1 of 1

Procedure "MODIFILE" Sub-System Drawn By Staffan Rosenborg



FUNCTIONS OF "MODIFILE"

1. To DELETE records:

   D Function: Deletes a single record (line) and its reciprocal, where applicable (e.g., category and scope note records have no reciprocals).

   Format: D in Col. 1

   Sequence number in Cols. 2-8

   All other columns blank

   Note: Do not employ on UF records; rather, employ the T function (described below) on the "main term" of the USE reference.

   R Function: Deletes all records between two sequence numbers, and all their reciprocals.

   Format: R in Col. 1

   Beginning sequence number (inclusive) in Cols. 2-8

   Ending sequence number (inclusive) in Cols. 9-15

   All other columns blank

   T Function Deletes an entire term entry and all reciprocals (i.e., all occurrences of a term throughout the thesaurus). This function is intended for use on Main-term line only (although it can, in fact, be used to delete the entire tail end of an entry).

   Format: T in Col. 1

   Sequence number of Main Term (or beginning of delete) in Cols. 2-8

   All other columns blank

2. To CHANGE records:

   C Function: Changes the relationship between a pair of terms. Except as noted below the change may be initiated at either of the two reciprocal occurrences of the pair, and is automatically carried through to the other.

Format: C in Col. 1

Sequence number of sub-term in Cols. 2-8

New desired relationship beginning in Col. 9 (i.e., BT, NT, RT, or USE).

All other columns blank

Note: When changing a term pair to the USE/UF relationship, it is preferable to initiate the change on the USE side.

E Function: Changes the spelling (or wording) of a term in all its occurrences throughout the thesaurus. The E change is always initiated at the Main-term occurrence.

Format: E in Col. 1

Sequence number of Main Term in Cols. 2-8

New spelling (or wording) in Cols. 9-50

All other columns blank

Note: Aside from normal spelling corrections, this function is also used for re-wording of terms beyond the 36-character limit of the "new input" card format. The maximum term length is 42 character positions.

Z Function: Changes the content of a Scope Note or a Subject Category record (line) to the new content indicated. Also has the option of changing or confirming the 1-digit sequence number of a Scope Note line, or the "C" designator of a Category record (i.e. tape position 15 in Magnetic Tape Record).

Format: Z in Col. 1

Sequence number in Cols. 2-8

New context of record in 9-50

Optional: SN1, SN2, SN3, etc., or SNC in Cols 78-80. (If option is not exercised, whatever is in tape position 15 of that record remains unchanged).

All other columns blank

Note: To change the wording or spelling of a Scope Note, only the line(s) affected need be changed (but the new wording of the entire line must be punched). Subject Category numbers are punched as 4-digit sets, beginning in Col. 9, leaving a blank column between sets.

3. To ADD records:

I Function: Inserts a record; adds its reciprocal, where applicable. The point of insertion is denoted by punching the sequence number of the record immediately preceding the insert. The inserted record automatically picks up the "main-term field" of the preceding record already on the file, unless the inserted record itself is denoted as a Main Term (MT).

Format: I in Col. 1

Sequence number in Cols. 2-8 (number of record immediately preceding the insert). If tne record being inserted is a Main Term (MT in Cols. 78-79), punch A instead of Ø in Col. 2. For subsequent records (scope note, category-number record, BT's, NT's, etc.), restore Ø in Col. 2, and keep repeating the sequence number.

Content of record (spelling or categories) in Cols. 9-50. If the insert is a category line, punch as described under Z Function above.

Type of record in Cols. 78-80. This is mandatory for inserts: the card must specify MT, BT, NT, RT, USE, UF, SNC, SN1, SN2, SN3, or whatever.

Note: The only significance of the insertion point is that it governs the "main-term field" that is picked up by an inserted record, as mentioned above. Hence, the point must be selected somewhere within the entry under the desired Main Term. The specific location within the term entry is irrelevant because all inserts and their reciprocals first get sorted and later merged back into the file, alphabetically and by relationship. New Main Terms are best inserted after the last sub-term of an entry in the file. All sub-terms inserted under that same

sequence number will pick up their "main-term field" from the inserted Main Term. To make sure that the Main-Term insert is processed ahead of its sub-terms, the letter A is substituted for the first leading zero in the sequence number. If several main-term entries are to be added, it is best to select different insertion points for each one. Again, the alphabetic location of these points within the thesaurus is irrelevant.

USE/UF Swap

Complex interactions can occur between MODIFILE functions, which make it necessary to follow a definite procedure. One such is the USE/UF swap, where a complete "about-face" is made concerning two synonyms (for example, "Trinitrotoluene USE TNT" is changed to "TNT USE Trinitrotoluene").

To achieve this, an E-change is performed on the Main Term of the entry containing the UF, and a T-delete is performed on the Main Term of the USE.

0075579 USE GEOMAGNETISM
0075580 MAGNETIC SURVEYS
0075581 MAGNETOVARIOGRAPHS
0075582 USE VARIOMETERS
0075583 MAGNETRONS
0075584 0701 0905
0075585 BT ELECTRON TUBES
0075586 MICROWAVE EQUIPMENT
0075587 MICROWAVE TUBES
0075588 OSCILLATIONS
0075589 *OSCILLATORS
0075590 *PLANOTRONS
0075591 VACUUM TUBE OSCILLATORS
0075592 *VACUUM TUBES
0075593 NT *PLANOTRONS
0075594 *VACUUM TUBES
0075595 RT CAVITY RESONATORS
0075596 CROSSED FIELD AMPLIFIERS
0075597 CROSSED FIELDS
0075598 ELECTROSTATIC GENERATORS
0075599 KLYSTRONS
0075600 MULTIMODE RESONATORS
0075601 *OSCILLATORS
0075602 RESONATORS
0075603 TRAVELING WAVE TUBES
0075604 MAGNETS
0075605 0801 0904 1504 1701 2307
0075606 NT CRYOGENIC MAGNETS
0075607 ELECTROMAGNETS
0075608 FERRIMAGNETS
0075609 HIGH FIELD MAGNETS
0075610 SUPERCONDUCTING MAGNETS
0075611 RT ELECTRETS
0075612 FERROMAGNETIC MATERIALS
0075613 FERROMAGNETISM
0075614 MAGNET COILS
0075615 MAGNETIC CORES
0075616 MAGNETIC FIELDS
0075617 MAGNETIC MATERIALS
0075618 MAGNETIC PROPERTIES
0075619 MAGNETIZATION
0075620 PERMALLOYS (TRADEMARK)
0075621 MAGNIFICATION
0075622 1002 2306 2310
0075623 UF MAGNIFIERS
0075624 RT AMPLIFICATION
0075625 INCREASING
0075626 LENSES
0075627 PROJECTION
0075628 MAGNIFIERS
0075629 USE MAGNIFICATION
0075630 MAGNITUDE
0075631 1902 3408

Specifications for Reproduction

Copy of NASA Thesaurus

1. Two-column format which will fit within a 16-inch vertical space (including headers and page numbers) and 11½-inch horizontal space.

2. Running head "NASA Thesaurus" on each page.

3. Left page - indicate first term on page in header. If a term has been continued from the previous page, pick up that term in header and indicate (continued).

4. Right page - indicate last main term entry on page in header.

5. Pagination - bottom center.

6. Twenty-six new starts - one for each letter of the alphabet. Each new start on right-hand page.

7. Use MYLAR ribbon printer.

8. Additional printouts should have following headings:

   a. Hierarchy - NASA Thesaurus (Appendix A)

   b. Category - NASA Thesaurus (Appendix B)

   c. Permuted Index - NASA Thesaurus (Appendix C)

   d. Postable Terms.

9. New term entry will not start on last line of any column.

MICROWAVE SPECTRA
2309 2402
UF INTERSTELLAR MICROWAVE SPECTRA
BT ELECTROMAGNETIC SPECTRA
RADIATION SPECTRA
RADIO SPECTRA
#SPECTRA
RT ABSORPTION SPECTRA
INFRARED SPECTRA
#MICROWAVE EQUIPMENT
MICROWAVE FILTERS
MICROWAVE FREQUENCIES
MICROWAVE TRANSMISSION
MOLECULAR ROTATION
MOLECULAR SPECTRA
MOLECULAR SPECTROSCOPY

MICROWAVE SWITCHING
0902 0904 2309
BT SWITCHING
RT FERROELECTRICITY
GYRATORS
#MICROWAVE EQUIPMENT
PHASE SHIFT
SWITCHING CIRCUITS
WAVEGUIDES

MICROWAVE TRANSMISSION
0701 2309 2402
BT ELECTROMAGNETIC WAVE TRANSMISSION
RADIO TRANSMISSION
#SIGNAL TRANSMISSION
#TRANSMISSION
WAVE PROPAGATION
NT MANDELSTAM REPRESENTATION
RT MICROWAVE SCATTERING
MICROWAVE SPECTRA
TELETYPEWRITER SYSTEMS

MICROWAVE TUBES
0905 2309 2402
BT #ELECTRON TUBES
#MICROWAVE EQUIPMENT
#OSCILLATORS
VACUUM TUBE OSCILLATORS
VACUUM TUBES
NT BACKWARD WAVE TUBES
CARCINOTRONS
CATHODE RAY TUBES
CELESCOPES
COLD CATHODE TUBES
GAS DISCHARGE TUBES
HELITRONS
IGNITRONS
IMAGE ORTHICONS
IMAGE TUBES
KLYSTRONS
MAGNETRONS
MONOSCOPES
ORTHICONS
PHOTOMULTIPLIER TUBES
PHOTOTUBES
PICTURE TUBES
PLANOTRONS
THERMICONS
THERMIONIC DIODES
THYRATRONS
TRAVELING WAVE TUBES
VIDICONS
RT MICROWAVE CIRCUITS
MICROWAVE RESONANCE
TRIODES
TUBES

MICROWAVES
2309
UF MICROWAVE RADIATION
BT #ELECTROMAGNETIC RADIATION
RADIO WAVES
SHORT WAVE RADIATION
NT MILLIMETER WAVES
RT COSMIC NOISE
ELECTROMAGNETIC NOISE
EXTRATERRESTRIAL RADIO WAVES
INFRARED RADIATION
MICROWAVE FREQUENCIES
SUBMILLIMETER WAVES
WHISTLERS

MICROWEIGHING
USE WEIGHT MEASUREMENT

MICTURITION
USE URINATION

MIDAS SATELLITES
1404 3107 3401
BT ARTIFICIAL SATELLITES
EARTH SATELLITES
#MILITARY SPACECRAFT
RECONNAISSANCE SPACECRAFT
#SATELLITES
NT MIDAS 3 SATELLITE
MIDAS 4 SATELLITE
MIDAS 5 SATELLITE
MIDAS 6 SATELLITE
MIDAS 7 SATELLITE
RT ATLAS AGENA B LAUNCH VEHICLE

MIDAS 2 SATELLITE
1404 3107
BT ARTIFICIAL SATELLITES
EARTH SATELLITES
#MILITARY SPACECRAFT
RECONNAISSANCE SPACECRAFT
#SATELLITES

MIDAS 3 SATELLITE
1404 3107
BT ARTIFICIAL SATELLITES
EARTH SATELLITES
MIDAS SATELLITES
#MILITARY SPACECRAFT
RECONNAISSANCE SPACECRAFT
#SATELLITES

MIDAS 4 SATELLITE
1404 3107
BT ARTIFICIAL SATELLITES
EARTH SATELLITES
MIDAS SATELLITES
#MILITARY SPACECRAFT
RECONNAISSANCE SPACECRAFT
#SATELLITES

MIDAS 5 SATELLITE
1404 3107
BT ARTIFICIAL SATELLITES
EARTH SATELLITES
MIDAS SATELLITES
#MILITARY SPACECRAFT
RECONNAISSANCE SPACECRAFT
#SATELLITES

MIDAS 6 SATELLITE
1404 3107
BT ARTIFICIAL SATELLITES
EARTH SATELLITES
MIDAS SATELLITES
#MILITARY SPACECRAFT
RECONNAISSANCE SPACECRAFT
#SATELLITES

MIDAS 7 SATELLITE
0705 3107
BT ARTIFICIAL SATELLITES
EARTH SATELLITES
MIDAS SATELLITES
#MILITARY SPACECRAFT
RECONNAISSANCE SPACECRAFT
#SATELLITES

MIDCOURSE GUIDANCE
2102 2103
BT #GUIDANCE (MOTION)
RT COMMAND GUIDANCE
INERTIAL GUIDANCE
INJECTION GUIDANCE
RENDEZVOUS GUIDANCE
SPACECRAFT GUIDANCE
TERMINAL GUIDANCE

MIDCOURSE TRAJECTORIES
1904 2102 2204 3101 3103 3105 3106
3108
BT #TRAJECTORIES
RT ASCENT TRAJECTORIES
BALLISTIC TRAJECTORIES
COASTING FLIGHT
DESCENT TRAJECTORIES
PARABOLIC FLIGHT

MIDDLE EAR
0404 0405 0408
BT #ANATOMY
EAR
SENSE ORGANS
RT SEMICIRCULAR CANALS

MIDDLE EAR PRESSURE
0404 0408
BT #PRESSURE

MIDLATITUDE ATMOSPHERE
1301 1302 1308 2003

NASA THESAURUS (POSTABLE TERMS)

AERODYNAMIC LOADS
AERODYNAMIC NOISE
AERODYNAMIC STABILITY
AERODYNAMIC STALLING
AERODYNAMICS
AEROELASTICITY
AEROEMBOLISM
AEROLOGY
AERONAUTICAL ENGINEERING
AERONAUTICS
AERONOMY
AEROS SATELLITE
AEROSINUSITIS
AEROSOLS
AEROSPACE ENGINEERING
AEROSPACE ENVIRONMENTS
AEROSPACE INDUSTRY
AEROSPACE MEDICINE
AEROSPACE SCIENCES
AEROSPACE SYSTEMS
AEROSPACE VEHICLES
AEROSPACEPLANES
AEROTHERMOCHEMISTRY
AEROTHERMODYNAMICS
AEROTHERMOELASTICITY
AEROZINE
AFFERENT NERVOUS SYSTEMS
AFFINITY
AFRICA
AFTERBODIES
AFTERBURNING
AFTERGLOWS
AFTERIMAGES
AGE FACTOR
AGENA A ROCKET VEHICLE
AGENA B RANGER PROGRAM
AGENA B ROCKET VEHICLE
AGENA C ROCKET VEHICLE
AGENA D ROCKET VEHICLE
AGENA ROCKET VEHICLES
AGENTS
AGGLOMERATION
AGGLUTINATION
AGGREGATES
AGING
AGING (BIOLOGY)
AGING (METALLURGY)
AGITATION
AGREEMENTS
AGRICULTURE
AIDS
AILERONS
AIR
AIR BREATHING ENGINES
AIR CARGO
AIR CONDITIONING
AIR CONDITIONING EQUIPMENT
AIR CONDUCTIVITY
AIR COOLING
AIR CURRENTS
AIR DEFENSE
AIR DUCTS
AIR FILTERS
AIR FLOW
AIR INTAKES
AIR JETS
AIR LAUNCHING
AIR LOCKS
AIR MAIL
AIR MASSES
AIR NAVIGATION
AIR POLLUTION
AIR PURIFICATION
AIR SAMPLING
AIR TO AIR MISSILES
AIR TO AIR REFUELING
AIR TRAFFIC
AIR TRAFFIC CONTROL
AIR TRANSPORTATION
AIRBORNE EQUIPMENT
AIRBORNE INFECTION
AIRBORNE RANGE AND ORBIT DETERMINATION
AIRCRAFT
AIRCRAFT ACCIDENT INVESTIGATION
AIRCRAFT ACCIDENTS
AIRCRAFT ANTENNAS
AIRCRAFT APPROACH SPACING
AIRCRAFT BRAKES
AIRCRAFT CARRIERS
AIRCRAFT COMMUNICATION
AIRCRAFT COMPARTMENTS
AIRCRAFT CONFIGURATIONS
AIRCRAFT CONTROL
AIRCRAFT DESIGN
AIRCRAFT DETECTION
AIRCRAFT ENGINES
AIRCRAFT FUEL SYSTEMS
AIRCRAFT FUELS
AIRCRAFT GUIDANCE
AIRCRAFT HAZARDS
AIRCRAFT HYDRAULIC SYSTEMS
AIRCRAFT INDUSTRY
AIRCRAFT INSTRUMENTS
AIRCRAFT LANDING
AIRCRAFT LAUNCHING DEVICES
AIRCRAFT LIGHTS
AIRCRAFT MAINTENANCE
AIRCRAFT MODELS
AIRCRAFT NOISE
AIRCRAFT PARTS
AIRCRAFT PERFORMANCE
AIRCRAFT PILOTS
AIRCRAFT PRODUCTION
AIRCRAFT RELIABILITY
AIRCRAFT SAFETY
AIRCRAFT SPECIFICATIONS
AIRCRAFT STABILITY
AIRCRAFT STRUCTURES
AIRCRAFT TIRES
AIRCRAFT WAKES
AIRFIELD SURFACE MOVEMENTS
AIRFOIL FENCES
AIRFOIL PROFILES
AIRFOILS
AIRFRAME MATERIALS
AIRFRAMES
AIRGLOW
AIRLINE OPERATIONS
AIRPLANE PRODUCTION COSTS
AIRPORT BEACONS
AIRPORT LIGHTS
AIRPORT PLANNING
AIRPORT SURFACE DETECTION EQUIPMENT
AIRPORT TOWERS
AIRPORTS
AIRSHIPS
AIRSPACE
AIRSPEED
AIRY FUNCTION
AJ- 10 ENGINE
AKERMANITE
ALAIS METEORITE
ALANINE
ALARM PROJECT
ALASKA
ALBEDO
ALBINISM
ALBUMINS
ALCOHOLS
ALDEHYDES
ALDOLASE
ALDOSTERONE
ALERTNESS
ALGAE
ALGEBRA
ALGERIA
ALGOL
ALGOL ENGINE
ALGORITHMS
ALIGNMENT
ALIPHATIC COMPOUNDS
ALKALI HALIDES
ALKALI METAL COMPOUNDS
ALKALI METALS
ALKALIES
ALKALINE BATTERIES
ALKALINE EARTH COMPOUNDS
ALKALINE EARTH OXIDES
ALKALOIDS
ALKALOSIS
ALKANES
ALKENES
ALKYD RESINS
ALKYL COMPOUNDS
ALKYLATES
ALKYLATION
ALKYLFERROCENE
ALKYLIDENE
ALKYNES
ALL SKY PHOTOGRAPHY

0300 AUXILIARY SYSTEMS

INCLUDES TERMINOLOGY ON ENERGY CONVERSION TECHNIQUES; FUEL CELLS AND CHEMICAL BATTERIES; AND AUXILIARY SYSTEMS (GENERAL). FOR RELATED TERMS SEE ALSO 0900 ELECTRONIC EQUIPMENT, 1400 INSTRUMENTATION AND PHOTOGRAPHY, AND 2800 PROPULSION SYSTEMS.

0301 ENERGY CONVERSION TECHNIQUES

- ABSORPTION CROSS SECTIONS
- ACCUMULATORS
- ASTEC SOLAR TURBOELECTRIC GENERATOR
- BURSTS
- CESIUM DIODES
- CESIUM ENGINES
- CONCENTRATORS
- DIRECT POWER GENERATORS
- DYNAMOMETERS
- ELECTRIC MOTORS
- ELECTRIC POWER
- ELECTRIC PROPULSION
- ELECTROMECHANICAL DEVICES
- ELECTROMECHANICS
- ELECTROSTATIC GENERATORS
- ENERGY
- ENERGY ABSORPTION FILMS
- ENERGY CONVERSION
- ENERGY CONVERSION EFFICIENCY
- ENERGY SOURCES
- ENGINES
- FURNACES
- GENERATORS
- GOLAY DETECTOR CELLS
- HEAT SOURCES
- HEAT STORAGE
- INDUCTION HEATING
- INTERNAL CONVERSION
- KNUDSEN GAGES
- LUNAR ENVIRONMENT
- MOTORS
- NICKEL ZINC BATTERIES
- ONBOARD EQUIPMENT
- PHOTOELECTRIC CELLS
- PHOTOELECTRIC GENERATORS
- PHOTOVOLTAIC CELLS
- PHOTOVOLTAIC CONVERSION
- PHOTOVOLTAIC EFFECT
- POWER
- POWER PLANTS
- RADIATION
- RADIOISOTOPE BATTERIES
- ROTATING GENERATORS
- SNAP
- SOLAR AUXILIARY POWER UNITS
- SOLAR CELLS
- SOLAR COLLECTORS
- SOLAR ENERGY ABSORBERS
- SOLAR GENERATORS
- SOLAR INSTRUMENTS
- SUNFLOWER POWER SYSTEM
- THERMIONIC CONVERTERS
- THERMOCOUPLES
- THERMONUCLEAR POWER GENERATION
- TURBINES
- TURBOGENERATORS

0302 FUEL CELLS AND CHEMICAL BATTERIES

- AC GENERATORS
- ALKALINE BATTERIES
- ALTERNATING CURRENT
- AMPLIDYNES
- ARC GENERATORS
- BIOCHEMICAL FUEL CELLS
- CELL ANODES
- CELL CATHODES
- CELLS
- CHEMICAL AUXILIARY POWER UNITS
- CHEMICAL ENERGY
- COLLOIDAL GENERATORS
- DEPOLARIZATION
- DIRECT CURRENT
- DIRECT POWER GENERATORS
- DRY CELLS
- ELECTRIC BATTERIES
- ELECTRIC CELLS
- ELECTRIC EQUIPMENT
- ELECTRIC GENERATORS
- ELECTROCATALYSTS
- ELECTROCHEMICAL CELLS
- ELECTROSTATIC GENERATORS
- FEED SYSTEMS
- FISSION ELECTRIC CELLS
- FUEL CELLS
- GENERATORS
- HEAT STORAGE
- HUMAN WASTES
- HYDROGEN OXYGEN FUEL CELLS
- ION EXCHANGE MEMBRANE ELECTROLYTES
- MAGNESIUM CELLS
- METAL AIR BATTERIES
- NICKEL CADMIUM BATTERIES
- NICKEL ZINC BATTERIES
- PHOTOVOLTAIC CONVERSION
- POWER LIMITERS
- POWER PLANTS
- PRIMARY BATTERIES
- PROVISIONING
- REFUELING
- REGENERATIVE FUEL CELLS
- SEPARATORS
- SILVER CADMIUM BATTERIES
- SILVER ZINC BATTERIES
- SPACE POWER UNIT REACTORS
- STORAGE BATTERIES
- THERMAL BATTERIES
- THERMOELECTRIC COOLING
- THERMOELECTRIC GENERATORS
- THERMOELECTRIC POWER GENERATION
- THERMOMAGNETIC COOLING
- THERMOPILES
- TRANSFORMERS
- VAN DE GRAAFF ACCELERATORS
- VOLTAGE GENERATORS
- WET CELLS

0303 AUXILIARY SYSTEMS (GENERAL)

- ACCESSORIES
- ACTUATOR DISKS
- ACTUATORS
- AIR LOCKS
- AIRCRAFT FUEL SYSTEMS
- AIRCRAFT HYDRAULIC SYSTEMS
- ALTERNATING CURRENT
- AMPLIDYNES
- AUXILIARY POWER SOURCES
- BASE HEATING
- BLEEDING
- BLOCKING
- BRAKES
- BRAKING
- BRAYTON CYCLE
- BYPASSES
- CAPILLARIES
- CARBURETORS
- CHEMICAL AUXILIARY POWER UNITS
- CLOCKS
- COMPRESSOR EFFICIENCY
- COMPRESSORS
- CONDENSERS (LIQUIFIERS)
- CONTROL RODS
- CORDAGE
- DISTRIBUTORS
- EARPHONES
- ELECTRIC CELLS
- ELECTRIC CONNECTORS
- ELECTRIC CONTACTS
- ELECTRIC CONTROL
- ELECTRIC ENERGY STORAGE
- ELECTRIC EQUIPMENT
- ELECTRIC GENERATORS
- ELECTRIC IGNITION
- ELECTRIC MOTORS
- ELECTRIC NETWORKS
- ELECTRIC POWER
- ELECTRIC WIRE
- ELECTROHYDRAULIC FORMING
- ELECTROMAGNETIC COMPATIBILITY
- ENERGY STORAGE
- FLUID POWER
- FLYWHEELS
- FUEL CELLS
- GAS GENERATORS
- GLANDS (SEALS)
- HEAT PUMPS
- HEAT SOURCES
- HYDRAULIC CONTROL
- HYDRAULIC EQUIPMENT
- HYDRAULIC FLUIDS

E
ARGO E-5 ROCKET VEHICLE
ATLAS E ICBM
E REGION
E- 1 LAYER
E- 2 LAYER
E- 3 LAYER
E-2 AIRCRAFT
SPORADIC E LAYER
EAR
EAR PRESSURE TEST
EAR PROTECTORS
MIDDLE EAR
MIDDLE EAR PRESSURE
EARLY
BALLISTIC MISSILE EARLY WARNING SYSTEM
EARLY BIRD SATELLITES
EARLY STARS
EARLY WARNING SYSTEMS
EARS
ARTIFICIAL EARS
EARTH
ALKALINE EARTH COMPOUNDS
ALKALINE EARTH OXIDES
EARTH (PLANET)
EARTH ALBEDO
EARTH ATMOSPHERE
EARTH AXIS
EARTH CORE
EARTH CRUST
EARTH HYDROSPHERE
EARTH MANTLE
EARTH MOTION
EARTH MOVEMENTS
EARTH ORBITAL RENDEZVOUS
EARTH ORBITS
EARTH PLANETARY STRUCTURE
EARTH ROTATION
EARTH SATELLITES
EARTH SURFACE
EARTH-MARS TRAJECTORIES
EARTH-MOON SYSTEM
EARTH-MOON TRAJECTORIES
MOON-EARTH TRAJECTORIES
RARE EARTH ALLOYS
RARE EARTH COMPOUNDS
RARE EARTH ELEMENTS
RETURN TO EARTH SPACE FLIGHT
EBERT
EBERT SPECTROMETERS
EC
EC-121 AIRCRAFT
ECHELETTE
ECHELETTE GRATINGS
ECHO
ECHO PROJECT
ECHO SATELLITES
ECHO SUPPRESSORS
ECHO 1 SATELLITE
ECHO 2 SATELLITE
ECHOES
AURORAL ECHOES
LUNAR ECHOES
LUNAR RADAR ECHOES
RADAR ECHOES
RADIO ECHOES
SOLAR RADAR ECHOES
VENUS RADAR ECHOES
ECLIPSE
ECLIPSE PROJECT
ECLIPSES
LUNAR ECLIPSES
SOLAR ECLIPSES
ECLIPSING
ECLIPSING BINARY STARS
ECOLOGICAL
CLOSED ECOLOGICAL SYSTEMS
ECONOMICS
DEMAND (ECONOMICS)
EDDINGTON
EDDINGTON APPROXIMATION
EDDY
EDDY CURRENTS
EDGE
EDGE DISLOCATIONS
EDGE LOADING
LEADING EDGE SLATS
LEADING EDGE SWEEP
TRAILING-EDGE FLAPS
EDGES
LEADING EDGES
EDGES
SHARP LEADING EDGES
TRAILING EDGES
EDITAR
EDITAR SYSTEM
EDUCATIONAL
EDUCATIONAL TELEVISION
EFFECT
AUGER EFFECT
BARKHAUSEN EFFECT
BAUSCHINGER EFFECT
BRILLOUIN EFFECT
CAPTURE EFFECT
COANDA EFFECT
COMPTON EFFECT
CORIOLIS EFFECT
CUSHIONCRAFT GROUND EFFECT MACHINE
DOPPLER EFFECT
DOPPLER-FIZEAU EFFECT
ELECTRO-OPTICAL EFFECT
ETTINGSHAUSEN EFFECT
FARADAY EFFECT
FIELD EFFECT TRANSISTORS
FIZEAU EFFECT
GREENHOUSE EFFECT
GROUND EFFECT
GROUND EFFECT MACHINES
GUNN EFFECT
HALL EFFECT
HOVERCRAFT GROUND EFFECT MACHINES
ISOTOPE EFFECT
JOULE-THOMSON EFFECT
KAG-3 GROUND EFFECT MACHINE
KERR ELECTROOPTICAL EFFECT
KERR MAGNETOOPTICAL EFFECT
KIRKENDALL EFFECT
LUXEMBOURG EFFECT
MAGNUS EFFECT
MOSSBAUER EFFECT
NERNST-ETTINGSHAUSEN EFFECT
NONOHMIC EFFECT
NUCLEAR EXPLOSION EFFECT
PENNING EFFECT
PHOTOELECTRIC EFFECT
PHOTOVOLTAIC EFFECT
PINCH EFFECT
POYNTING-ROBERTSON EFFECT
RAMSAUER EFFECT
SCALE EFFECT
SCREEN EFFECT
SEEBECK EFFECT
STARK EFFECT
SUHL EFFECT
SWEEP EFFECT
UMKEHR EFFECT
VA-3 GROUND EFFECT MACHINE
VOIGT EFFECT
WESTLAND GROUND EFFECT MACHINES
ZEEMAN EFFECT
ZENER EFFECT
EFFECTIVENESS
RELATIVE BIOLOGICAL EFFECTIVENESS (RBE)
EFFECTS
BIOLOGICAL EFFECTS
CHEMICAL EFFECTS
COMPRESSIBILITY EFFECTS
GALVANOMAGNETIC EFFECTS
GRAVITATIONAL EFFECTS
JET BLAST EFFECTS
KERR EFFECTS
LONG TERM EFFECTS
LUNAR EFFECTS
LUNAR GRAVITATIONAL EFFECTS
MAGNETIC EFFECTS
MOIRE EFFECTS
PATHOLOGICAL EFFECTS
PELTIER EFFECTS
PHOTOELECTROMAGNETIC EFFECTS
PHOTOMAGNETIC EFFECTS
PHYSIOLOGICAL EFFECTS
PRESSURE EFFECTS
PSYCHOLOGICAL EFFECTS
RADIATION EFFECTS
REENTRY EFFECTS
RELATIVISTIC EFFECTS
SOLAR ACTIVITY EFFECTS
SURFACE ROUGHNESS EFFECTS

Sample of an early Permuted Index produced by ARIES for DoD Project LEX

PERMUTED INDEX PAGE 0009

AGGLOMERATION
⊙AGGLOMERATORS
•AGGRADATION
AGGREGATE
EXPOSED AGGREGATE CONCRETE
LIGHTWEIGHT AGGREGATE CONCRETES
PREPLACED AGGREGATE CONCRETE
AGGREGATES
EXPANDED CLAY AGGREGATES
EXPANDED SHALE AGGREGATES
EXPANDED SLAG AGGREGATES
EXPANDED SLATE AGGREGATES
LIGHTWEIGHT AGGREGATES
SOIL AGGREGATES
•AGGREGATION
AGGRESSION
AGING
AGING (METALLURGY)
AGING TESTS (MATERIALS)
ARTIFICIAL AGING (METALLURGY)
QUENCH AGING
•STRAIN AGING
AGITATION
AGITATORS
DISPERSERS (AGITATORS)
•AGM
AGONISTIC
AGONISTIC BEHAVIOR
AGRANULOCYTOSIS
•AGREEMENTS
ARMS CONTROL AGREEMENTS
BILATERAL AGREEMENTS
MULTILATERAL AGREEMENTS
UNILATERAL AGREEMENTS
AGRICULTURAL
•AGRICULTURAL AIRCRAFT
AGRICULTURAL ECONOMICS
AGRICULTURAL ENGINEERING
AGRICULTURAL MACHINERY
AGRICULTURAL PRODUCTS
AGRICULTURAL WASTES
•DISKS (AGRICULTURAL)
AGRICULTURE
AGROBACTERIUM
AGRONOMY
AID
FIRST AID
FIRST AID KITS
FOREIGN AID
AIDS
HEARING AIDS
LANDING AIDS
NAVIGATIONAL AIDS
PENETRATION AIDS (COMBAT)
VISUAL AIDS

A dot preceding a term denotes it is a USE reference in the Thesaurus

Bold print (overstrike) denotes a permuting word which itself is a term in the Thesaurus

Light print denotes a permuting word which is not itself a term in the Thesaurus

IBM FORM NO. 813

LUNAR MODULE
MARS EXCURSION MODULE

SPACECRAFT CONTROL
SATELLITE CONTROL
SATELLITE ATTITUDE CONTROL

SPECTRA
ENERGY SPECTRA
ELECTRONIC SPECTRA
NEUTRON SPECTRA
MASS SPECTRA
MOLECULAR SPECTRA
ELECTRONIC SPECTRA
RAMAN SPECTRA
VIBRATIONAL SPECTRA
NOISE SPECTRA
OXYGEN SPECTRA
PLASMA SPECTRA
POWER SPECTRA
RADIATION SPECTRA
ABSORPTION SPECTRA
FRAUNHOFER LINES
HERZBERG BANDS
TELLURIC LINES
ELECTROMAGNETIC SPECTRA
INFRARED SPECTRA
LINE SPECTRA
BALMER SERIES
D LINES
ELECTRONIC SPECTRA
FRAUNHOFER LINES
H LINES
H ALPHA LINE
H BETA LINE
H GAMMA LINE
K LINES
LYMAN SPECTRA
PASCHEN SERIES
RYDBERG SERIES
TELLURIC LINES
RADIO SPECTRA
MICROWAVE SPECTRA
RAMAN SPECTRA
STELLAR SPECTRA
SOLAR SPECTRA
ULTRAVIOLET SPECTRA
VIBRATIONAL SPECTRA
EMISSION SPECTRA
SHOCK SPECTRA
SPECTRAL BANDS
ABSORPTION SPECTRA
FRAUNHOFER LINES
HERZBERG BANDS
TELLURIC LINES
PHOTOLUMINESCENT BANDS
SCHUMANN-RUNGE BANDS
SWAN BANDS
VEGARD-KAPLAN BANDS

SPECTROSCOPY
ASTRONOMICAL SPECTROSCOPY
AURORAL SPECTROSCOPY
GAS SPECTROSCOPY
INFRARED SPECTROSCOPY
MAGNETIC SPECTROSCOPY
MASS SPECTROSCOPY
MOLECULAR SPECTROSCOPY
RAMAN SPECTROSCOPY
NUCLEAR RADIATION SPECTROSCOPY
OPTICAL EMISSION SPECTROSCOPY
RADIO SPECTROSCOPY
SPECTROPHOTOGRAPHY
SPECTROPHOTOMETRY
STELLAR SPECTROPHOTOMETRY
SPECTROSCOPIC ANALYSIS
ULTRAVIOLET SPECTROSCOPY
VACUUM SPECTROSCOPY
X RAY SPECTROSCOPY

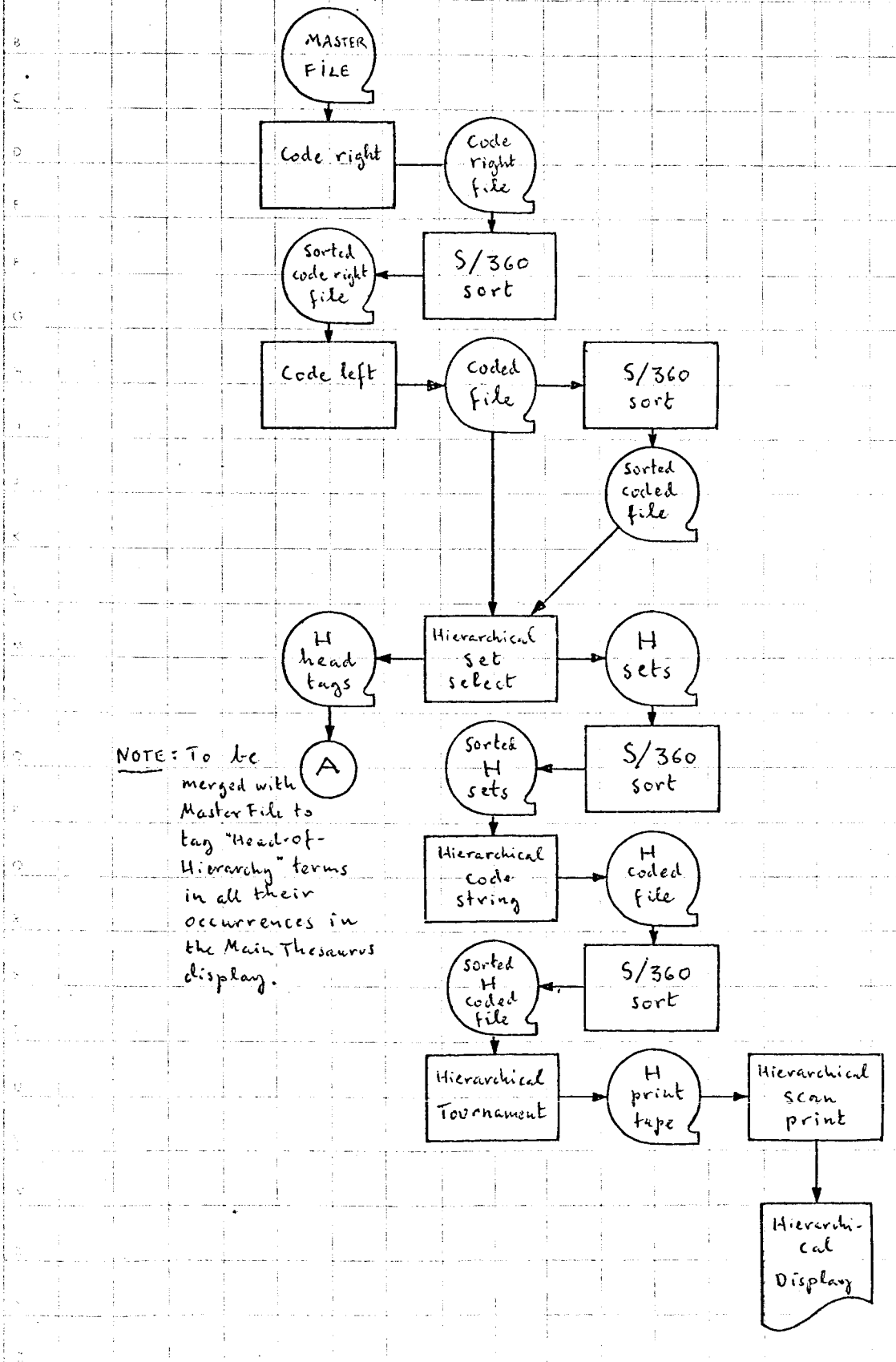
SPEECH
ARTICULATION
CONVERSATION
PHONEMES
PHONETICS

281-684 O-67—10

ARIES CORPORATION DIAGRAMMING AND CHARTING WORKSHEET

Application ARIES Thesaurus Package Date _____ Page 1 of 1

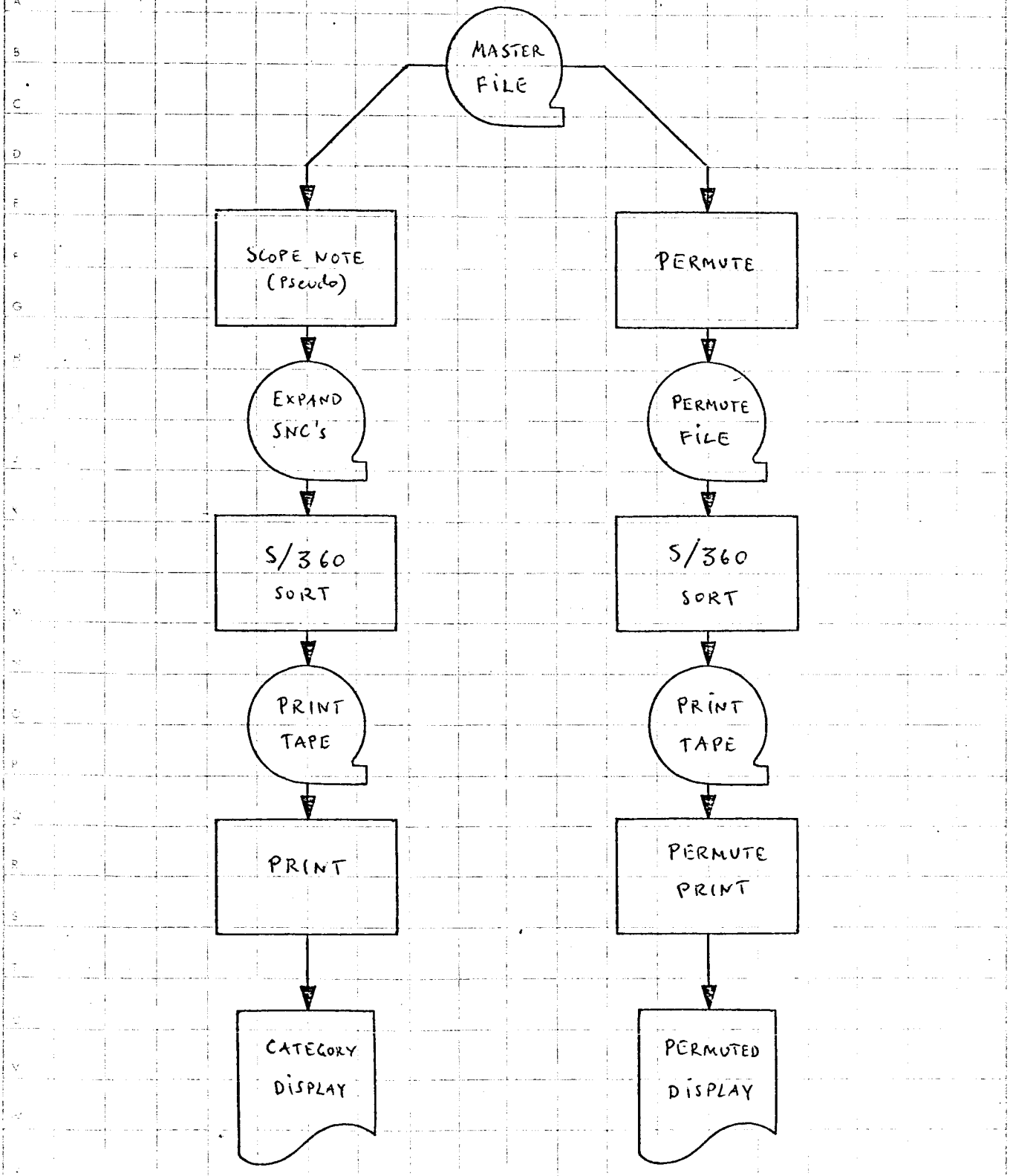
Procedure Hierarchical Display Drawn By Staffan Rosenborg



ARIES CORPORATION DIAGRAMMING AND CHARTING WORKSHEET

Application ARIES Thesaurus Package Date _____ Page 1 of 1

Procedure Category and Permuted Displays Drawn By Staffan Rosenborg



JUL 18 1967 LEX EDIT COPY 18 JUL 67 PAGE 980

```
0116172   BT  Arthropoda
0116173       Invertebrates
0116174  Pentenes  0703
0116176   Acyclic hydrocarbons with molecular
0116177   formula C'5H'10'0
0116178   BT  Aliphatic acyclic hydrocarbons
0116179       Aliphatic hydrocarbons
0116180       Alkene compounds
0116181       Alkene hydrocarbons
0116182       Hydrocarbons
0116183       Unsaturated hydrocarbons
0116184  Pentobarbital  0615 0703
0116186   BT  Barbiturates
0116187   RT  Hypnotics and sedatives
0116188  Pentodes  0901
0116190   UF  Pentode tubes
0116191   RT .Electron tubes
0116192      Soniconductor devices
0116193      .Transistors
0116194  Pentode tubes
0116195   USE Electron tubes
0116196   and Pentodes
0116197  Pentolite  1901
0116199   BT  Explosives
0116200  Pentoses  0601
0116202   BT  Carbohydrates
0116203       Monosaccharides
0116204       Sugars
0116205   NT  Arabinose
0116206       Ribose
0116207       Xylose
0116208  Pentylenetetrazol  0615 0703
0116210   BT  Heterocyclic compounds
0116211       Nitrogen heterocyclic compounds
0116212       Nitrogen heterocyclics with 4 N
0116213   RT  CNS stimulants
0116214  People
0116215   USE Personnel
0116216  Pepsin  0601
0116213   BT  Enzymes
0116219       Hydrolases
0116220       Peptide hydrolases
0116221       Peptide peptidohydrolases
0116222   BT  Gastric juice
0116223  Peptic ulcers  0605
0116225   BT  Gastrointestinal diseases
0116226       Stomach and duodenal diseases
0116227       Ulcers
0116228   RT  Dyspepsia
0116229       Esophagitis
0116230       Gastrectomy
0116231       Hematemesis
0116232       .Psychosomatic disorders
```

Scope note

Trigger for sub-script

Subject classification

Reciprocal of multiple USE

Use for

Term has narrower term(s)

Multiple USE

Broader term

Narrower term

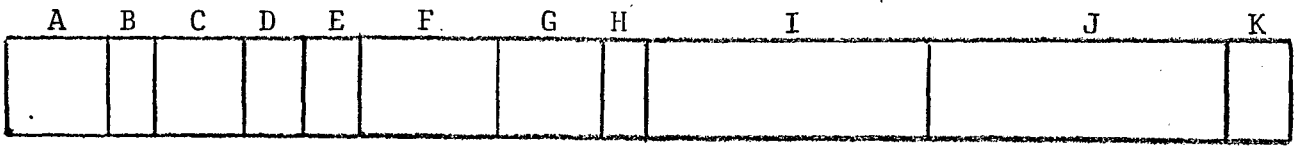
Related term

0006791	PHERONOME		...T	QT
	QT			
0006792	BT	ANIMAL COMMUNICATION		QT
0006793		SOCIAL BEHAVIOR ANIMAL		QT
0006794	PHILOSOPHY		MP.T	PZ
	PZ05			
	RT	ESTHETICS		
	NT	LOGIC		
0006798	PHOBIA			PZ
	PZ05			
	RT	ANXIETY		
		FEAR		
0006801	PHOBIAS		MR.T	RZ
	RZ05			
0006802	BT	MENTAL DISORDERS		QT
0006803		NEUROSES		MZ
0006804		OBSESSIVE COMPULSIVE NEUROSES		MZ
0006805	PHONEME		.P.T	PZ
	PZ05			
0006806	USE	PHONETICS		PZ
0006807		WORD		PZ
0006808	BT	CULTURAL ANTHROPOLOGY		QT
0006809		GENERAL LINGUISTICS		QT
0006810		PHONOLOGY		QT
	RT	DISTINCTIVE FEATURES		
0006812	PHONEMIC REGRESSION		...T	QT
	QT			
0006813	BT	LANGUAGE/SPEECH DISORDERS		QT
0006814		PSYCHOLINGUISTICS		QT
0006815	PHONETICS		MPRT	PZ
	PZ05			
	UF	PHONEME		
0006817	BT	ACCUSTICS		QT
0006818		ANIMAL COMMUNICATION		QT
0006819		GENERAL LINGUISTICS		QT
0006820		LANGUAGE		QT
0006821		SPEECH		PZ
0006822	PHONOLOGICAL STRESS		...T	QT
	QT			
0006823	BT	CULTURAL ANTHROPOLOGY		QT
0006824		GENERAL LINGUISTICS		QT
0006825		PHONOLOGY		QT
12 0006826	PHONOLOGY		...T	QT
11	QT			
10 0006827	BT	ANTHROPOLOGY AND SOCIOLOGY		QT
9 0006828		CULTURAL ANTHROPOLOGY		TZ
8 0006829		GENERAL LINGUISTICS		QT
7 0006830		LINGUISTICS		QT
6	NT	DISTINCTIVE FEATURES		
5		INTONATION PATTERN		
4		PHONEME		
3		PHONOLOGICAL STRESS		
2 0007947	PHYCOSIS AND PSYCHOTICS		.T..	
	RT	GENETICS OF ALCOHOLISM		
0006841	PHYSICAL AGGRESSION			QT

DESCRIPTION

OF

MAGNETIC TAPE LAYOUT FOR LEX DATA STORE



109 CHARACTER RECORD -- 10 FIELDS -- IBM MODE

(10 records per block; 800 BPI recording density; 9-track)

FIELD CONTENT

FIELD	POSITION(S)	CONTENT
A	1-7	Reserved for line sequence code employed in maintenance.
B	8	Term relationship code; see code key below.
C	9-12	Reserved for system use
D	13-14	Source vocabulary code (repeated for scope notes only)
E	15	Line sequence code for scope note
F	16-19	COSATI subject category code
G	20-21	Source vocabulary code - all entries
H	24	Reserved for LEX Disposition Code
I	25-66	42-character term entry, scope note line, or subject category codes; this field is also referred to as the sub-term field; see key below.
J	67-108	42-character term entry; this field also referred to as main-term field.
K	109	Record Mark (retained from IBM 1401 computer generation).

APPENDIX C

PROGRESS REPORT FROM SUB-CONTRACTOR (LEX-INC)

PROGRESS REPORT

Subcontract 212056 Under NASA Contract NASW-1629

Discussion

A. Editing of the NASA Thesaurus Review Copy

1. Materials Provided by NASA and ARIES Corporation

As input to the thesaurus editing operation, NASA provided the following materials:

NASA Thesaurus -- Review Copy.

Copies of Term Review Forms (TRF's)
completed during the period 6-1-67 to 7-5-67.

Subject Authority List (dated June 1967).

Main Term (and Use Reference) List, based on the Thesaurus Review Copy.

Phase 3 (Pre-Review Copy)
Computer printout of the Thesaurus.

List of Terms explicitly deleted.

In addition, so that the editing effort could be expedited and made more effective, the prime contractor (ARIES Corporation) provided the following materials:

NASA Thesaurus -- Computer printout ("Phase 7")
essentially equivalent to the Review Copy.

A list of changes made to the "Phase 7"
printout to create the Review Copy.

A loan copy of the latest computer printout of the
LEX-EJC Thesaurus, including permuted term index.

2. Editing Objectives

The primary objective of the editing effort, as detailed in conversations with NASA-STID personnel, was to review all work performed by the "NASA task force" between June 1, 1967 and July 5, 1967. This task force effort consisted of a review and updating of the products of a previous contract effort

and the resultant addition or deletion of thesaurus entries. An additional objective of the editing effort, however, was the "overview" of the entire thesaurus, such that recommendations for improvements in content and internal consistency could be provided.

3. Editing Procedures

The initial editing procedure consisted of reviewing each TRF copy, as it had been translated into the Thesaurus Review Copy, for accuracy and completeness. About 2500-3000 TRF copies were provided. Notations (i.e., corrections and additional entries) were entered in red ink in the "Phase 7" computer printout, rather than in the review copy, to facilitate clarity of entry (the Review Copy itself was too greatly size - reduced during reproduction to permit voluminous notations) and to increase the efficiency of later keypunching of the notations (in that changes to the Review Copy are handled by line numbers -- which appear on the "Phase 7" printout but not on the Review Copy itself). The LEX-EJC Thesaurus printout was used as a reference in detecting missing crossreference entries; i.e., if a crossreference term appeared in the LEX-EJC Thesaurus, and if it also appeared in the NASA Main Term List, then the cross-reference term was a candidate for addition to the NASA Thesaurus. Because the Main Term List was available only after August 8, 1967, the NASA Subject Authority List had to be used in lieu of the Main Term List prior to that time. This means that some of the added crossreferences may no longer be valid NASA terms; they may have been deleted or changed in word form.

The overview of the Review Copy indicated that a number of internal inconsistencies existed, as follows:

a. "Array" terms with BT or NT crossreferences.

b. "Use Reference" terms with category numbers.

c. Essentially duplicate main terms (i.e., singular and plural, misspelled and spelled correctly, etc.)

d. Many invalid BT/NT crossreferences were caused by a few erroneous but basic BT or NT entries; for example, X-19 Aircraft was erroneously (via TRF entry or key punching) given a NT (rather than a BT) crossreference to Research Aircraft, and the computer promptly made all Research Aircraft (X-1 Aircraft, X-2 Aircraft, etc.) narrower terms for X-19 Aircraft, and X-19 Aircraft a broader term for all of the various Research Aircraft.

Because the above items (a) and (c) were easily and quickly corrected, such action was completed as a part of the Thesaurus overview, and the appropriate entries were made, again in red, in the "Phase 7" printout. Item (b) was made known to NASA-STID personnel, who agreed to see that category numbers were removed from "use references"; this can probably best be done by computer. Item (d) was also reported to NASA-STID personnel, who asked us to make the appropriate corrections partially in lieu of the educational seminar (especially because many of the seminar comments would have been concerned with the numerous invalid BT/NT crossreferences).

When it became apparent that NASA would be unable to organize the educational seminar for the NASA research centers, and thereby to obtain "face-to-face" comments by the research centers on the Review Copy, we were asked to perform several tasks to provide at least a part of the input which would have been provided by such comments. The first of these additional tasks was to validate all BT crossreferences (thereby also validating all NT crossreferences, in view of the capabilities of ARIES' thesaurus assembly programs). The second additional task was to review carefully all "multiple use references". The third additional task was to provide the bases for a procedure for converting existing machine-sensible indexes from the original vocabulary to the "new" one. The performance of these additional tasks is described below.

The validity of BT crossreferences was determined by scanning the entire 152,743 lines of the "Phase 7" printout for BT crossreferences, evaluating each such crossreference for validity, and (if invalid) marking the entry for

appropriate change or deletion (e.g., X-19 Aircraft was deleted as a BT for X-1 Aircraft and X Rays was deleted as a BT for Aerospace Medicine). Many invalid BT crossreferences were appropriate as RT (i.e., related term) cross-references, and in such instances the crossreference type was changed. All entries were marked in red in the "Phase 7" printout.

The validity of the "multiple use references" was determined by scanning the Main Term List for occurrences of such terms, by reevaluating the utility of the term (employing usage statistics available in the Subject Authority List when appropriate), and by marking in the "Phase 7" printout any changes. A number of the "multiple use references" were found properly to be "single use references", but most of them were retained. The "multiple use reference" serves a valuable function when a term is seldom (or -- at most -- moderately) used in indexing and when its referent belongs to two or more classes. Consider the following example:

Boeing Military Aircraft	(used 8 times in indexing)
Use Boeing Aircraft	("new" term)
(and) Military Aircraft	(used 228 times in indexing)

The "new" term, Boeing Aircraft, is a much more useful term than was Boeing Military Aircraft. In fact, we are able to assign to Boeing Aircraft a total of 16 narrower terms, each the name of a specific aircraft model. Thus we have moved our active, "postable" terminology from the overly specific to the more appropriately general. Nevertheless, should the need ever arise to reconstitute Boeing Military Aircraft as a "postable" term, such can be done by searching for the logical intersection of Boeing Aircraft and Military Aircraft and thereby reconstituting the postings for Boeing Military Aircraft. The "multiple use reference" technique thereby serves as a "safe" means of "demoting" less useful terms without destroying irretrievably their postings. Reconstruction of such a term's postings is essentially a mechanical task, whereas reconstruction of a "single use reference's" postings would be an almost totally cerebral task.

The problem of converting the existing machine-sensible indexes to the "new" vocabulary is discussed in the following section of this report.

B. Index Conversion

1. Statement of the Problem

NASA's existing machine-sensible indexes are now based on its "original" (pre-Thesaurus) vocabulary. It will ultimately be necessary to convert these indexes to the "new" vocabulary. It is thus necessary to provide a "trail" from each term in the "original" vocabulary to a term in the "new" vocabulary. The problem assigned to us is "How can this terminological trail be created?"

During the original contract work (and during the subsequent NASA task force work) which created the raw material for the Thesaurus (the completed TRF's), each TRF was given a two-digit "trail" code, indicating the action to be taken with respect to each "original" term's postings. Codes beginning with "4" indicated no change in term word-form; "41" or "42" indicated that the term would now be "unpostable", and the term (or terms) "used for" the term-at-hand and to which the term-at-hand's postings were to be transferred, was written on the TRF. A "40" code indicated no change in either word-form or postings. Codes beginning with "5" indicated only a change in term word-form; a "50" indicated that the "new" term was "postable", and "51" or "52" meant that the term was a "single" or "multiple use reference", with action on postings indicated as for the "41" and "42" codes. Codes beginning with "6" indicated terms to be deleted; "60" indicated deletion of the term and its postings, while "61" and "62" indicated that the postings should be transferred to one or more other terms.

At the time of TRF completion, these "trail" codes were still valid. However, between that time and the time when we began work under this sub-contract, hundreds of changes (corrections, deletions, etc.) were made to the computer Thesaurus record without making corresponding changes to the "trail" code record. In not a few instances, changes were no doubt made to changes, with the result that consistent correspondence no longer exists between the terms to which "trail" codes were assigned and the present terms, and between the terms to which postings were to be transferred 'per the TRF record) and the present terms. Hence the problem "How can the terminological trail be re-created"?

We discard immediately the alternative of trying to recapture, to reconstitute, and to act upon all changes made to the computer Thesaurus file as being too complex and uncertain. The best remaining alternative involves use of both computer and cerebral techniques, but it is at least reasonably straightforward, is assured of success, and is therefore recommended. The other alternative is to re-code all 19,000 original terms by reference to the TRF's and to the Thesaurus itself; this approach is not recommended because of its great manpower requirements.

2. Index Conversion Procedure

a. Nomenclature

T_o Terms in "originial" (pre-Thesaurus) vocabulary.

T_t Terms in Thesaurus vocabulary.

T_u Terms to be used in lieu of others in Thesaurus vocabulary.

T_x Terms to which postings on "original" terms were to be transferred (per TRF's).

T_{xt} Terms which are in both T_x and T_t.

T_{ot} Terms which are in both T_o and T_t.

T_{xy} Terms which are in T_x but not T_t.

T_{oy} Terms which are in T_o but not T_t.

(Symbols in parentheses and separated with a dash indicate double-field records.)

b. Computer Files Available

File 1 -- T_o with associated T_x, assumed available from keypunching of the TRF's; "trail" code (also available) is not needed, except for the "60" code. These records should be carefully proofread from the TRF's.

File 2 -- T_o, the "original" vocabulary term list; this may even be the most up-to-date vocabulary list available.

File 3 -- $T_t - T_u$, the Thesaurus main term list, together with any terms which are to be "used" in lieu of any T_t.

c. Detailed Procedure

We wish to create four final files:

(1) File U -- terms (with their postings) unchanged from the original.

(2) File X -- "original" terms together with the Thesaurus terms to which the postings are to be transferred.

(3) File D -- "original" terms which are to be deleted together with their postings.

(4) File R -- an ultimate residue of "original" terms which have not been assigned to one of the first three files and which must be "researched" individually.

The first step is to match the T_o records of File 1 with the T_t records of File 3. Matching records should be transferred from File 3 to File 4 ($T_{ot} - T_u$). Those records of File 4 which have no T_u record should be transferred to File U (T_{ot}). Those records of File 4 which have T_u records should be transferred to File 5, with the T_u records converted to the T_t field, such that File 5 ($T_{ot} - T_t$) is the beginning of File X.

Non-matching records from the first step (matching of File 1 with File 3) should be written from File 1 ($T_o - T_x$) onto File 6 ($T_{oy} - T_x$). File 4 ($T_{ot} - T_u$) should then be matched with File 2 (T_o) and the non-matching T_o records of File 2 should be written as File 7 (T_{oy}), which in turn should be merged with File 6 to create File 8, a complete record of ($T_{oy} - T_x$). Duplicate T_{oy} records should not be written in File 8.

File 8 ($T_{oy} - T_x$) should then be matched with the T_o records in File 1 and the matching entries in File 1 which are coded with a "60" trail code should be deleted, with non-matching records of File 8 written on File 9 ($T_{oy} - T_x$) and matching records written on File 10, which is the beginning of File D.

File 9 should then be processed to write on File 11 ($T_{oy} - T_x$) those records which have T_x components and on File 12 (T_{oy} -- the beginning of File R) those records which do not have T_x components. File 11 should then be sorted by T_x to create File 13 ($T_x - T_{oy}$).

File 13 is now matched with File 3 ($T_t - T_u$) to find records with matching T_x and T_t entries. On File 14 ($T_t - T_{oy}$) write the entries from File 13 wh which have T_x matching T_t on File 3, except convert the resulting T_{xt} entry on File 14 to the T_u entry taken from File 3 when a T_u entry is present. On File 15 ($T_{xy} - T_{oy}$) write the residue of records from File 13.

Until this point all processes have been via computer, and File 15 will be a relatively small file. Cerebrally convert the T_{xy} components of the File 15 records to T_t entries by finding appropriate T_t terms in the printout of File 3 (Main Term and "use reference" list). Merge the "converted" File 15 with File 14 to obtain File 16 ($T_t - T_{oy}$). Sort File 16 by the T_{oy} component to create File 17 ($T_{oy} - T_t$) and merge File 17 with File 5 to create File 18 ($T_o - T_t$), the continuation of the creation of File X.

We are now left with only File 12 (T_{oy}), a record of all "original" terms which are not in the Thesaurus vocabulary, which did not have assigned to them any kind of term to which their postings were to be transferred, and which had not been explicitly deleted from the vocabulary. File 12, when combined with any residue from the cerebral conversion effort performed on File 15 (see above) thus becomes File R -- subjects to be "researched" individually and assigned to File D (original terms deleted with their postings), File X (original terms not now postable but "tagged" with postable Thesaurus terms to which postings are to be transferred), or File U (which implies adding the "original" term to the Thesaurus, with all its crossreferences). These various reassignments of File R terms, when added to File 18 and File 10 become Files X and D respectively. It is expected that few reassignments will be made from File R to the already-existing File U.

d. Remarks

Other than a small amount of human (rather than computer) effort in proofing File 1 and in converting Files 12 and 15 (and all these files will be reasonably small), all work in the above procedure is via computer. Nearly

all the computer programs require only standard, "canned" routines, and those which do not are quite simple. There are 11 computer runs involved, all on small files; hence running time will not be a significant cost factor.

Once Files U, X and D have been assembled, they can be used respectively to (1) transfer terms and postings unchanged from the present machine-sensible indexes to the new ones, (2) to transfer postings only from terms in the present indexes to terms in the new ones, and (3) to delete both terms and postings from the old indexes. Whereas the procedures to create Files U, X and D are not costly in time or funds, the actual index conversion involves very large files (some of which will have to be sorted twice) and will therefore be relatively costly in computer time.

Lex - Inc
Information Systems Consultation

7037 Wolftree Lane
Rockville, Md. 20852
Telephone 301 - 881-1623

August 30, 1967

ARIES Corporation
Westgate Research Park
McLean, Virginia 22101

Gentlemen:

Progress Report
Subcontract 212056 -- NASA Contract NASW-1629

Summary

This report, and the materials delivered herewith or previously, constitutes completion of our duties under Tasks 1 and 2 of the subject subcontract. We have reviewed the Thesaurus printout and Term Review Forms (TRF's) and have marked corrections in the Thesaurus printout, in accordance with ARIES' requirements for keypunching input to effect the corrections (Task 1). Because NASA was unable to schedule the seminar for the research centers, we were asked to provide at least a part of the input which had been expected as a result of the seminar (Task 2) -- to validate all BT crossreferences, to re-examine all "multiple use references", and to provide the bases for a procedure for converting existing machine-sensible indexes from the original vocabulary to the "new" one. Corrections to the BT crossreferences and to the "multiple use references", when necessary, were also marked in the Thesaurus printout. The bases for the index conversion procedure are recommended in detail in the "Discussion" section of this report. Briefly, a largely computerized procedure is recommended, with input consisting of existing computer records.

We return herewith the following materials:

1. Marked (i.e., corrected) Main Term and "Use Reference" List.

2. List of terms explicitly deleted during the creation of TRF's.

3. Marked (i.e., corrected) NASA Thesaurus printout -- "Phase 7".

4. List of changes made to the "Phase 7" printout to create the Review Copy.

We have previously returned to you the following:

1. Loan copy of the LEX-EJC thesaurus printout.

2. Copies of TRF's supplied by NASA.

We have previously returned to NASA the no-longer-needed copy of an earlier (Phase 3) NASA thesaurus printout. This leaves in our hands worn-out and mutilated copies of a Subject Authority List and a NASA Thesaurus Review Copy, which we intend to retain until after the completion of Task 3 -- the planning and conducting of the indexer training program, scheduled for October 1967.

We estimate that we have made about 15,000 corrections to the printout -- about 5900 crossreference deletions, about 5200 crossreference additions, about 3200 crossreference changes, about 500 changes in term word-form, and about 200 term deletions; many of the latter were duplicates, etc. After these corrections (as noted in the Thesaurus printout) are made to the computerized Thesaurus file, together with any other corrections desired by NASA, it will be necessary to repeat the entire set of computer runs because of the massive nature of the revision. The first runs should detect non-existent main terms to which crossreferences have been made, and inconsistencies and conflicts in cross-referencing. These corrections should be made and verified (proofread) and if necessary re-corrected before the hierarchical expansion run is attempted. Hierarchical expansion must begin from its first phase, as it has been impossible to enter all BT's of BT's in the correction and review procedure to date. Again, the printouts and corrections must be carefully reviewed after each phase of hierarchical expansion.

The work performed under Tasks 1 and 2 of this subcontract is described in detail in the "Discussion" section of this report.

Respectfully submitted,

LEX-ING

Eugene Wall
President

EW:cpw

APPENDIX B

THESAURUS RULES AND CONVENTIONS

Facsimile from
ONR-26, Appendix 7
(AD 661 001)

Alphabetization

A-1

Alphabetize descriptors letter-by-letter, according to the following rules:

(1) Ignore all spaces between words.

(2) Ignore all characters other than left parenthesis, numerals, and letters.

(3) File according to the sequence:

(a) left parenthesis

(b) numerals in usual order

(c) letters in usual order

A representative sequence of terms filed according to the above rules is:

mercury (metal)
mercury (planet)
mercury amalgams
mercury arc rectifiers
mercury lamps
metal finishing
metallurgy
metals
metal working.

NOTE: *Metal working* is sometimes spelled as one word, *metalworking*. In letter-by-letter alphabetization, the sequential position of *metalworking* is unaffected by the spelling selected for the authorized descriptor. Similarly, the sequential position of *metal-working* is unaffected by the use of a hyphen.

(Verso Blank)

the hyphen) to indicate that it is not the most specific concept of its class. The placement of such terms in the thesaurus should be such that the term is left justified according to its first alphabetical character (so that the alphabetical portion of the entry lines up with the terms above and below it in the display). (See also Rules C-5, C-6, and C-8.) For example in the thesaurus display:

impedance

NT	acoustic impedance
	capacitive reactance
	characteristic impedance
	– electrical impedance
	– electrical resistance
	inductive reactance
	input impedance
	– reactance
	transfer impedance
RT	bandwidth
	conductivity
	– damping
	dead time
	– dynamic response
	hold-up
	– resistivity
	transient response,

the hyphen designator shows that, for the terms so designated, there are more specific terms available in the vocabulary.

C-5
Narrower Terms

The NARROWER TERM reference (NT) is the reciprocal of the BROADER TERM reference (See Rule C-4) and is employed to refer from a term symbolizing a concept class to all terms symbolizing concepts that are members of that class; for example, *iron alloys* NT *gray iron, mottled iron, steels.* For each NARROWER TERM reference, there must be provided a corresponding BROADER TERM reference.

C-6
Hierarchy

BROADER TERM references and NARROWER TERM references are hierarchical references. If there exist more than two levels in such hierarchies, the cross references for all levels should be completed for each term. This is done to enable the thesaurus user to ascertain the appropriate level of specificity in a family of generically related concepts and to promote editorial consistency during thesaurus revision, or in cases where portions of the thesaurus are extracted as specialized indexing vocabularies.

In a few instances, terms will be so broad in meaning that their utility as indexing terms will be doubtful, yet they must be retained for use in disciplines of peripheral interest or merely as a guide to more specific terminology. Under these circumstances, list under the very broad descriptor the terms to which it is desirable to refer the thesaurus user omitting cross reference designators. Append to the broad descriptor the scope note *Use of a more specific term is recommended — consult the terms listed below.* For example, the term *materials* is of little use in indexing documents that deal with materials in any but the most general way, but in a vocabulary in which many specific materials types are represented by indexing terms, the term is a useful point at which to display certain more specific terms or related terms for further study without carrying a useless BROADER TERM reference to *materials* on each of many specific terms. No reciprocal cross references should be made from listed specific descriptors to the broad descriptor, although such records should be maintained for housekeeping purposes. (See Rule C-9.)

C-7
Hierarchy Overlap

Terms may be members of more than one hierarchy. (See Rules C-4 and C-5.) For example, consider the entries *avalanche diodes* BT *diodes* and *semiconductor devices; diodes* NT *avalanche diodes.* The term *avalanche diodes* represents a concept that is properly a member of the two different classes of concepts represented by the terms *diodes* and *semiconductor devices.*

C-8
Related Terms

The RELATED TERM references (RT) is employed as a guide from a given descriptor to other descriptors that are closely related conceptually, but not hierarchically. (See Rules C-4 to C-7.) In general, any two descriptors are cross referenced RT if it is believed that the user, when examining one descriptor, might want to be reminded of the existence of the other.

RELATED TERM references may be used to identify:

a. descriptors that are closely related in meaning or concept including those in different hierarchical structures.

b. descriptors that are near synonyms.

c. descriptors that have viewpoint interrelationships, such as a relationship based on usage; for example, *alcohols* RT *solvents,* RT *antifreezes.*

d. descriptors representing concepts bearing a part-whole relationship to each other.

C-9
Reciprocal Entries

Reciprocal entries are required in every instance of cross-referencing except as noted in Rule C-6. The three pairs of reciprocal cross references are:

USE	USED FOR
BROADER TERM	NARROWER TERM
RELATED TERM	RELATED TERM

Cross References

Appendix 7

C-1
Cross References

Relationships among terms will be shown by cross references, which will aid users in selecting descriptors from the thesaurus. Types of cross references and their symbols that will be used as required are:

Cross References	Symbols
USE	USE
USED FOR	UF
BROADER TERM	BT
NARROWER TERM	NT
RELATED TERM	RT

These are described in the following rules.

C-2
Use

The USE reference is intended to lead users of the thesaurus to appropriate descriptors and should be employed to refer from a term that is not an authorized descriptor to the term which has been chosen as the descriptor, as follows:

a. to indicate a preferred synonym; for example, *lysozyme* USE *muramidase; secondary batteries* USE *storage batteries;*

b. to refer from a specific term to a more general term which has been selected to represent (that is, subsume) the specific concept; for example, *plant waxes* USE *waxes; sand blasting* USE *abrasive blasting;*

c. to indicate a preference between spelling variations, or to expand or explain abbreviations; for example, *pi mesons* USE *pions; inflammability* USE *flammability; pentaerythritol tetranitrate* USE *PETN; IFF* USE *identification systems;*

d. to prescribe the use of two or more descriptors to express a concept; for example, *ferromagnetic films* USE *ferromagnetic materials* and *films; antitank rockets* USE *antitank ammunition* and *rockets; optical illusions* USE *illusions* and *vision;*

e. To express concepts that can be considered synonyms for purposes of indexing and retrieval; for example, *heredity* USE *genetics; semantemes* USE *semantics;*

f. to bring together different viewpoints of a conceptual continuum; for example, *fluidity* USE *viscosity; smoothness* USE *roughness; instability* USE *stability.*

g. to explain variations in word order; for example, *tables (mathematics)* USE *mathematical tables; propellers (aerial)* USE *aerial propellers; propellers (marine)* USE *marine propellers;*

h. to reflect current terminology; for example, *electrical condensers* USE *capacitors;*

i. to eliminate jargon; for example, *whirly bird* USE *helicopters.*

C-3
Used For

The USED FOR reference (UF) is the mandatory reciprocal of the USE reference and accompanies the descriptor to which the USE reference refers. Accordingly, the USE references in two of the examples given in Rule C-2b and C-2c would generate the following USED FOR references:

abrasive blasting

UF sand blasting

pions

UF pi mesons

When a USE reference has prescribed two or more descriptors to represent a concept (See Rule C-2d), a unique symbol such as the dagger sign (↑) should be placed in front of the unauthorized term in the USED FOR reference; for example, *ferromagnetic films* USE *ferromagnetic materials* and *films;*

ferromagnetic materials

UF ↑ferromagnetic films

films

UF ↑ferromagnetic films

C-4
Broader Terms

The BROADER TERM reference (BT) is employed to refer from a term representing a member of a class (or classes) of concepts to any term(s) in the thesaurus representing that class or classes; for example, *steels* BT *iron alloys.* For each BROADER TERM reference there must also be provided a corresponding NARROWER TERM reference. (See Rule C-5.) The part-whole relationship is usually not a broader-narrower relationship; for example, *gear teeth* BT *gears* is incorrect. However, in certain specific areas, part-whole generics can be usefully employed. These areas are anatomy and geographic locations. Also specifically excluded from the broader-narrower term category are relationships based on the possible applications or uses of an entity; for example, *platinum* is not considered to be a member of the generic family *catalysts* because, although it is sometimes used as a catalyst, it has too many other applications to list all as broader terms. *Platinum* is, however, always a member of the class metals, so that the reference *platinum* BT *metals* should be entered.

Wherever a descriptor for which there are narrower terms in the vocabulary appears in the thesaurus display as either a narrower term or a related term, it should be preceded by an unambiguous symbol (such as

ations. Abbreviated terms can be used when meanings are well established and when significant gains in convenience can be demonstrated. Examples are: *ACTH* for *adrenocorticotropic hormone; PETN* for *pentaerythritol tetranitrate; VTOL aircraft* for *vertical take off and landing aircraft.* Abbreviated and unabbreviated forms of a given term should be treated as synonyms and cross referenced accordingly. (See Rule C-2c.) Well established acronyms are acceptable as descriptors; for example, *shoran, radar, maser.*

T-10
Specialized Vocabularies

Effective indexing and retrieval of information in certain specialized subject fields will require descriptor vocabularies that differ in some ways from the natural language approach of the thesaurus as a whole as in the following examples.

a. Chemistry. To avoid proliferation of terms in the field of chemistry, the names of specific chemical compounds as descriptors should be restricted. Instead, a vocabulary of descriptors representing generic compounds classes, functional groups, and structural features should be devised. This will permit indexing and searching by coordinating appropriate descriptors to denote specific compounds as well as classes of compounds.

Names of specific compounds that occur frequently in the literature may be entered as descriptors; for example, *sulfuric acid, carbon tetrachloride, morphine, progesterone.*

b. Alloys. Descriptors should be established for certain generic alloy families; for example, *aluminum copper alloys, molybdenum steels, zinc alloys.* This will permit indexing and retrieval on a somewhat general level, but will prevent proliferation of descriptors to represent specific alloy systems.

c. Biological nomenclature. Where possible, consistent use should be made of established nomenclature systems for describing plants and animals. Where departures are necessary, cross references should be provided to maintain continuity.

T-11
Multiword Descriptors

A descriptor represents a concept. In general, a descriptor should reflect the terminology found in the literature regardless of the number of words needed to express the concept. However, many candidate descriptors will represent, or appear to represent, concepts that are combinations of two or more specific descriptors or potential descriptors. In these instances, a determination must be made of the most effective way in which to incorporate the candidate descriptor, that is, by adding the specific multiword term to the thesaurus or by prescribing a combination of two or more existing descriptors.

Specific multiword descriptors often facilitate retrieval of specific information, but may add to the cost of indexing by increasing substantially the number of descriptors in the index. On the other hand, proper use of individual descriptors in indexing for later combination upon retrieval will serve to control the size of the vocabulary and to promote consistency in the use of terminology. Decisions on the formation of specific descriptors require consideration of the following factors. If reasonable doubt remains, a specific multiword descriptor should be established because, if it eventually proves to be of little utility, it can easily be reduced to a combination of individual descriptors. The converse is not easily accomplished.

a. A specific multiword descriptor should be established when suitable more general descriptors are not available in the vocabulary. To provide adequate representation of the specific concept, a combination of general descriptors must include at least one descriptor that is a member of the same hierarchical class as the specific concept. Observation of this principle will promote more consistent and complete retrieval, whereas the use of a combination in which neither descriptor bears a generic relationship to the specific concept may lessen retrieval efficiency.

b. A specific, multiword descriptor should be established when the specific concept is encountered so frequently that the ability to index and search directly would be both expeditious and economical or when one or both of the more general descriptors is so often used in indexing as to make searches awkward or inaccurate.

c. Two or more individual descriptors should be used instead of establishing a specific multiword descriptor when the specific concept is a member of the same generic class (See Rule C-4) as each of the more general descriptors.

TABLE 1. GUIDELINES TO SINGULAR-PLURAL USAGE

TYPE OF TERM	*USE SINGULAR FORM*	*USE PLURAL FORM*
Material terms, such as: chemical compounds mixtures materials	When term is specific, as: urea cellophane beeswax	When term is generic, as: amines solvents plastics
Terms representing properties, conditions, characteristics	When term is specific, as: viscosity temperature purity opacity	When term is generic, as: physical properties process conditions
Terms representing equipment, devices, physical objects, and elementary particles	Do not use singular	Use plural, as: pulverizers regulators mesons teeth stars
Class of use terms	Do not use singular	Use plural, as: adhesives catalysts
Process terms	Use singular, as: constructing installing modulating	Do not use plural
Proper names (A proper name is defined as the name for a *single unique* item)	Use singular, as: Hookes Law Pluto	Do not use plural
Disciplines, fields, subject areas	Use singular according to common usage, as: chemistry hydraulics engineering	Do not use plural (Words such as "hydraulics" are actually singular)
Events or occurrences	Do not use singular	Use plural, as: ambushes explosions discharges

preferred term should be entered as a descriptor, the other as a USE reference. (See Rule C-2f.)

Terms representing concepts that overlap significantly may be treated as quasi-synonyms; for example, *lighting* and *illumination; duration* and *time; genetics* and *heredity*. The preferred term should be indicated by a USE reference. (See Rule C-2e.)

T-8 Punctuation

Punctuation marks in descriptors should be restricted. Highly specific systematic names, such as some chemical names, that require elaborate punctuation should be treated specially when they fall within the scope of the thesaurus. (See Rule T-10.) Parentheses should be used to enclose qualifying expressions which are included in descriptors to prevent ambiguity. (See Rule T-5.) Commas, periods, apostrophes and most hyphens should be excluded since they are difficult to handle consistently, complicate machine processing of the thesaurus, and are not necessary to convey the meaning of the terms. Hyphens should be used only in terms whose intended meaning would be altered by omission of the hyphen. In omitting a normally occurring hyphen, the space occupied by the hyphen should be handled according to these criteria; (1) retain the space for compound adjectives, noun-noun combinations, and letter-word combinations; (2) drop the space in attaching prefixes to the base words. Examples are:

(1) high temperature testing
man machine systems
n body problem

(2) countermeasures
microanalysis
ultrahigh frequency

T-9 Abbreviated Forms

In general, abbreviated terms should be avoided as descriptors because their understanding may not be universal, their meaning may be dependent on context, or their recognition may be dependent on capitalization and periods, which are constraints in computer oper-

b. As construction of the thesaurus progresses and descriptors are selected, an ad hoc vocabulary framework will emerge. This structure will help form a basis for the selection of additional descriptors. Candidate descriptors should be examined to determine that they reflect a level of specificity commensurate with that of the existing structure and that they represent discreet concepts. Avoid the selection of descriptors whose meanings coincide so closely with those of established descriptors that indexers (and searchers) will have difficulty in distinguishing between or among them.

c. The acceptability of terms can be determined by consulting dictionaries, encyclopedias, other indexing vocabularies, and the opinions of subject specialists. Slang, jargon, coined terms and deprecated terminology should be excluded.

d. It is usually desirable to establish a maximum term length for the purpose of maintaining succinctness of terms or for other special purposes such as maintaining the capability for a particular page format in a printed thesaurus. A maximum of 36 characters per term is recommended.

T-2
Noun Form

Use noun forms wherever possible; for example, *roughness* rather than *rough*. In a limited number of instances, needed retrieval concepts can be represented only by adjectives or equivalent expressions. These usually take the form of words or phrases that describe in some manner the operation of equipment or systems; for example, *airborne, mobile, portable*.

Never use verbs; for example, use *catalysis* rather than *catalyze*, the gerund *pouring* rather than the verb *pour*.

T-3
Singular vs. Plural

In choosing between singular and plural noun forms, a useful rule of thumb may be applied as follows: use the plural form when the proposed term is a count noun, that is, a noun which may be used to answer the question "how many?" (for example, devices such as *gages, nozzles, fuzes);* use the singular form for mass nouns, those that express "how much?" (for example, *iron, wood, charcoal)*; use the singular for specific processes, properties, or conditions. Table 1 provides a summary of the recommended procedure. Common usage should be followed for term types not covered in the above general rule or in the table.

T-4
Direct Entry

Descriptors consisting of two or more words should be listed in their natural word order, that is, the order normally used in English sentences; for example, *radar antennas* rather than *antennas, radar; refractory materials* rather than *materials, refractory*.

T-5
Descriptor Definition

Terms that have more than one accepted meaning, that are intended to be used in a somewhat different way than ordinarily defined, or for which distinctions from other descriptors must be drawn, should be accompanied by an explanation. The meanings of terms can be clarified or made more specific in the following ways:

a. Modifying terms can be used to preface a given term, as in *metal tubing* to make the meaning of the term *tubing* more specific. This is the method underlying the construction of direct entries (See Rule T-4) and is subject to the limitations of the rule on multiword descriptors (See Rule T-11).

b. A parenthetical qualifying expression can be appended to a homograph to clarify its meaning; for example, *mercury (metal)* and *mercury (planet)*. Such a qualifying expression becomes a part of the descriptor.

c. Employ the "–ing" suffix for processes and the "–ion" suffix or other appropriate suffixes for materials, characteristics, etc., when necessary to distinguish clearly among them. Examples are: *concentration* and *concentrating; precipitation* and *precipitating*.

d. When a qualifying expression cannot adequately convey the intended meaning, a short explanation called a scope note should accompany the term. Precise dictionary definitions usually should not be attempted. The scope note merely indicates the way in which the descriptor should be used. It is not a part of the descriptor, but follows on a succeeding line, as:

water cooling
cooling with water

e. Since trademarks are recognized by law as being proprietary, they should be identified in the thesaurus by the qualifying expression "(trademark)." When the meaning of a trademark is not self-evident, a scope note may also be added to clarify its meaning.

T-6
Synonyms

When two candidate descriptors are true synonyms, one should be selected as the descriptor, the other entered as a USE reference. (See Rule C-2a.)

T-7
Quasi-synonyms

To prevent scattering of like information in indexing and to obviate multiple searches for effective retrieval of information, it is both practical and desirable to consider terms having certain special relationships to be synonyms for indexing and retrieval purposes, that is, quasi-synonyms.

Terms that represent different viewpoints of the same property continuum may be considered quasi-synonyms; for example, *smoothness* and *roughness*. The

Thesaurus Rules and Conventions

Introduction

Building a thesaurus for vocabulary control in information storage and retrieval systems requires that certain decisions be made and recorded so that the thesaurus structure is internally consistent and reasonable. Moreover, users of a thesaurus need to know the rules and conventions upon which it is based for efficient and intelligent use.

These rules and conventions were used as the basic reference in the joint EJC-Project LEX development of an identical thesaurus for both Engineers Joint Council and the Department of Defense. The principles set forth should be generally applicable to thesaurus development and use. Uniform application of widely accepted thesaurus rules and conventions in information storage and retrieval systems may be expected to facilitate compatibility of systems for more efficient operation.

The thesaurus rules and conventions were prepared by a Task Group of the EJC Engineering Vocabulary Panel, Robert Dodds of Gibbs and Hill, chairman. Members of the Task Group were:

David M. Liston — Battelle Memorial Institute, • chairman

Margaret Hicks — CFSTI, assigned to Project LEX

Jack W. Hilf — Bureau of Reclamation, Dept. of the Interior

Terry Gillum — DDC, assigned to Project LEX

Eugene Wall — Consultant to Project LEX

Indexing Terms—Descriptors[1]

The description of technical documents (indexing) for information storage and retrieval requires the use of two types of terms: (a) those that describe the information and data contained in the document, and (b) bibliographic terms—terms that describe the document itself, not the information in the document. Bibliographic terms, examples of which are personal authors, corporate authors, and publication dates, should not be included in the thesaurus. Terms that describe the information contained in the documents may include (1) project names, (2) military nomenclature, (3) identification symbols or numbers, (4) nicknames or jargon, (5) geopolitical names, (6) trademarks, (7) other proper names, (8) terms of an analytical nature, such as *anemometers, boundary layer, cardiovascular system, density, energy conversion, heat resistant alloys, spectroscopy;* and (9) terms of an abstract nature, such as *tests, measurement,* and *calibration.* The index terms of primary concern to the thesaurus are types 8 and 9. For convenient reference, terms selected to be entered into the thesaurus for use as indexing terms will be called "descriptors."

The rules and conventions for constructing the thesaurus are of three types: (1) those dealing with fundamental terms (T-1 to T-12); (2) those dealing with cross references (C-1 to C-8); and (3) the alphabetization rule (A-1).

[1] Examples of descriptors included in these rules and conventions are chosen for illustration only, and do not necessarily appear in the thesaurus.

T-1
Descriptor Selection

Descriptors should be selected for inclusion in the thesaurus on the basis of their estimated usefulness in communication, indexing, and retrieval. In general, utility of terms can be estimated by considering (a) the relative frequency of occurrence in the literature, (b) the relative frequency of use within an operating system, (c) relationships to descriptors that have been selected previously, and (d) scientific or technical precision and acceptability. These factors are very much interdependent and should be considered together in the selection of descriptors.

a. Frequency of prior use of a term in indexing and searching within a particular vocabulary gives a rough quantitative indication of its usefulness. The importance of frequency of use of a term depends upon the usage of other terms, the relative age of the term, and the age and scope of the collection. Terms that have been used relatively often within a given vocabulary may represent concepts that are poorly defined or too general to be useful in describing subject matter, whereas those that have been used very infrequently may represent concepts that are obscure or overly specific. Low frequency of use should not necessarily cause the rejection of a term that represents a novel concept and is a recent addition to the original vocabulary. The general rule is to establish descriptors that convey specifically the subject matter indexed. The utility of the terms can then be evaluated by reviewing their frequency of use in indexing.

LEX DATA STORE

RECORD IDENTIFICATION CODE KEY

- Main-term record (MT) is identified by a 1 in position 8.
- Scope Note (SN) is identified by a 2 in position 8 and a numeric line sequence code (1 through 9) in position 15. Frequency of use is also included as a scope note line thus "FRQ 135".
- Subject Category Code. Subject category is represented by a set of 4 numeric digits: numerics in positions 16 and 17 represent COSATI Fields and those in positions 18 and 19 represent COSATI groups (included where information became available).
- A "USE" cross-reference is identified by a 3 in position 8.
- A "Use for" (UF) cross-reference (reciprocal of USE) is identified by a 4 in position 8.
- A "Broader term" (BT) cross-reference is identified by a 5 in position 8.
- A "Narrower term" (NT) cross-reference is identified by a 6 in position 8.
- A "Related term" (RT) cross-reference is identified by a 7 in position 8.
- Some contributors also submitted a non-reciprocal related term. This is identified in the file by an 8 in position 8. In ARIES Programs this will print out as "RRT" unless suppressed.