DOCUMENT RESUME

ED 260 735 IR 051 241

AUTHOR	Silvester, June P.; And Others
TITLE	An Operational System for Subject Switching between Controlled Vocabularies: A Computational Linguistics Approach.
INSTITUTION	Planning Research Corp., McLean, Va.
SPONS AGENCY	National Aeronautics and Space Administration, Washington, DC. Scientific and Technical Information Branch.
REPORT NO	NASA-CR-3838
PUB DATE	Oct 84
CONTRACT	NASW-3330
NOTE	105p.
PUB TYPE	Reference Materials - Vocabularies/Classifications/Dictionaries (134) -- Reports - Descriptive (141)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	*Computational Linguistics; Information Science; *Information Systems; *Language Processing; Linguistic Theory; Program Descriptions; Programing Languages; *Scientific and Technical Information; Structural Analysis (Linguistics); *Subject Index Terms; *Vocabulary Development

ABSTRACT

This report describes a new automated process that pioneers full-scale operational use of subject switching by the NASA (National Aeronautics and Space Administration) Scientific and Technical Information (STI) Facility. The subject switching process routinely translates machine-readable subject terms from one controlled vocabulary into the equivalent terms of another controlled vocabulary using a system called the NASA Lexical Dictionary (NLD). This report also describes the NLD, discusses how to build a lexical dictionary system, details the resources needed, and explains how to maintain the system after it is built. A description of the NASA STI Facility's experiences with their first input vocabulary, that of the Defense Technical Information Center (DTIC), is included. Following a preface and executive summary, this report is divided into seven sections: (1) introduction (purpose, significance, definition of the NASA Lexical Dictionary, scope of NLD, preliminary results, presentation, and project personnel); (2) system description; (3) history; (4) procedures for building a lexical dictionary; (5) data file maintenance; (6) results and conclusions; and (7) summary. A glossary, two appendices, and references are included. (THC)

NASA Contractor Report 3838

An Operational System for Subject Switching Between Controlled Vocabularies: A Computational Linguistics Approach

June P. Silvester, Roxanne Newton,
and Paul H. Klingbiel
Planning Research Corporation
McLean, Virginia

Prepared for
National Aeronautics and Space Administration
under Contract NASW-3330

National Aeronautics
and Space Administration

Scientific and Technical
Information Branch

1984

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF ILLUSTRATIONS

PREFACE

This report describes a new automated process that pioneers full-scale operational use of subject switching by the NASA Scientific and Technical Information (STI) Facility. The subject switching process routinely translates machine-readable subject terms from one controlled vocabulary into the equivalent terms of another controlled vocabulary. To do subject switching, we use a system called the NASA Lexical Dictionary (NLD). The report also describes the NLD, how to build a lexical dictionary system, what resources are needed, and how to maintain the system after it's built. The experience of the NASA STI Facility with their first input vocabulary, that of the Defense Technical Information Center (DTIC), is included in the section labeled HISTORY.

We would like to acknowledge the help given to this project by personnel at DTIC. Without their cooperation the construction of the NLD would have been more difficult and costly.

Work on the NLD, done by Planning Research Corporation/Government Information Systems, was supported by the National Aeronautics and Space Administration's Scientific and Technical Information Branch under contract NASw-3330. The period of performance covered by this report is from November 2, 1981 to December 31, 1983.

EXECUTIVE SUMMARY

The NASA Lexical Dictionary (NLD), a system that automatically translates input subject terms to those of NASA, was developed in four phases. Phase One provided Phrase Matching, a context sensitive word-matching process that matches input phrase words with any NASA Thesaurus posting (i.e. index) term or Use reference. Other Use references have been added to enable the matching of synonyms, variant spellings, and some words with the same root. Phase Two provided the capability of translating any individual DTIC term to one or more NASA terms having the same meaning. Phase Three provided NASA terms having equivalent concepts for two or more DTIC terms, i.e. coordinations of DTIC terms. Phase Four was concerned with indexer feedback and maintenance. Although the original NLD construction involved much manual data entry, ways were found to automate nearly all but the intellectual decision-making processes. In addition to finding improved ways to construct a lexical dictionary, new applications for the NLD have been found and are being developed.

INTRODUCTION

Purpose

The purpose of the NLD is to minimize the indexing of documents already indexed by another agency. Approximately half of the report literature added to the NASA STI Facility data bases each year has been previously cataloged, abstracted, and indexed by another agency. See Figure 1. Much of this previously processed material is received at the NASA STI Facility (hereafter referred to as the Facility) in machine-readable form on magnetic tape. The Facility's objective is to accept as much as possible of this work and the NLD is part of the overall effort. The NLD accepts, in machine-readable form, words and phrases from the document record created by another agency and translates them into valid NASA index terms. The words and phrases that are run through the NLD are normally terms from a controlled vocabulary such as DTIC's. However, it is possible to take a title or a line or two of text and treat it as if it were a long phrase. While the Access Routine was not designed to select phrases from text, it can be used to generate posting terms from a limited amount of text such as a title, title supplement, note of content, or from words and phrases from any machine-readable source. The terms generated in this manner must be reviewed and may need to be edited by the indexer.

Significance

The NLD Subject Switching system is a flexible tool. It could be implemented as a time-saving device by any organization that accessions and reindexes documents that have been indexed by another organization. The components of the NLD provide the basis on which to build either a system for automatically indexing text, or a system for the automatic translation of index terms from any controlled vocabulary to another.

Definition of the NASA Lexical Dictionary

A lexical dictionary has been defined in several ways. Paul H. Klingbiel, who initiated the NLD, defines it two ways in his latest work (ref. 1): as "a phrase structure rewrite system" and as "a matrix." Roxanne Newton defined the NLD (in the "System Overview" written for Facility use) as "a translation device." These different descriptions represent different points of view. To a mathematician, a lexical dictionary is a matrix; to a linguist, it's a grammar; to an accountant, the system may resemble a spreadsheet; but to those dealing with operating systems, the lexical dictionary is a translation device. June Silvester adds that, to the indexer, the lexical dictionary is a tool.

This report addresses itself primarily to the operating system definition--that is, that the lexical dictionary is a translation device, and secondarily to the idea that the lexical dictionary is an indexer tool.

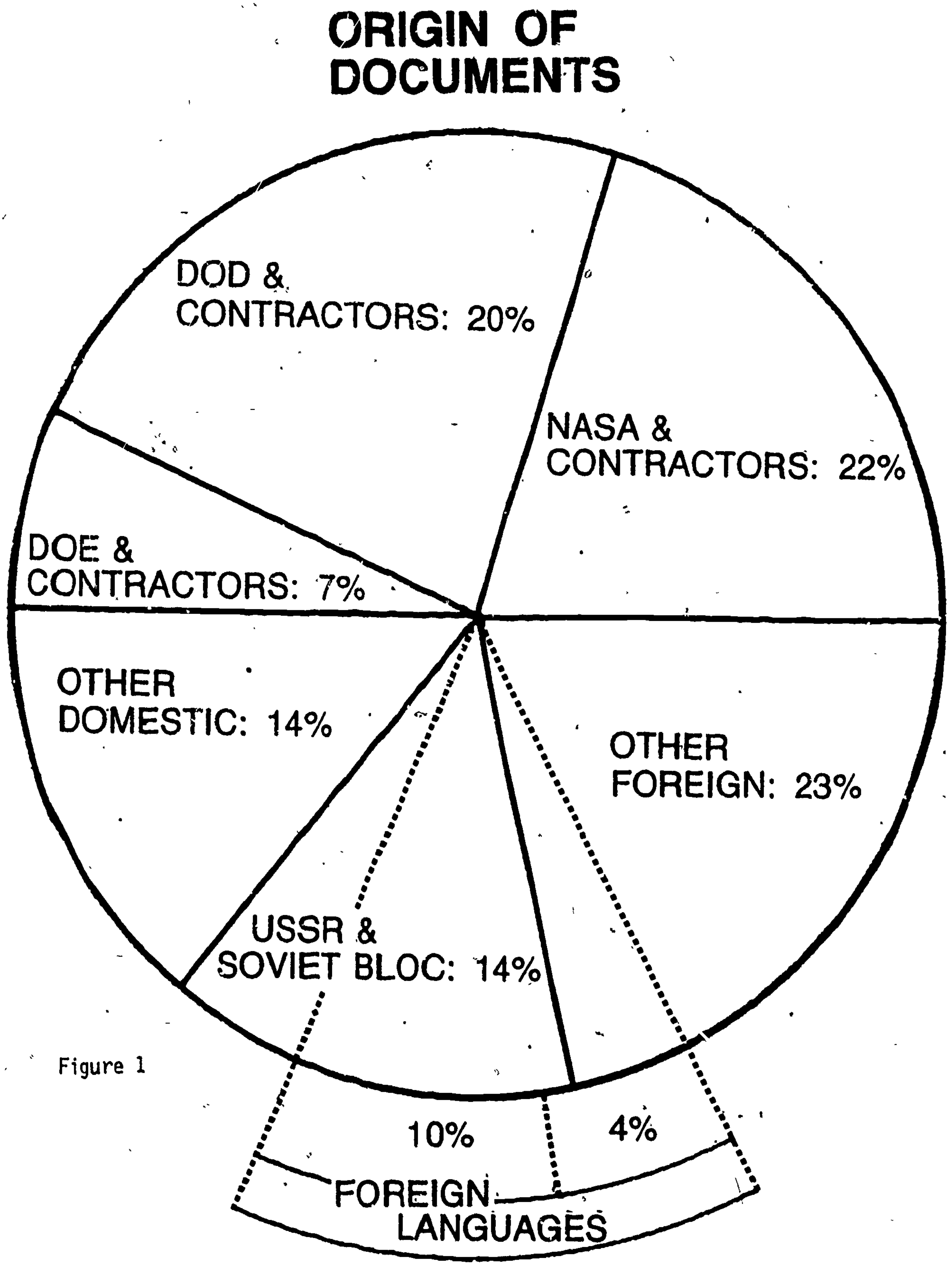

ORIGIN OF DOCUMENTS
DOD & CONTRACTORS: 20%
NASA & CONTRACTORS: 22%
DOE & CONTRACTORS: 7%
OTHER DOMESTIC: 14%
OTHER FOREIGN: 23%
USSR & SOVIET BLOC: 14%
10%
4%
FOREIGN LANGUAGES

Figure 1

Scope of the NLD

The original NLD project, which aimed to translate DTIC indexing to NASA indexing, involved two procedures. First, the translation of the concept of every individual term in DTIC's controlled vocabulary to one or more NASA terms that express the same concept. Second, the translation of two or more coordinated DTIC terms to the NASA term or terms expressing the same concept. If the concept of coordinated terms required more than one NASA term for its translation, the NASA terms must be different from the DTIC terms. For example:

DTIC	NASA
administrative personnel	management, personnel
aeroelasticity	aeroelasticity
aircraft;drones	drone aircraft
army planning;army research	project planning, armed forces (United States)

The DTIC terms and their NASA equivalents are given to the indexers for review. Indexer review consists of four functions:

- Accept or reject NASA terms listed by the NLD.
- Add any terms not listed by the NLD that are necessary for the NASA environment.
- Indicate which terms are major terms.
- Recommend NLD changes such as: ways of improving translations, deleting irrelevant translations, or adding terms that should be coordinated to translate to a single term.

Preliminary Results

Management expected that the use of the NLD would shorten indexing time, and it has. Based on a questionnaire that was used as an evaluation tool, the NLD saves at least three minutes per document indexed.

The project personnel expected that the use of the NLD would make indexing more of a decision-making process and less of a lookup job -- lookup the term, lookup the word form, lookup the spelling, etc. The NLD has done that, too.

The indexers had mixed expectations. Some feared that the NLD would eliminate their jobs. The NLD has not done that. Many terms are context sensitive and to maintain high quality indexing at the Facility, the decision was made to require indexer review. The NLD provides a team-approved* translation of any input word or phrase. The indexer provides a check on context sensitive selections, a choice of pertinent NLD-suggested terms, and any additional index terms needed to serve the NASA environment. We feel that this combination makes the NLD an expert system (see section on SYSTEM DESCRIPTION).

Indexers who understand best how the NLD operates and what its output means are the most enthusiastic. Indexers must be trained to achieve optimal use of the NLD, but the training required is minimal.

Presentation

This report details the resources required for implementing the NLD Subject Switching System and provides a step-by-step implementation plan. It includes a system overview, the NASA experience with DTIC-to-NASA Subject Switching, the three-phase implementation plan which we followed, and some recommendations for doing it more easily. It also describes system maintenance--what is involved and how to do it. Finally, we discuss the benefits, problems, and future of the NASA Lexical Dictionary.

Project Personnel

The NLD has had three project directors since its inception: Paul H. Klingbiel, Roxanne Newton, and former analyst June P. Silvester. Programming has been provided by Elaine Sellman, succeeded by Duchesne "Duke" Clark, and Patricia Carroll. They were assisted by Midori Keech and Nina Kit. Posting term translations were done by senior retrieval analyst Edna Fleek, lexicographer Ron Buchan, abstracting/indexing supervisor Jacqueline Streeks, Klingbiel, Newton, and Silvester. The project was also supported by the publications and data entry staff.

*The NLD team for this task consisted of the project director, analysts, lexicographer, and the abstracting/indexing supervisor. Translations done by any team member were reviewed by other team members.

SYSTEM DESCRIPTION

Significance Within the Larger System

As we have stated, the lexical dictionary is a tool. Although it is used to automate the translation of one agency's vocabulary to that of another, the lexical dictionary does more. Its use can alter indexing procedures by relieving indexers of purely mechanical tasks. The maintenance of the lexical dictionary stimulates increased communication and cooperation between agencies involved. Lexical dictionary construction brings a new awareness of the shortcomings and strengths of various thesauri and the need to improve terminological standards within the government. Thinking and talking about ways to communicate better is better communication--or at least some communication where often little or none existed before.

Expert System Concept

General Description. The lexical dictionary might be classed as an expert system, although it is a somewhat rudimentary one. By an expert system we mean a system that can emulate human reasoning. William B. Gevarter describes the components of an expert system as follows:

> (1) a knowledge base (or knowledge source) of domain facts and heuristics associated with the problem;
>
> (2) an inference procedure (or control structure) for utilizing the knowledge base in the solution of the problem;
>
> (3) a working memory--"global data base"--for keeping track of the problem status, the input data for the particular problem, and the relevant history of what has thus far been done.
>
> (ref. 2, p. 80).

The NLD has data files created by NASA vocabulary experts. The files include logic codes. These files are constantly added to, corrected, and improved by the experts who created the original knowledge base and by others who interface with the system. The logic codes in the files provide direction to the Access Routine (see subsection on System Components, Lexical Dictionary Access Routine). These files with their domain facts satisfy the requirement for the knowledge base and heuristics associated with the problem.

The Access Routine approximates an inference procedure by using the NLD files or knowledge base in the solution of the problem. The problem is defined as the determination of acceptable combinations of words from an input source to be translated into authorized NASA index terms.

The working memory for the NLD traces input material through the

translation process. The memory makes it possible to print lists of input terms with their translations, input terms with partial translations, input terms with no translations, and statistics about the logic codes used.
Thus the NLD meets the basic criteria for an expert system.

Another way in which the NLD system might be considered expert is its use of the best of two approaches: Indexer Simulation and Indexer Feedback.

Indexer Simulation. A lexical dictionary simulates the indexer's translation of input terms into target vocabulary terms. The printout provided by the lexical dictionary lists the terms from the input vocabulary and the terms from the target vocabulary, side by side, ready for indexer review and selection. This process uses the computer to do the repetitious, uninteresting, time-consuming indexer tasks that are largely mechanical, and provides expert, consistent translations for a final review by humans.

Indexer Feedback. The indexer reviews the suggested target terms and provides the lexical dictionary personnel with recommendations for improved translations. These recommendations are studied, usually approved, and the needed changes are made. These changes improve the indexer simulation for subsequent runs of the program, but the system uses humans to make decisions not yet possible with available software. Humans also upgrade the computer system which results in improved indexer simulation for subsequent runs of the program.

As in the old chicken and egg go-round, each produces the other. Both are essential, and they work in a kind of endless loop. Together the system has the best of both automatic and human input, and it keeps building on itself, hence an expert system.

System Functions

As stated before, the NASA Lexical Dictionary system is a translation device. The NLD translates words and phrases from machine-readable input material into corresponding NASA Thesaurus posting terms. The mode of operation, either Phrase Matching or Subject Switching, depends upon the type of input material being processed.

- The Phrase Matching mode is a general purpose matching routine which attempts to find context sensitive word-by-word matches between any input phrases and NASA posting terms or Use references. Matches may be complete or partial. In some cases, no match will be found. For example:

Input Phrase	NASA Posting Term(s)
Salaries	No match found
Fuel consumption	Fuel consumption
Inorganic acids	Acids
Cellulose acetates	Cellulose, Acetates
Chance-Vought military aircraft	Chance-Vought aircraft, Military aircraft

- The Subject Switching mode is a special purpose routine which translates the concepts expressed by the posting terms assigned to a document by a particular contributing source (such as DTIC) into the equivalent concept expressed in NASA posting terms. Subject Switching treats each input posting term as a unit, in contrast to Phrase Matching where the unit is the word. A unique translation table is built for the posting terms of each contributing source. An entry is created for every contributed posting term, but in some cases, the translation may indicate that the term is out of scope or not able to be translated. For example:

DTIC Posting Term(s)	NASA Posting Term(s)
Regiment level organization	NIS (Not In Scope for NASA)
Complementary metal oxide semiconductors	CMOS
Internal combustion engine noise	Engine noise, Internal combustion engines
Abrasion, Resistance	Abrasion resistance
Self treatment	00 (No NASA translation)

A more detailed explanation of these two translation modes is provided in the section on DATA FILE MAINTENANCE.

The Lexical Dictionary system can be used as the basis for an automatic indexing system to process text fields, such as abstracts. Automatic indexing, if it were instituted, would require the addition of a word recognition file to assign syntax codes and a program or programs to break text into logical words and phrases for processing. DTIC uses a similar system for automatic indexing of several data bases.

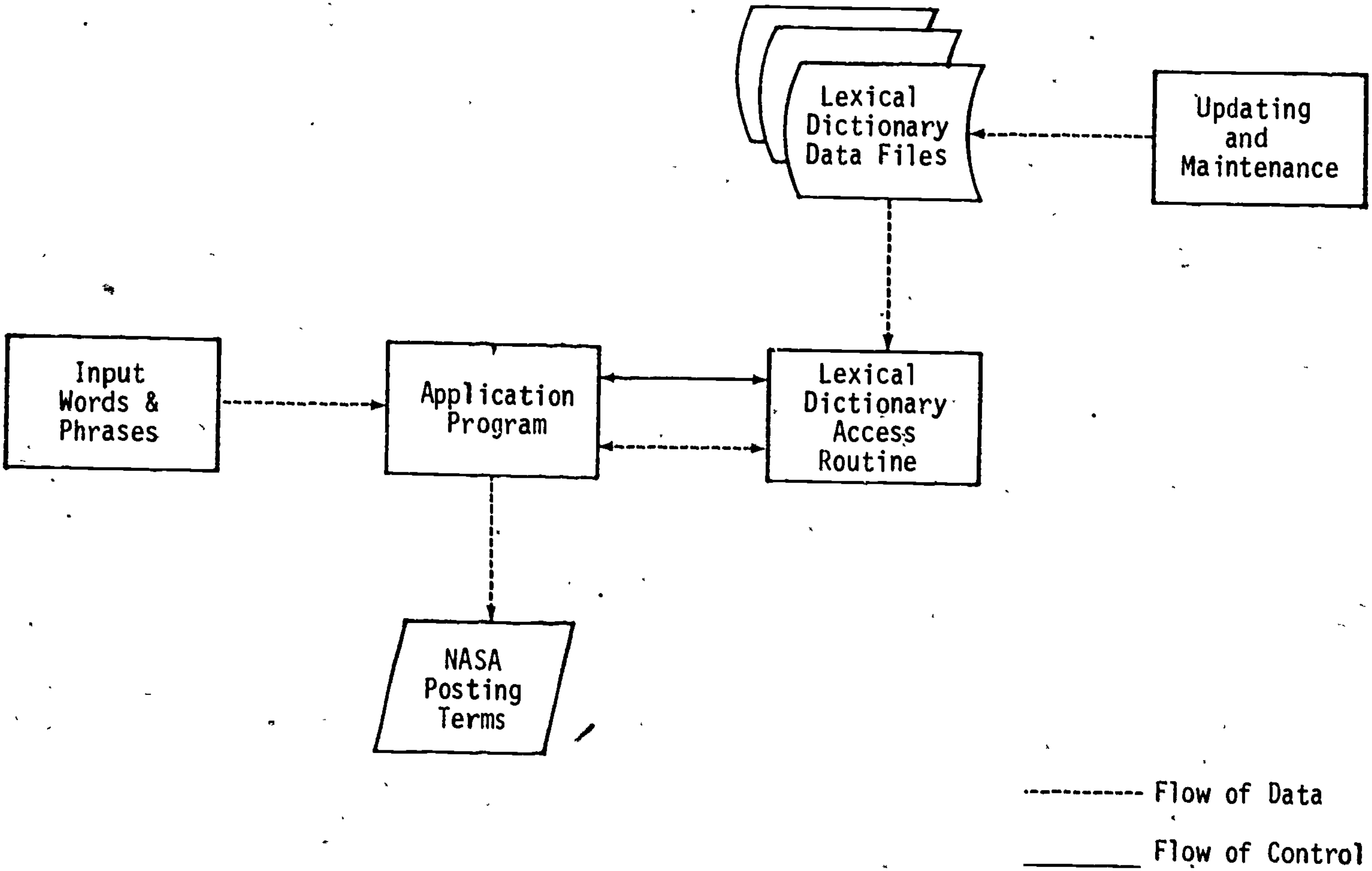
System Components

The NLD has three major components: data files which act as translation tables, an Access Routine which manipulates the input words and phrases, matches them against the data files, and returns the NASA translation to the application program, and applications programs which call the Access Routine.

Figure 2 gives an overview of the NLD system operation, and this section will describe briefly the three NLD components.

Figure 2

Overview of Lexical Dictionary System Operation



Data Files. The Lexical Dictionary system employs two types of Virtual Storage Access Method (VSAM) files:

- A general purpose Phrase Matching file and
- special purpose Subject Switching files for the controlled vocabulary (thesaurus terms) of each contributing source.

The file organization and record layout for both types of files are the same. Each NLD file record consists of the following fields:

- Key
  Each key is unique and consists of terms that may be encountered in the input material. The key can consist of a single element, followed by a semicolon and two zeros (;00), or of multiple elements separated by semicolons (;). In the Phrase Matching file, these elements are the individual words that make up the target vocabulary posting terms or Use references. In the Subject Switching file, each element is an entire posting term from the vocabulary of the contributing source for that file. Terms may be single or multiple words.

- Logic Code
  The Logic Code is a one character code that indicates how the key is to be processed. Single element keys are assigned one of the following logic codes:

  E - (Equal) The key translates to a single posting term that is identical to the key.

  C - (Change) The key translates to a single posting term that is different from the key.

  L - (List) The key translates to multiple posting terms that should be used in combination.

  I - (Indexer Choice) The translation of the key is context dependent. The meaning appropriate for the document at hand must be selected and a choice of posting terms is offered.

  O - No translation is available for the key.

  When there are multiple elements in the key, the logic code T (Table) is always used.

- Posting Term
  The posting term field contains the NASA posting term or terms to which the key is to be translated. The field may also contain the following special symbols, which serve as an aid to indexers:

  @ - NASA posting term is an array or ambiguous term.

> - NASA posting term is broader than the contributing source term in the key.

+ - NASA posting term has narrower terms that the indexer should consider. These are terms that the contributing source does not have.

? - Indexer should choose one or more of the NASA posting terms as appropriate.

00 - No appropriate NASA translation is available.

NIS - The contributing source term is NOT IN SCOPE for NASA.

In some cases, there are more than two elements in a key. NLD System processing requires that intermediate records be created which build to these multi-element keys. The first entry will consist of the first two elements. Each successive entry will add one more element until the entire phrase is complete. Since the intermediate keys do not have translations, the posting term fields for these records contain a special symbol as a place holder. For example:

Logic Code	Key	Posting Term
T	Body;Centered	*
T	Body;Centered;Cubic	**
T	Body;Centered;Cubic;Lattices	Body centered cubic lattices

Samples of records from the Phase Matching file and the Subject Switching file are shown below:

Phrase Matching File Sample

Logic Code	Key	Posting Term
E	Bleeding;00	Bleeding
C	Blends;00	Mixtures
E	Blight;00	Blight
T	Blind;Landing	Blind landing
T	Block;Band	Block band

(Logic codes I and O; and symbols > , ?, +, 00, and NIS are not normally used in the Phrase Matching file.)

Subject Switching File Sample

Logic Code	Key	Posting Term(s)
E	Filters;00	Filters
E	Financial management;00	Financial management
C	Fingernails;00	Fingers

Subject Switching File Sample (continued)

Logic Code	Key	Posting Term(s)
O	Fingerprint recognition;00	00
E	Fins;00	Fins+
L	Fire alarm systems;00	Fires,Warning systems
I	Fire protection;00	Fire prevention?, Fireproofing?
T	Floating bodies;Sea ice	Ice floes

Lexical Dictionary Access Routine. The NLD Access Routine is a general purpose program that accesses the Lexical Dictionary files. Its product is a list of index terms from the NASA Thesaurus which was the target vocabulary.

The Access Routine never operates independently; it is always called by an application program. The application program passes the Access Routine two things:

- a code that indicates whether the Phrase Matching or Subject Switching mode should be employed and

- a character string that is either a word or phrase for Phrase Matching or the set of posting terms assigned to a record by a contributing source for Subject Switching.

As the first processing step, the Access Routine creates an array from the input character string. For Phase Matching, each word in the phrase is treated as an individual element, and the words are left in the natural order of the phrase. For Subject Switching, each posting term (which may be single word or multiple word) is treated as an element, and the posting terms are sorted in alphabetical order.

The following examples show a Phrase Matching and Subject Switching input array:

Phrase Matching
Input phrase: Engine Endurance Testing Research Laboratories

Phrase Matching Array:
Engine
Endurance
Testing
Research
Laboratories

Subject Switching
Input DTIC Posting Terms: Engines, Laboratory Tests, Endurance (General), Laboratories

Subject Switching Array:
Endurance (General)
Engines
Laboratories
Laboratory Tests

Aside from the initial difference in creating the input array, processing by the Access Routine is basically the same for the Phrase Matching and Subject Switching modes. A general description of this processing may be found in Appendix A.

Application Interface Programs. The NLD system is designed so that the application program determines the translation mode to be used and the files to be accessed. The Access Routine performs a standard processing routine based on these requirements and returns all matches that it finds to the application program. The application program determines which of the matches will be used. Because of this design, adding new applications or modifying requirements of existing applications does not generally require changes to the NLD system itself. Normally only the application program must be created or modified.

HISTORY

DTIC's Role

Paul Klingbiel, first director of the NLD Project, had been active for 18 years in linguistic research at DTIC. While there, he had initiated a lexical dictionary which became part of DTIC's machine-aided indexing system.

NASA had been studying methods of reducing duplication of work done by other agencies. In 1981, it was decided to move ahead with plans for a NASA Lexical Dictionary, designed to switch automatically the subject terms selected by DTIC's indexers to NASA terminology.

Klingbiel, by then retired from DTIC, agreed to organize the project. Copies of the lexical dictionary software were obtained from DTIC, and programmer Elaine Sellman began a study of NLD requirements.

DTIC's programs were written in COBOL for a UNIVAC mainframe while the Facility used a different programming language, PL1, and an IBM mainframe. So, although the DTIC software was available, it served primarily as an example and the basis for the new NLD programs.

A tape of DTIC's lexical dictionary file also was obtained. This was used to determine how DTIC would translate NASA posting terms into DTIC posting terms and was helpful in constructing entries that translated coordinations of DTIC terms into single NASA terms.

NASA KWOC and Data Entry

Klingbiel began the NLD with a list of NASA posting terms in a special Key Words Out of Context (KWOC) format. A KWOC listing had been used at DTIC to review and correct inconsistencies that had entered into the Natural Language Database. By starting the NLD with a KWOC printout of all of NASA's posting terms and Use references, the problems experienced at DTIC were avoided. In fact, the KWOC became the basic tool for coding NLD entries. (See Figure 3 for a sample page of the NASA KWOC.) Column 1 lists the unique words in the NASA controlled vocabulary in alphabetical order. Column 2 shows all NASA terms and Use references that are in the Thesaurus and that contain the word in column 1. Column 3 lists only NASA Posting terms. These are either the same terms that appear in column 2 or authorized NASA posting terms that are to be used for those in column 2.

Entries for the Lexical Dictionary were selected from column 2. Only entries that began with the word in column 1 were selected; all of the others in that array were selected for coding as they appeared in other sections of the alphabet where the initial word in column 2 and the unique word in column 1 matched.

For example, in Figure 3, note the term OPERATIONS in the second column. It matches the word OPERATIONS in the first column and should be posted to the term appearing in column 3, namely OPERATIONS. The first

	OPERATIONAL AMPLIFIERS	OPERATIONAL AMPLIFIERS
	OPERATIONAL CALCULUS	OPERATIONAL CALCULUS
	OPERATIONAL HAZARDS	OPERATIONAL HAZARDS
	OPERATIONAL PROBLEMS	OPERATIONAL PROBLEMS
	TIROS OPERATIONAL SATELLITE SYSTEM	TIROS OPERATIONAL SATELLITE SYSTEM
OPERATIONS	AIR DROP OPERATIONS	AIR DROP OPERATIONS
	AIRLINE OPERATIONS	AIRLINE OPERATIONS
	FLIGHT OPERATIONS	FLIGHT OPERATIONS
	LOADING OPERATIONS	LOADING OPERATIONS
	MILITARY OPERATIONS	MILITARY OPERATIONS
	OPERATIONS	OPERATIONS
	OPERATIONS RESEARCH	OPERATIONS RESEARCH
	PREFLIGHT OPERATIONS	PREFLIGHT OPERATIONS
	RESCUE OPERATIONS	RESCUE OPERATIONS
OPERATL	GEOSTATIONARY OPERATL ENVIRON SATELLITE B	GOES B (NOAA)
OPERATOR	BERGMAN OPERATOR	BERGMAN OPERATOR
	OPERATOR PERFORMANCE	OPERATOR PERFORMANCE
	STURM-LIOUVILLE OPERATOR	STURM-LIOUVILLE THEORY
OPERATORS	DIFFERENTIAL OPERATORS	DIFFERENTIAL EQUATIONS OPERATORS (MATHEMATICS)
	FREDHOLM OPERATORS	FREDHOLM EQUATIONS OPERATORS (MATHEMATICS)
	LAPLACE OPERATORS	LAPLACE TRANSFORMATION
	OPERATORS	OPERATORS
	OPERATORS (MATHEMATICS)	OPERATORS (MATHEMATICS)
	OPERATORS (PERSONNEL)	OPERATORS (PERSONNEL)
OPHTHALMODYNAMOMETRY	OPHTHALMODYNAMOMETRY	OPHTHALMODYNAMOMETRY
OPHTHALMOLOGY	OPHTHALMOLOGY	OPHTHALMOLOGY
OPIK	OPIK THEORY	OPIK THEORY
OPOSSUM	OPOSSUM	OPOSSUM
OPTICAL	MINITRACK OPTICAL TRACKING SYSTEM	MINITRACK SYSTEM
	OPTICAL ABSORPTION	ELECTROMAGNETIC ABSORPTION LIGHT TRANSMISSION
	OPTICAL ACTIVITY	OPTICAL ACTIVITY
	OPTICAL AMPLIFIERS	LIGHT AMPLIFIERS
	OPTICAL COMMUNICATION	OPTICAL COMMUNICATION
	OPTICAL CORRECTION PROCEDURE	OPTICAL CORRECTION PROCEDURE
	OPTICAL COUNTERMEASURES	OPTICAL COUNTERMEASURES
	OPTICAL COUPLING	OPTICAL COUPLING
	OPTICAL DATA PROCESSING	OPTICAL DATA PROCESSING
	OPTICAL DATA STORAGE MATERIALS	OPTICAL DATA STORAGE MATERIALS
	OPTICAL DENSITY	OPTICAL DENSITY
	OPTICAL DEPOLARIZATION	OPTICAL DEPOLARIZATION
	OPTICAL EMISSION	LIGHT EMISSION
	OPTICAL EMISSION SPECTROSCOPY	OPTICAL EMISSION SPECTROSCOPY
	OPTICAL EQUIPMENT	OPTICAL EQUIPMENT

Figure 3

word of the term immediately following OPERATIONS, i.e. OPERATIONS RESEARCH, also matches the word in column 1, and this item should be posted to the term appearing on the corresponding line in column 3, i.e. OPERATIONS RESEARCH.

The KWOC listing also was used to determine the proper logic code. In the case of OPERATIONS in column 2 which is posted to OPERATIONS in column 3, it would appear that the two are equal and the logic code should be E. However, notice that the next term after OPERATIONS, i.e. OPERATIONS RESEARCH, consists of two words making two elements in the key to the record. For any key with two or more elements or for any single element key that matches the first element of a longer key, the logic code must contain a T. And so the KWOC helped the person coding entries to select the proper logic code.

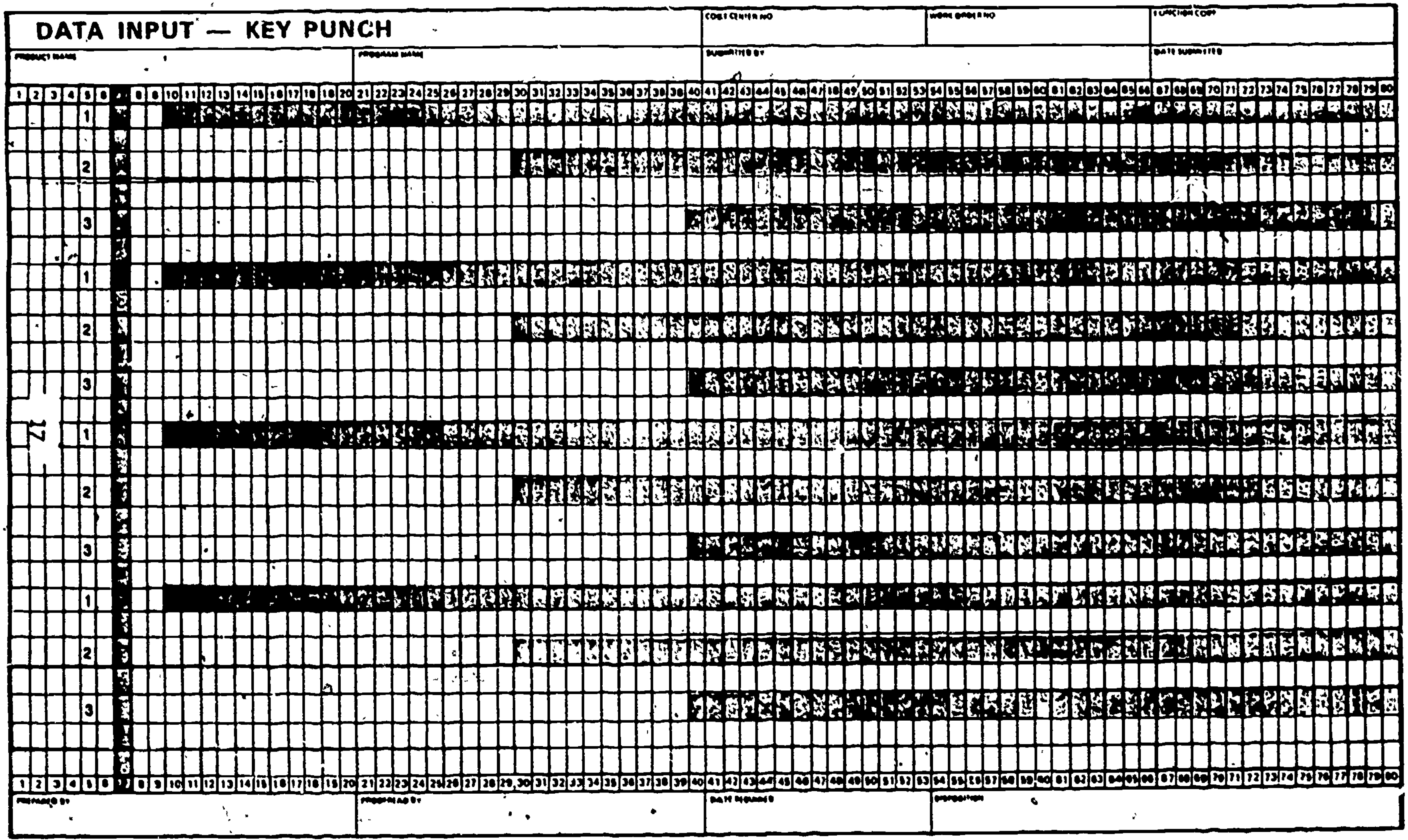
Entries for the Lexical Dictionary were coded for keypunching. Specially printed coding sheets were used (see Figure 4) to keep the various parts of the entry in the proper columns. Three lines (and therefore three cards) were required for each one- or two-element key. For each additional word in a key, three additional cards were coded, punched, and added to the deck. All cards contained an identifying five digit number. The first four digits were assigned consecutively except that the same four digits appeared on three cards before the number changed. When 9999 was reached, the sequence returned to 0001. Since the original record that had been numbered 0001 had already been loaded onto magnetic tape, the duplication of numbers was not confusing. The final or fifth digit of the identifying number was either a 1, 2, or 3. It indicated which of the three parts of the record the card contained. All cards with numbers ending in 1 contained the logic code. For a one- or two-element term, card 1 also contained the first element. Card 2 contained the second element or two zeroes. Card 3 held the posting term for that record. For terms with three or more elements, card 3 contained a continuation symbol, card 4 held the first two elements (separated by a semicolon), card 5 furnished the third element, and card 6 the posting term for a three-element key or another continuation symbol if any additional words were required for the key, and so on.

It can be seen that for a seven-word term -- the longest in the NASA controlled vocabulary -- it was necessary to code and keypunch (n-1)3 cards (where n equals the number of words or elements in a term) or a total of 18 cards. Fortunately, quicker ways are now available for this job.

Logic codes that were being used at that time also were more complicated than those used now and contained some additional intelligence.

At that time card 1 for a three element term would have the logic code of T; card 4 would have a logic code of TT to indicate that a table entry existed within a table entry. If the NASA posting consisted of two or more terms, the T or TT on the first of the final three cards required for the entry would be followed by an L making the logic code TL or TTL.

Several programmers recommended that the initial procedure of creating the NLD entries be automated. However, project director Klingbiel decided that stopping the manual process to reduce the manual procedures to program



DATA INPUT — KEY PUNCH

COST CENTER NO

WORK ORDER NO

FUNCTION CODE

PRODUCT NAME

PROGRAM NAME

SUBMITTED BY

DATE SUBMITTED

PREPARED BY

PROOFREAD BY

DATE RECEIVED

DISPOSITION

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
1
2
3
1
2
3
1
2
3
1
2
3

Figure 4

specifications; brief the programmers, write, test, and debug the programs, and automatically generate the NLD entries would take more time and be less cost effective than finishing the job manually. Therefore, the manual coding and keypunching continued. (For the next effort, candidate entries were created automatically.)

The primary job of coding and keypunching entries was finally completed, but since there were a number of errors to be corrected, Sellman devised a way of doing this online to speed up the process.

During the time when the entries were being coded and the data entered into the file, the records were changed from four fixed-length fields storing the logic code, first element, last element, and posting term, to a VSAM file containing three fields: the logic code, the key, and the posting term. The key for each record was and is unique. Any record in the file could and can be replaced by overlaying another record having the same key. In this way, logic codes and posting terms can be changed. If the error is in the key, it is necessary to delete the record and add it in its correct form.

In the spring of 1982 there were some personnel changes. On April 1, Klingbiel retired from the Facility, but was retained as a consultant to the project. June Silvester became assistant and acting project director, but this job was taken over in late May for eight weeks by Ron Buchan while Silvester was on extended leave. In the meantime, Edna Fleek completed the job of getting the file ready for use. The excellence of her and Buchan's work was attested to by the confidence NASA indexers soon had in the accuracy of the NLD output.

When all errors were corrected, the Phrase Matching file became operational. This meant that the NLD would find and print out the NASA translation for each DTIC term that matched, character for character, either a NASA posting term or Use reference. For example:

Matched	DTIC Posting Term	NASA Posting Term
Posting term	DECODING	DECODING
Use reference	DECOMPRESSION	PRESSURE REDUCTION

The June progress report on the NLD included the following statements:

> The Lexical Dictionary now has about 14,000 records out of a projected 20,000 in the NASA Thesaurus. After the NASA terms have been coded, a tape of NASA Terms will be made that can be run against the NASA Lexical Dictionary to determine misspelled terms as well as missing terms.
>
> Of the computer identified errors, over 300 have been corrected with manual coding and data entry keying. Nearly 200 corrections have been made using the TSO direct-entry program which consumes 1/3 of the labor of the old method.

> A recovery command was developed for the TSO entry system for the Lexical Dictionary, enabling the entry of data more than once a day.

First Operational System

The June report also stated that:

> The Access Routine was tested and has proven workable leaving only questions of format to be considered. This means that we have actually achieved subject switching between DTIC and NASA terms.

On the other hand, Klingbiel's September trip report stated that:

> At this point in time there has been no Subject Switching with either NASA or DTIC data, except in the most trivial and incomplete sense, because neither file as now constituted contains Subject Switching data. Subject Switching cannot occur until the present NLD is upgraded with data to be obtained from successful DTIC/NASA, NASA/DTIC runs.

This seeming disagreement with the statement from the progress report stemmed from a misunderstanding as to the nature of subject switching. We reiterate that subject switching is translating concepts expressed by one or more posting terms from the controlled vocabulary of a contributing organization to the same concept expressed in the posting terms from the target vocabulary, also controlled.

The system had achieved the capability of matching input phrases, character by character -- the first operational segment of the NASA Lexical Dictionary system -- but the translation of concepts was instituted later.

In early September 1982, the NLD file was transferred from magnetic tape to disk files. Also programmer Sellman left the Facility, turning over the NLD development to Duchesne Clark, assisted by Midori Keech.

Klingbiel visited the Facility September 13-24, ironing out problems that had arisen during the summer and laying out in detail the steps to be taken before his next visit in December. These tasks were carried out by the NLD team of Buchan, Fleek, Silvester, Streeks, Clark, Keech, and programmer Patricia Carroll who joined the project in September. The tasks included updating and slightly changing the DTIC Lexical Dictionary, fixing a problem that had been discovered with the way in which glosses were handled, updating the DTIC thesaurus authority listing, doing many error checks and corrections, and finally producing four printouts and a copy each of the NLD and DTIC's Lexical Dictionary.

The first of the four printouts was the result of running DTIC's posting terms through the NLD which, so far, consisted of just one file and a program that could phrase match. This program provided a printout of DTIC terms and matching NASA terms, not only when the entire DTIC term matched, character for character, but also when only part of the term matched. The listing was sorted by the input posting terms, in this case DTIC's.

The second printout was the same information but sorted by the output (NASA's posting terms).

The third printout was the result of running NASA's posting terms through the new version of DTIC's lexical dictionary. The printout was sorted by DTIC's posting terms (the output).

Finally, the fourth printout was the same as the third but sorted by NASA's posting terms.

Collectively these printouts totalled over 2,500 pages. When Klingbiel returned to the Facility on December 6, 1982, it was determined that a more compact presentation of the data was required in order to expedite analysis and data entry.

Discussions with Clark resulted in some changes and reprints of the four printouts. To avoid cumbersome nomenclature, the printouts were referred to as Books 1 through 4, and identified as follows:

Book 1 - DTIC/NASA sorted alphabetically by DTIC terms
Book 2 - DTIC/NASA sorted alphabetically by NASA terms
Book 3 - NASA/DTIC sorted alphabetically by DTIC terms
Book 4 - NASA/DTIC sorted alphabetically by NASA terms

Two of these books were re-sorted. The re-sort analysis conducted by Clark resulted in another software change and finally five copies of each book were printed on 8 1/2" x 11" photocopy paper for use by the NLD team.

The conclusion of Klingbiel's visit on December 10, 1982 coincided with the announcement of Roxanne Newton's appointment to the position of project director. She had joined the project on November 29.

Implementation of Subject Switching

Second Operational System. The data analysis tasks that were to occupy the next few weeks were identified and assigned as follows:

- Analysis of DTIC terms with no mechanically derivable NASA counterpart (Buchan, Streeks).

- Identification of identities between NASA and DTIC terms (Fleek, Newton).

- Compilation of Tables, i.e. coordinated DTIC terms (Silvester).

Another Klingbiel visit to the Facility was scheduled for January 3-7, 1983. In the meantime, the team did some analysis and obtained some hands-on experience with translating DTIC concepts to the same concepts expressed in NASA's terms.

Klingbiel recommended that as the assigned tasks were being carried out, the team:

1. Note anomalous machine translations for subsequent evaluation.

2. Evaluate alternative data entry methods.

3. Collect pertinent statistics which would help in estimating the total workload.

Newton recognized and pointed out that the NLD entries selected from the KWOC could be identified even more easily from the NASA Thesaurus. This is because logic codes are determined by the initial word position and the presence or absence of significant following words. That is, significant words in the medial or final positions in a posting term or Use reference were of interest only to the extent that they existed or did not exist.

A new data entry method was devised and instituted by Newton, Silvester, Clark, and Carroll. At the time of Klingbiel's December visit, the four books of data had been categorized by the type of match that they supplied between the DTIC and NASA vocabularies (i.e., no match, exact match, change, and coordination -- or tables). Except for the "no match" entries, each kind of data was transferred to a dataset that could be edited online. Building the datasets in this way kept the files accurate since the input had been checked and corrected repeatedly throughout the fall months.

The data in the four printouts, books, or datasets presented a variety of problems - most of them anticipated. For instance, "no matches" were expected because DTIC's and NASA's vocabularies are designed to support two different missions. Human analysis of the "no matches" was able to resolve about 80% of the cases leaving 20% of DTIC's terms with no translation. These were zeroed out. As expected, problems in generic level occurred in two ways: DTIC had specific terms for which there was no equally specific NASA counterpart and vice versa.

A problem not explicitly recognized prior to the acquisition of the four books of data was that which was presented by chemical terms. DTIC uses a highly coordinated (Boolean) method of indexing with chemical terms that can produce significant false coordinations when more than one chemical term is indexed for the same document. No obvious solution was apparent.

It was noted in Klingbiel's January trip report that about 10% of the data had been analyzed, major problem areas and solutions had been identified, an efficient data entry technique had been devised, and anomalous data had been noted and either deleted or corrected.

The translations of individual DTIC posting terms to NASA posting terms continued as assigned.

Meanings of all DTIC terms were examined. Meanings of terms that appeared to be identical were compared and translations corrected when homonyms were discovered. The evaluation of candidate coordinated entries also was begun, as was internal documentation and a preliminary study and test of the NLD. As part of the study, the indexers were interviewed individually and confidentially. In addition a test NLD was created, enabling a comparison of DTIC, NLD, and NASA-indexer indexing for a sample of 100 documents.

By April 1, 1983 all DTIC terms had been examined and a translation for each had been entered into the DTIC Subject Switching file. With the loading of these entries into the NLD, the DTIC tapes could be run through the second operational system. That did provide Subject Switching on a limited basis.

The entries consisted of the following:

Type of Entry	Number Coded
Exact Match	5400
Partial Match	4500
No Match	3200

Third Operational System. By April 28 all of the 6,300 table -- or coordination -- entry candidates had been examined. Over 3,000 entries were accepted as presented. Others were accepted with additions or alterations. The remainder were deleted. The table entries then were loaded into the NLD and full Subject Switching became not only available but also operational.

Review and Feedback. The final phase of developing the DTIC/NASA Subject Switching capability of the NLD system began at the end of April 1983 and is ongoing. This consists of adding and revising entries based on feedback from the NASA indexers and on a systematic review of the file by the Lexical Dictionary staff.

PROCEDURES FOR BUILDING A LEXICAL DICTIONARY

Overview of Lexical Dictionary Implementation for Subject Switching

The NASA STI Facility has already developed the following major components of the NLD system:

- file structures for Phrase Matching and Subject Switching,
- coding procedures for Phrase Matching and Subject Switching entries,
- programs for generating candidate Subject Switching entries,
- the Access Routine program,
- online file maintenance and validation programs, and
- application programs suitable for the Facility's uses of the NLD

In order to implement the NLD system for another organization, the following efforts would be required:

- modification of the entry creation programs, the Access Routine, and the online file-maintenance and validation programs to run on a different host system,
- development of application programs suitable for that organization's uses of the NLD system, and
- coding of translation entries to create the Lexical Dictionary data files.

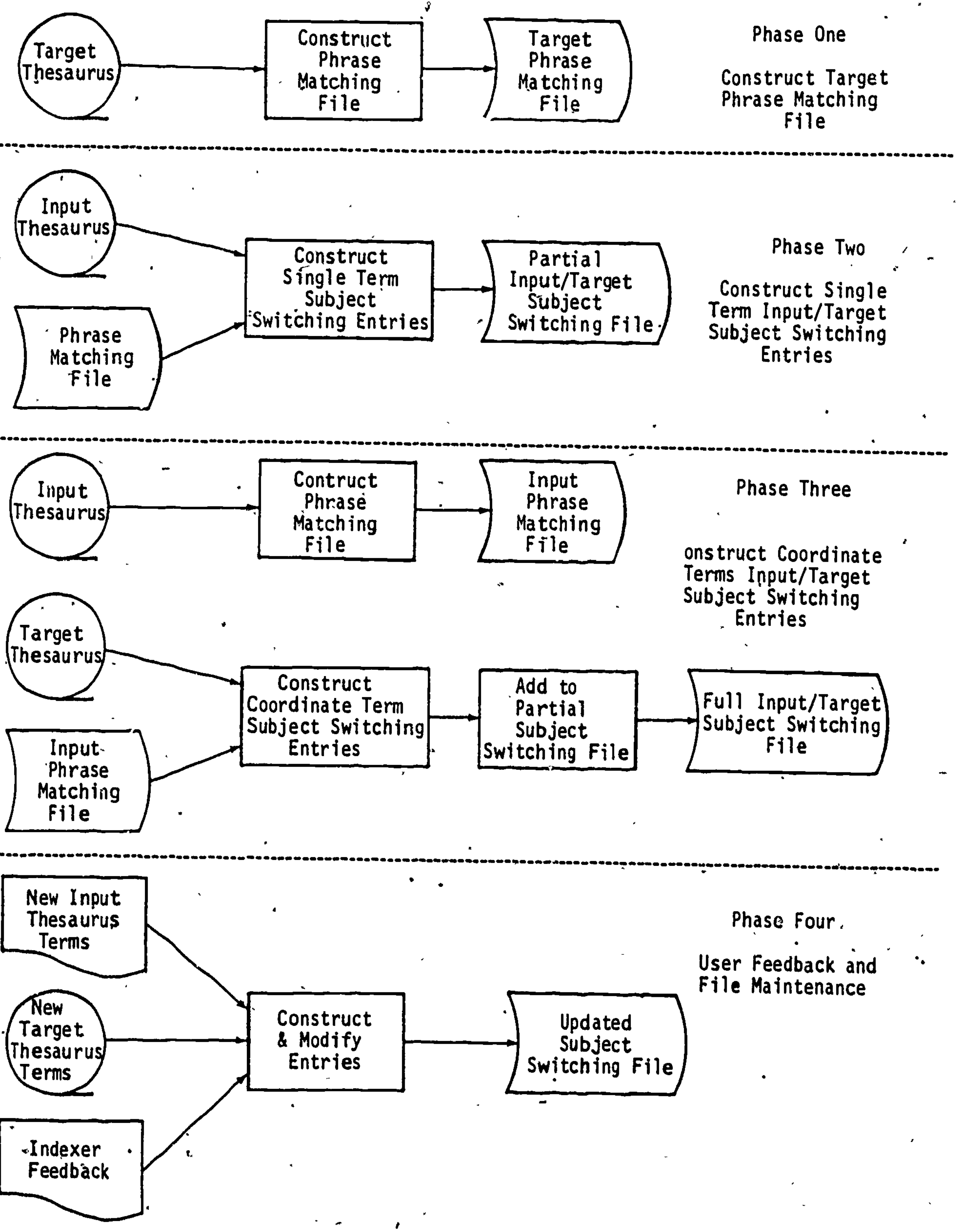
Automated Subject Switching from one vocabulary to another using the NLD system can be implemented in four phases. Figure 5 presents an overview of these four phases.

Phase One centers on the construction of a Phrase Matching file for the target vocabulary (the vocabulary into which input phrases are to be translated). This file consists of entries for every posting term and Use reference in the target thesaurus, as well as additional Use references constructed specifically for the NLD system. The entries for the file can be coded manually or a program can be written to generate them automatically from a machine-readable file of the thesaurus. Using the Phase One or Phrase Matching file, the NLD system will attempt to match any input term or phrase with entries in the file and translate them into target vocabulary posting terms.

In Phase Two, a Subject Switching file is begun. This file is basically a translation table between the posting terms of a contributing source (the input vocabulary) and the posting terms of the target

Figure 5

Overview of Lexical Dictionary Implementation for Subject Switching


Target Thesaurus
Construct Phrase Matching File
Target Phrase Matching File
Phase One
Construct Target Phrase Matching File
Input Thesaurus
Phrase Matching File
Construct Single Term Subject Switching Entries
Partial Input/Target Subject Switching File
Phase Two
Construct Single Term Input/Target Subject Switching Entries
Input Thesaurus
Contruct Phrase Matching File
Input Phrase Matching File
Phase Three
onstruct Coordinate Terms Input/Target Subject Switching Entries
Target Thesaurus
Input Phrase Matching File
Construct Coordinate Term Subject Switching Entries
Add to Partial Subject Switching File
Full Input/Target Subject Switching File
New Input Thesaurus Terms
New Target Thesaurus Terms
Indexer Feedback
Construct & Modify Entries
Updated Subject Switching File
Phase Four
User Feedback and File Maintenance

vocabulary. Entries in the file pair each input vocabulary posting term with the posting term or terms from the target vocabulary that express the equivalent concept. Candidate entries for this file are created by processing the input vocabulary posting terms through the target vocabulary Phrase Matching file created in Phase one. Analysts then evaluate and edit these entries to create the final Subject Switching file. A separate file is built for each input vocabulary to be translated.

Phase Three adds entries for coordinations between posting terms of the input vocabulary to the Subject Switching file created in Phase Two. These coordination entries represent two or more posting terms from the input vocabulary which, when used in combination, translate to a posting term or terms in the the target vocabulary. One way in which Phase Three can be implemented is by creating a Phrase Matching file for the input vocabulary, processing the target vocabulary through this file, and analyzing and editing the resulting candidate entries. The completion of Phase three makes possible full Subject Switching from the input vocabulary to the target vocabulary.

Phase Four is concerned with user feedback and file maintenance. New terms added to both the input thesaurus and the target thesaurus require additions and modifications to entries in the data files. In addition, users can supply feedback as to translations that should be added or modified.

The following sections describe these four phases in detail.

Phase One: Phrase Matching File

Purpose. The creation of the Phrase Matching file makes it possible to attempt to match terms and phrases from any source (see the subsection on Purpose in the INTRODUCTION) with the target vocabulary. Additional Use references from varying forms of target vocabulary terms, such as singulars, plurals, spelling variants, and gerunds, also are put into the Phrase Matching file. The match capability of the system increases with the number of Use references in the file. The Phrase Matching capability can be used for any application requiring the translation of words or phrases into the target vocabulary. The Phrase Matching file is used in building the Subject Switching file and is an essential part of a machine-aided indexing system.

Record Description. Each record in the Phrase Matching file consists of three fields: the logic code, the key, and the posting term.

The logic code in the Phrase Matching file is entered in the first column of the record. This code is selected according to prescribed rules and provides a weak form of syntax for use by the Access Routine in its search for multi-element terms. The logic code also indicates the relationship between the key and the posting term(s).

The key consists of one or more elements. In Phase One, these elements are the individual words that make up the target vocabulary posting terms or Use references. A single element key will end with a semicolon and two zeroes. The key for each entry must be unique and must be combined with only one posting term field. Input for the Phrase Matching file consists of the target vocabulary posting terms, thesaurus Use references, synonyms for and variants of the terms, which become additional Use references.

The posting term field contains one or more posting terms from the target vocabulary. When an input word or phrase matches a key, it is translated to the term or terms in the posting-term field.

For each entry in the Phrase Matching file, it is necessary to determine the key, the posting term(s), and the logic code.

Key. The key of the record being constructed is unique. It is the subject of the record and consists of the words of the term or Use reference being described. The Phrase Matching file is based on keys created from the target thesaurus posting terms and Use references. Additional Use references, such as singulars and plurals, may also be added. Each word in the posting term or Use reference is a separate element in the key.

- If the key consists of only one word, a semicolon and two zeroes are added following this single element. For example:

Term:	Controllability
Key:	Controllability;00

- If the key consists of more than one word, the words (or elements) are separated by semicolons. For example:

  Term: Geological surveys
  Key: Geological;Surveys

- If the key is identical to the first two or more elements of a longer key, then in addition to separating the words by semicolons, a semicolon and two zeroes are added following the final element. For example:

  Terms: Charge transfer devices
  Charge transfer
  Keys: Charge;Transfer;Devices
  Charge;Transfer;00

Some specific formatting rules follow:

- Hyphenated words or two words separated by a slash are treated as a single element. For example:

  Terms: Government/industry relations
  Key: Government/industry;Relations

- An ampersand (&) is treated as a word. For example:

  Term: Atmospheric & Oceanographic Information System
  Key: Atmospheric;&;Oceanographic;Information;System

- Parentheses are dropped from around words in the key. For example:

  Term: Hudson River (NY-NJ)
  Key: Hudson;River;NY-NJ

Posting Term. The posting term field represents the target vocabulary's equivalent of the elements that appear in the key. The posting term or terms are entered exactly as they appear in the target vocabulary thesaurus. In the Phrase Matching file, posting terms listed in the key field are posted to the same term in the posting term field. The Use references in the key field go to one or more valid posting terms in the posting term field. Multiple posting terms are separated by commas. A space is left between words in a posting term, but not between multiple posting terms in the posting field. For example:

Key	Posting Term(s)
Controllability;00	Controllability
Chrome;00	Chromium
Geoastrophysics;00	Astrophysics,Geophysics
Geological;Surveys	Geological surveys
Gold;Plate	Gold coatings

Key	Posting Term(s)
Gold;00	Gold
Government/industry;Relations	Government/industry relations
Hudson;River;NY-NJ	Hudson River (NY-NJ)

Logic Code. The logic code indicates the relationship between the key and the posting term. The first three logic codes are used with single word keys. E indicates that the single word key and the posting term are EQUAL or exact matches. For example:

E	Controllability;00	Controllability

C indicates that the posting term shows a CHANGE from the single word key. For example:

C	Chrome;00	Chromium

L indicates that the single word key is posted to a LIST or multiple posting terms. For example:

L	Geoastrophysics;00	Astrophysics,Geophysics

If the key contains two or more words, the logic code is a T. The T refers to the TABLE format of the coded file entries. For example:

T	Geological;Surveys	Geological surveys
T	Gold;Plate	Gold coatings
T	Hinged;Rotor;Blades	Hinges,Rotary wings

Continuation Entries. When a key exceeds two words, special continuation entries must be made for use in NLD system processing. The key for the first of these continuation entries is made up of the first two words of the term. The next key is created by adding the next word from the term to the key. Additional entries are created in this way until the entire term appears in the key.

A symbol is used in the posting term field to indicate that the program must continue to look for additional key elements in order to reach the proper posting term. The format for the entries required for a term of multiple words is a table. For example:

A term consists of seven words, ABCDEFG, and it is to be posted to a term of three words, HIF. The entries are as follows:

Logic Code	Key	Posting Term(s)
T	A;B	*
T	A;B;C	**
T	A;B;C;D	%
T	A;B;C;D;E	%%
T	A;B;C;D;E;F	%%%
T	A;B;C;D;E;F;G	HIF

The asterisks and percent signs in the posting term field not only tell the Access Routine that additional elements must be located, but also tell the analyst how many elements belong in the entry, how many entries the term requires, and in the case of omissions, which entries need to be added. A program is available that will create the continuation entries, so they do not need to be manually coded.

Special Symbols. In addition to the asterisk and percent sign, discussed under Continuation Entries, other special symbols may be used in the posting term field if they are helpful for a given application. For example, the NASA Thesaurus designates certain ambiguous or very broad terms as Array terms. The Thesaurus recommends use of a more specific term in place of the Array term. When these terms appear in the posting term field of the Phrase Matching file, they are followed by the @ symbol. This symbol alerts the indexers to the fact that the posting term is an Array term. For example:

```
E   Analysis;00                 Analysis@
E   Lifts;00                    Lifts@
```

Coding for Input. the NLD system has an online update program used for adding new entries to the file. For online update, the entry is coded as follows:

```
Logic code$Key$Posting term
```

Elements in the key are separated by semicolons, and single element keys are followed by ";00". Multiple posting terms are separated by commas. The "$" indicates the end of a field.

Examples of Entries Coded for Online Update:

```
E$Controllability;00$Controllability
C$Chrome;00$Chromium
L$Geoastrophysics;00$Astrophysics,Geophysics
T$Geological;Surveys$Geological surveys
T$Hinged;Rotor;Blades$Hinges,Rotary wings
```

When the entries are loaded into the Phrase Matching file, the "$"s that are used as field delimiters are dropped. The fields are entered in the record as follows:

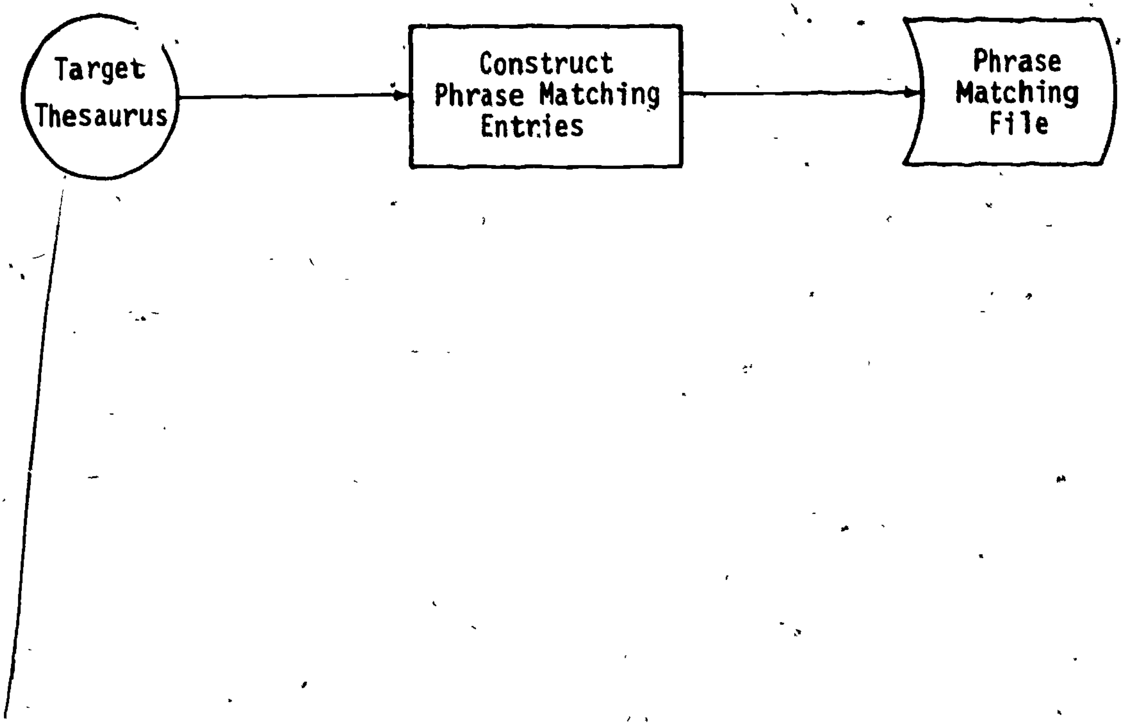
- the logic code in Column 1
- the key in Columns 4 through 127, and
- the posting term in Columns 130 through 400 (variable length).

A full description of the procedures for coding and loading new entries may be found in the section on DATA FILE MAINTENANCE.

Implementation. Figure 6 presents a graphic view of creating the target vocabulary Phrase Matching file. It is a fairly simple process which involves:

Figure 6

Phase One: Creating the Phrase Matching File for the Target Vocabulary



- procuring a copy of the target thesaurus,
- constructing Phrase Matching entries from the target thesaurus using the procedures just described, and
- loading the entries into the Phrase Matching file.

The entries for NASA's original Phrase Matching file were coded manually by analysts, keypunched, and loaded into the file using a batch program. The section on HISTORY provides a detailed description of this development. However, based on the experience gained from building the original file, this process can now be automated to a large extent. If a machine-readable file of the thesaurus is available, a program can be written to generate all of the entries for the posting terms and Use references in the thesaurus. Analysts would still be required to construct additional Use references for variant forms of thesaurus terms. An online program is available that allows direct online data entry to replace keypunching.

Validation. A number of programs have been written that aid in the validation of the Phrase Matching file. One preliminary program compiles an alphabetical list of terms appearing in the posting term field of the NLD file. These terms are referred to as the Lexical Dictionary posting terms. The Lexical Dictionary keys and posting terms are compared with the authority files for thesaurus terms and for Use references for possible errors. Programs exist for the following comparisons:

Check	Against	To Locate
Thesaurus posting terms and Use references	Lexical Dictionary keys	Omissions
Thesaurus posting terms	Lexical Dictionary posting terms	Omissions
Lexical Dictionary posting terms	Thesaurus posting terms	Non-matches
Lexical Dictionary posting terms	Lexical Dictionary keys	Non-matches

Once located, discrepancies are corrected using online maintenance software.

Product. The product of Phase One is the target vocabulary Phrase Matching file and the capability for Phrase Matching input terms and phrases with target vocabulary terms.

Required Programs/Tools. If the manual method of construction is used, the following will be required:

1. Phrase Matching File - A VSAM file with the record structure described in the section labeled Record Description.

2. Online Maintenance Software - A program that creates a load file from online data entry of new records, changes to existing records, and deletions of records.

3. Load Program - A program that loads the load file created by the Online Maintenance Software into the Phrase Matching file.

4. Access Routine - A program that accepts input words and phrases from an application program and returns the posting terms into which the input phrases translate.

5. Continuation Entry Generation Program - A program that creates the continuation entries that are required for keys of three or more elements.

6. Phrase Matching File Validation Programs - A set of error checking programs that validate the entries in the Phrase Matching file.

All of the above programs are available, but may require modification to run on a different computer system.

7. If the automated construction method is used, all of the above programs are required, and a new program must be written to generate the Phrase Matching entries.

8. If the Phrase Matching file is to be used for any translation applications in addition to building candidate Subject Switching entries in Phase two, then an application program must be written for each intended use.

Manpower Estimates. If the automated approach to file construction is selected, it will require an estimated 10 manweeks of labor to build the Phrase Matching file. This represents approximately 4 weeks of programming effort, and 6 weeks of analysis and data entry effort.

If the manual approach is selected, less programming time will be required, but the analysis and data entry time will be approximately tripled, based on the size of the input vocabulary.

Phase Two: Subject Switching File for Individual Input Vocabulary Terms

Purpose. Phase Two provides a limited Subject Switching capability. It involves the creation of a translation for every individual posting term in the input vocabulary expressed in terms of the target vocabulary. The input and output may or may not be the same words, but they must convey the same concept. Phase Two is geared to handle simple individual term switches such as those shown below, but not the complex coordinations that are addressed in Phase Three. The following examples are taken from the DTIC/NASA Subject Switchina file:

Logic Code	Key (DTIC Posting Term)	Posting Term (NASA translation)
E	Radar;00	Radar
C	Adenine;00	Adenines
C	Bases chemistry;00	Bases (chemical)
C	Carbon carbon composites;00	Carbon-carbon composites
C	Drilling machines;00	Boring machines
I	Estimates;00	Estimates?,Estimating?
L	Fluorescent dyes;00	Dyes,Fluorescence

Record Description. Each record in the Subject Switchina file consists of the same three fields already described for the Phrase Matching file:

- Key
- Posting Term
- Logic Code

This record differs from the records in the Phrase Matching file in the following ways:

- the logic code is recorded in the second column of the record rather than the first,
- the elements of the Subject Switching key consist of posting terms (which may be single or multiple words) rather than individual words, and
- the posting terms that constitute the elements of the key come from the thesaurus of a contributing organization.

The keys for all entries created in Phase Two consist of a single element followed by a semicolon and two zeroes. As stated above in Phase Two these elements are the single and multi-word posting terms that make up the input vocabulary. Each key is unique because the contributing organization's posting terms are each unique.

The posting term field represents the target vocabulary posting term or terms that express the concept equivalent to the input vocabulary posting term in the key.

The logic code provides a weak syntax for use by the Access Routine in its processing and indicates the relationship between the key and the posting term.

For each entry in the Subject Switching file, it is necessary to determine the logic code, the key, and the posting term or terms.

Logic Code. Phase Two logic codes E, C, and L are determined in essentially the same way as in Phase One. However, in Phase Two the logic code will be entered in the second column of the record.

Logic Code ƀE, or blank E, indicates that each organization has identically spelled terms with identical meanings as used in the context of each environment, and therefore the key and the posting term are exact matches. For example:

ƀE	Europe;00	Europe
ƀE	Aircraft carriers;00	Aircraft carriers

Logic ƀC indicates that the posting term in the target vocabulary shows some change from the posting term in the input vocabulary. The input term may be singular, while the target term is plural. For example:

ƀc	Adenine;00	Adenines

The input term may have a different form of a word. For example:

ƀC	Bases chemistry;00	Bases (chemical)

One term may have a hyphen which the other omits. For example:

ƀC	Carbon carbon composites;00	Carbon-carbon composites

The target term may be different from the input term, but it means essentially the same thing. For example:

ƀC	Drilling machines;00	Boring machines

In each case, there is a change in the term but not in the concept or subject described by the term.

Logic Code ƀL indicates that a list of multiple posting terms from the target vocabulary are necessary to convey the same meaning as the term from the input vocabulary. For example:

ƀL	Femoral arteries;00	Arteries,Femur

Each of the above logic codes is used for single term entries only. That is, the key contains only one element which in Phase Two is a posting term from the input vocabulary, followed by a semicolon and two zeroes.

Two new codes are used in Phase Two in the Subject Switching file.

Logic Code ƀI indicates that the proper translation is context dependent and therefore indeterminate and must be an indexer choice. An indeterminate translation is flagged with a question mark. For example:

ƀI   Estimates;00   Estimates?,Estimating?

The input vocabulary has only the term "estimates" to cover both of the concept of "estimates" and "estimating" that are found in the target vocabulary. The correct translation must be selected by the indexer based on the document at hand.

In another case, the terms appear to be the same but have a slight difference in meaning. For example:

ƀI   Performance tests;00   Performance tests?

The target vocabulary's thesaurus limits the use of "performance tests" to apply only to operating equipment. The organization contributing the input vocabulary uses "performance tests" for equipment, systems, or human performance. Therefore, the terms may or may not be equivalent depending upon the context. The indexer will have to choose.

Logic Code ƀ0 is the only numeric logic code used. Whenever a translation of a term from the input vocabulary is not wanted or when the target vocabulary does not have an acceptable translation, the logic code used is zero (0). For example:

ƀ0   Peer groups;00   00

Key. As stated, in Phase Two, the elements of the key are terms from the input vocabulary, not the words of a term as in Phase One. The key contains only one posting term, and two zeros are added as a place holder for the second element. An entry is created for every individual posting term in the input vocabulary.

Posting Term. In the Subject Switching file, the posting term is selected by analysts familiar with both the input and the target vocabularies. The posting term field contains one or more posting terms from the target vocabulary or the codes 00 or NIS. The contents of the posting term field reflect the best translation that can be made of the concept expressed by the individual term from the input vocabulary which is in the key. Sometimes there will be an exact match between an input vocabulary posting term and a target vocabulary posting term. In some instances, the translation will reflect only the addition or subtraction of an "s" or a hyphen. In other cases, the term may change to a different term or to a list of terms. A translation may not be possible or not be wanted and the term is "zeroed out." The logic code is entered as zero and the posting term as two zeroes. A term considered Not In Scope is posted to "NIS".

Symbols. In Subject Switching, three new symbols are introduced into the posting term field in addition to the Array symbol (@) described under "Special Symbols" in Phase One. When one of these symbols is used, it immediately follows the term to which it applies.

- Indexer Choice (?)

  The question mark, discussed under logic code I, is used when the proper translation is context dependent and therefore indeterminate. The indexer is presented with a choice of terms, each flagged with a question mark.

- Broader Term Translation (>)

  When the suggested target term is of a broader generic level than the input term, the Lexical Dictionary posting term is followed by a "greater than" (>) symbol. For example:

  ƀC Jugular vein;00 Veins>

- Additional Target Vocabulary Narrower Terms (+)

  When the suggested target posting term has narrower terms which are not covered by the vocabulary of the contributing organization, a plus sign (+) immediately follows the target posting term. For example:

  ƀE Bolts:00 Bolts+
  (The input vocabulary has no narrower terms to "Bolts", but the target vocabulary has narrower terms "Rock bolts" and "Tie bolts".)

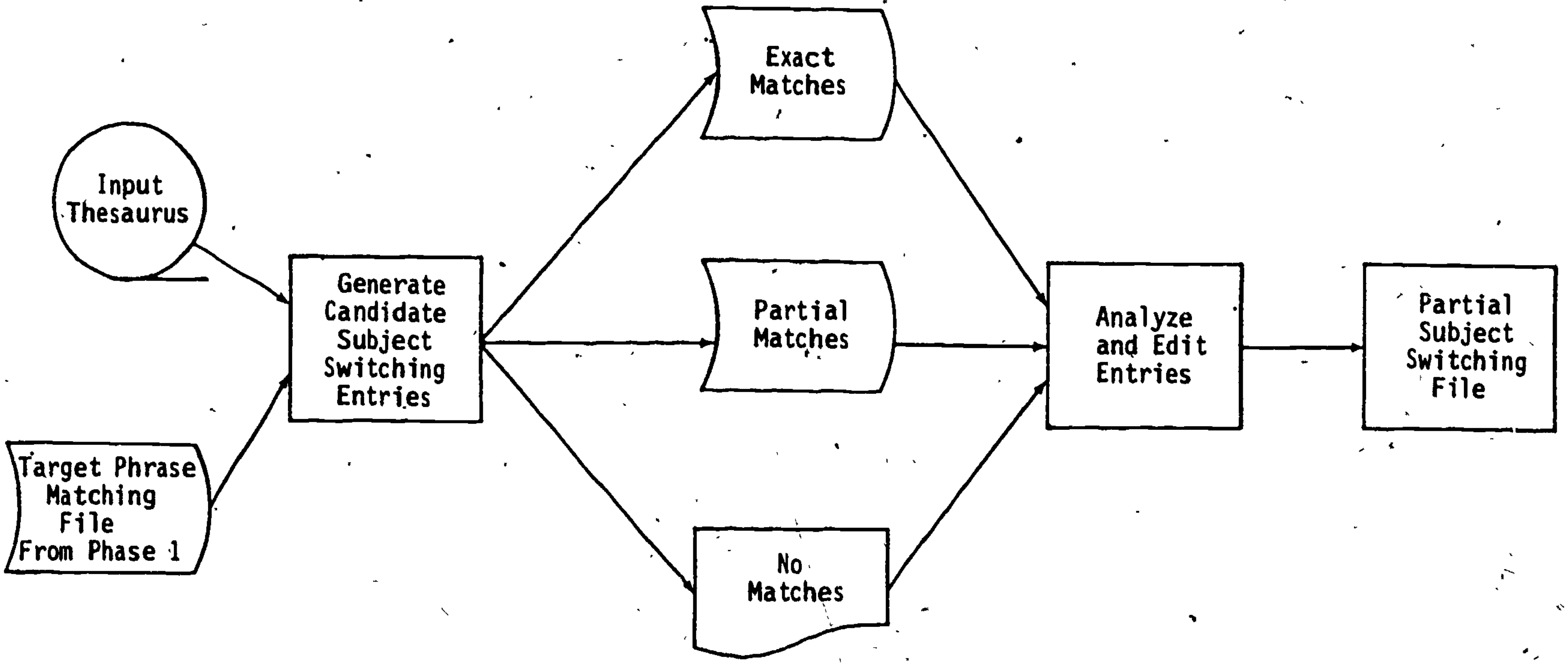
Implementation. Figure 7 presents an overview of Phase Two implementation. A machine-readable file of the posting terms of the input vocabulary is required. This file is processed through the NLD system using the target-vocabulary Phrase Matching file constructed in Phase One and Phrase Matching logic. For each input posting term either an exact match, a partial match, or no match is found. By computer program, base files are created that contain candidate entries for the exact matches and the partial matches. These files are printed, reviewed by analysts, then edited online. When editing is complete, they are loaded into the Subject Switching file. The no-match group is printed and researched by analysts. These no matches are translated into target vocabulary equivalents, if possible, or are "zeroed out", that is, translated to a posting term of 00. In a few instances, new terms may be added to the target vocabulary to translate these terms. A no-match file is then created using the online-update program. When all entries are edited, they are loaded onto the master Subject Switching file.

Validation. Programs exist for the following comparisons in the completed Subject Switching file:

Check	Against	To locate
Input vocabulary posting terms	Keys	Omissions
Keys	Input vocabulary posting terms	Non-matches
Lexical Dictionary posting terms	Target thesaurus posting terms	Non-matches

Figure 7

Phase Two: Creating the Subject Switching File for Individual Input Posting Terms


Input Thesaurus
Target Phrase Matching File From Phase 1
Generate Candidate Subject Switching Entries
Exact Matches
Partial Matches
No Matches
Analyze and Edit Entries
Partial Subject Switching File

When discrepancies are found, corrections are made using online-maintenance software.

Product. The product of Phase Two is a partial Subject Switching file and the capability for Subject Switching from individual input posting terms to target-vocabulary posting terms.

Required Programs/Tools. Phase Two development requires four programs described in Phase One:

1. Phrase Matching File - now completed
2. Online Maintenance Software
3. Load Program
4. Access Routine.

In addition, Phase Two development requires:

5. Subject Switching Build Program - A set of programs which process a machine-readable file of the input posting terms through the Phrase Matching file and creates:
   - a file of candidate entries for exact matches,
   - a file of candidate entries for partial matches, and
   - a printout of no matches.
6. Software for editing the files of candidate entries - software package with text editing capabilities such as TSO, SPF, or WYLBUR is helpful.
7. Subject Switching File Validation Programs - Error checking routines which validate that there is a key to match every input posting term, that all elements of the key are valid input posting terms, and that all entries in the posting term field are valid target posting terms.
8. An application program for each Subject Switching application, if not already developed in Phase one.

Manpower Estimates. Coding for Phase Two will require approximately two manweeks per 1000 terms in the input vocabulary. If the additional programs required for Phase Two must be modified to run on a different system, some programming time will be required. In addition, programming is required to develop software for the specific applications for which the Subject Switching capability is being developed.

Phase Three: Subject Switching File for Coordinated Input Vocabulary Terms

Purpose. Phase Three concentrates on translating concepts expressed by coordination of multiple input vocabulary posting terms into target vocabulary posting terms. Completion of Phase Three provides full Subject Switching capability.

Record Description. Phase Three is an expansion of the Subject Switching file created in Phase Two; therefore, the record is the same as that described in Phase Two. The record consists of the same three fields: the logic code, the key, and the posting term. The logic code is recorded in the second column, and the elements of the key are posting terms from the input vocabulary. The records created in Phase Three differ from those in Phase Two in that the key will always contain at least two elements and that the logic code is always T. The posting term field contains the target vocabulary posting term or terms which express the concept equivalent to the coordination of input posting terms in the key. For example:

ƀT Accident investigations; Aircraft	Aircraft accident investigation

For each entry in the Subject Switching file, it is necessary to determine the logic code, the key, and the posting term. Appendix B contains the procedures followed for creating DTIC/NASA Subject Switching entries for DTIC term coordinations, which can be used as a guide.

Logic Code. The logic code is always ƀT.

Key. Determining the key is a decision-making process performed by an analyst. It is based upon a study of the vocabulary and the indexing practices and policies of the contributing organization. The key always contains at least two input posting terms which, when taken together (coordinated) convey the same concept as the target vocabulary posting term or terms in the posting term field for that entry. Continuation entries, discussed under Phase One, are required for entries with three or more elements in the key.

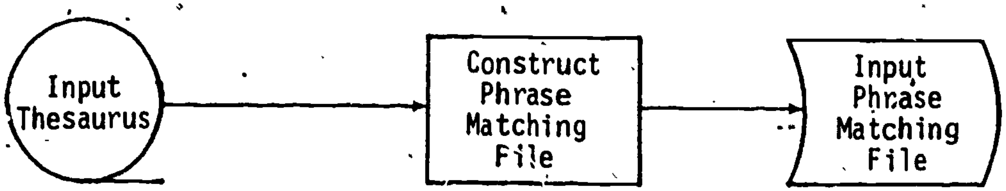
Posting Term. The posting term field may contain one or more target vocabulary posting terms. The concept expressed by the posting term field should be the same as that expressed by the key.

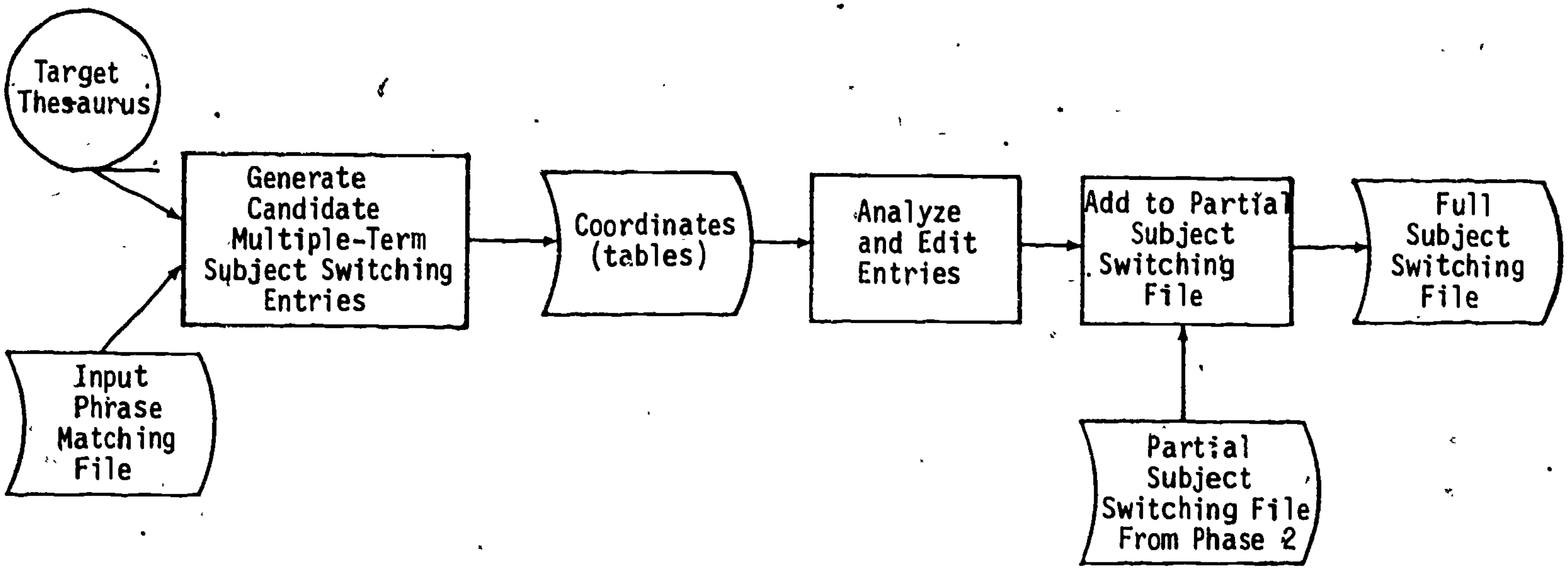
Implementation. Figure 8 presents an overview of one implementation option for Phase Three. This option consists of generating candidate term coordination entries by processing the target vocabulary through the input vocabulary Phrase Matching file. All target vocabulary terms which translate into two or more input vocabulary terms are selected as candidate entries. The program formats these entries according to the rules for the Subject Switching file. The input vocabulary terms (which were the output of the Phrase Matching file) become the keys of the Subject Switching entry. The target vocabulary posting term (which was the input to the Phrase Matching file) becomes the Subject Switching posting term.

Figure 8

Phase Three: Creating the Subject Switching File for Coordination of Input Posting Terms

To construct the Input Phrase Matching File if one does not exist:




Target Thesaurus
Input Phrase Matching File
Generate Candidate Multiple-Term Subject Switching Entries
Coordinates (tables)
Analyze and Edit Entries
Add to Partial Subject Switching File
Partial Subject Switching File From Phase 2
Full Subject Switching File

For example:

Target Vocabulary Posting Term:	Abrasion resistance
Input Vocabulary Translation: (from Phrase Matching file)	Abrasion, Resistance
Creates Subject Switching Entry; BT Abrasion;Resistance	Abrasion resistance

Analysts review and edit the candidate entries generated by the program. If the contributing organization has a lexical dictionary, it can be used to create the candidate entries for Phase Three. (DTIC and the NASA STI Facility have lexical dictionaries.) If no Phrase Matching file exists for the input vocabulary, one can be created using the procedures described in Phase One.

The table entries can also be created by feedback from indexers who spot combinations while indexing.

Another possibility is making a study of documents which have been indexed independently using the input vocabulary and the target vocabulary. By comparing the lists of posting terms assigned by the two vocabularies, coordination should become apparent to a trained analyst.

Any one of these options, or some combination of them, may be used to create table entries. When all entries are in a file, reviewed, and edited, they are loaded onto the master Subject Switching file for the contributing organization.

Validation. The same validation routines and correction procedures used for Phase Two files may be used for Phase Three.

Products. With the addition of the coordinated DTIC terms to the Subject Switching file, the NLD System achieved the capability for full Subject Switching from DTIC indexing to NASA indexing. The product as the indexer sees it is a printout with two lists of terms, one from DTIC's fields 23 and 25 and the other of the NASA terms to which DTIC's terms have been translated. See Figure 9.

Required Programs/Tools. Phase Three requires the following components already described in Phases One and Phase Two:

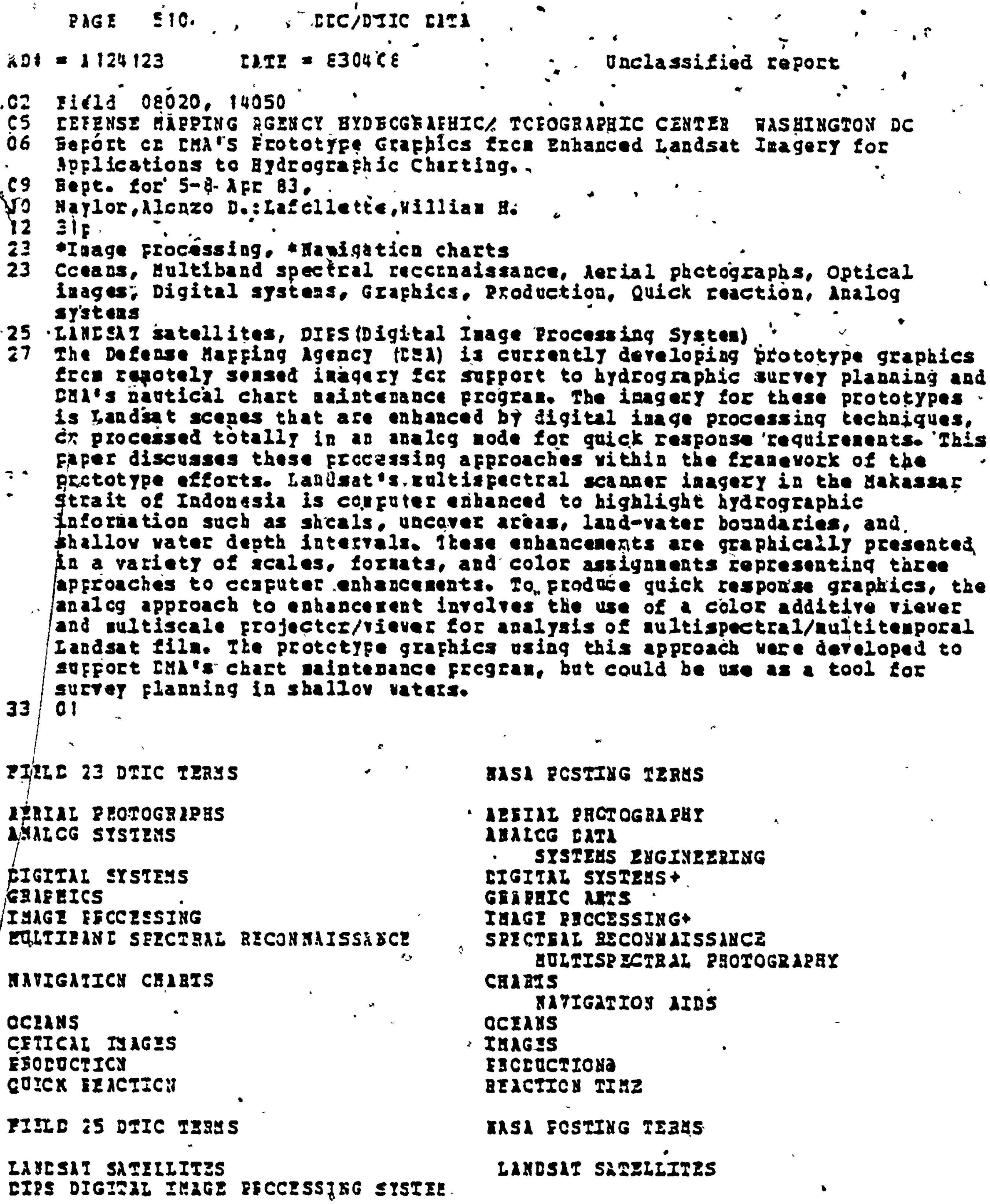
1. Online Maintenance Software
2. Load Program
3. Access Routine
4. Subject Switching File Validation Programs
5. Continuation Entry Generation Program
6. Application Program

In addition, if the automated approach to creating and editing candidate entries is selected, the following programs will be required:

7. Program to generate the input vocabulary Phrase Matching file (if not created for Phase One).

Figure 9

DTIC/NASA Subject Switching Output



```
     PAGE   510        DTC/DTIC DATA

AD# = A124123       DATE = 830408                 Unclassified report

 02  Field  08020, 14050
 05  DEFENSE MAPPING AGENCY HYDROGRAPHIC/ TOPOGRAPHIC CENTER  WASHINGTON DC
 06  Report on DMA'S Prototype Graphics from Enhanced Landsat Imagery for
     Applications to Hydrographic Charting.
 09  Rept. for 5-8 Apr 83,
 10  Naylor,Alonzo D.;Lafollette,William H.
 12  31p
 23  *Image processing, *Navigation charts
 23  Oceans, Multiband spectral reconnaissance, Aerial photographs, Optical
     images, Digital systems, Graphics, Production, Quick reaction, Analog
     systems
 25  LANDSAT satellites, DIPS(Digital Image Processing System)
 27  The Defense Mapping Agency (DMA) is currently developing prototype graphics
     from remotely sensed imagery for support to hydrographic survey planning and
     DMA's nautical chart maintenance program. The imagery for these prototypes
     is Landsat scenes that are enhanced by digital image processing techniques,
     or processed totally in an analog mode for quick response requirements. This
     paper discusses these processing approaches within the framework of the
     prototype efforts. Landsat's multispectral scanner imagery in the Makassar
     Strait of Indonesia is computer enhanced to highlight hydrographic
     information such as shoals, uncover areas, land-water boundaries, and
     shallow water depth intervals. These enhancements are graphically presented
     in a variety of scales, formats, and color assignments representing three
     approaches to computer enhancements. To produce quick response graphics, the
     analog approach to enhancement involves the use of a color additive viewer
     and multiscale projector/viewer for analysis of multispectral/multitemporal
     Landsat film. The prototype graphics using this approach were developed to
     support DMA's chart maintenance program, but could be use as a tool for
     survey planning in shallow waters.
 33  01


FIELD 23 DTIC TERMS                         NASA POSTING TERMS

AERIAL PHOTOGRAPHS                          AERIAL PHOTOGRAPHY
ANALOG SYSTEMS                              ANALOG DATA
                                                SYSTEMS ENGINEERING
DIGITAL SYSTEMS                             DIGITAL SYSTEMS+
GRAPHICS                                    GRAPHIC ARTS
IMAGE PROCESSING                            IMAGE PROCESSING+
MULTIBAND SPECTRAL RECONNAISSANCE           SPECTRAL RECONNAISSANCE
                                                MULTISPECTRAL PHOTOGRAPHY
NAVIGATION CHARTS                           CHARTS
                                                NAVIGATION AIDS
OCEANS                                      OCEANS
OPTICAL IMAGES                              IMAGES
PRODUCTION                                  PRODUCTION
QUICK REACTION                              REACTION TIME

FIELD 25 DTIC TERMS                         NASA POSTING TERMS

LANDSAT SATELLITES                           LANDSAT SATELLITES
DIPS DIGITAL IMAGE PROCESSING SYSTEM
```

8. Table Entry Build Program - A program which takes the output from running the target thesaurus through the input Phrase Matching file and creates a file of candidate entries from the partial matches.

9. Software for editing the file of candidate entries.

Manpower Estimates. The level of effort required for Phase Three will depend upon the implementation option selected. It is estimated that the automated approach of creating and editing candidate entries will require approximately 2 manweeks per 1000 automatically generated table entries. In addition, programming effort may be required for the programs numbered 7 and 8 listed above. No time estimates are available for the other options.

Phase Four: User Feedback and Maintenance

Purpose. The purpose of Phase Four is the establishment of procedures for handling updates to the Lexical Dictionary data files based on updates to the input and target vocabularies and feedback from the indexers. Figure 10 presents an overview of Phase Four activities.

Updates to the Input Vocabulary. Whenever a contributing organization adds terms to or changes terms in its controlled vocabulary, changes must be made in the Subject Switching file. At the very least, one entry must be made for each new individual term, together with its logic code and a translation into the target vocabulary.

In addition, the input vocabulary should be studied for possible additional tables or improved tables which should be entered. Procedures for making changes are covered in the section on DATA FILE MAINTENANCE.

It is desirable to make arrangements with the contributing organization for automatic receipt of information on thesaurus changes. Without continuing communication between the organizations and the necessary information for updates, there will be no way to distinguish between new terms which should be added to the Subject Switching file and errors.

Updates to the Target Vocabulary. Whenever terms are added or changed in the target vocabulary, this must be reflected in the Phrase Matching file. An entry will be made for each new posting term and Use reference, and also for variant forms and synonyms. Complete procedures for adding entries are covered in the section on DATA FILE MAINTENANCE.

The Subject Switching file will also be affected by the addition of new target vocabulary terms. Analysts must look for possible additional tables, improved translations, or for new translations of terms previously zeroed out.

Updates should be made on a regular basis.

User Feedback. User feedback, such as from indexers, is an important part of the intellectual effort in Subject Switching. With specific documents in hand, indexers are uniquely able to verify whether suggested translations are appropriate. Indexers can spot new coordinations which should be added to the Subject Switching file, or coordinations which should be modified or deleted.

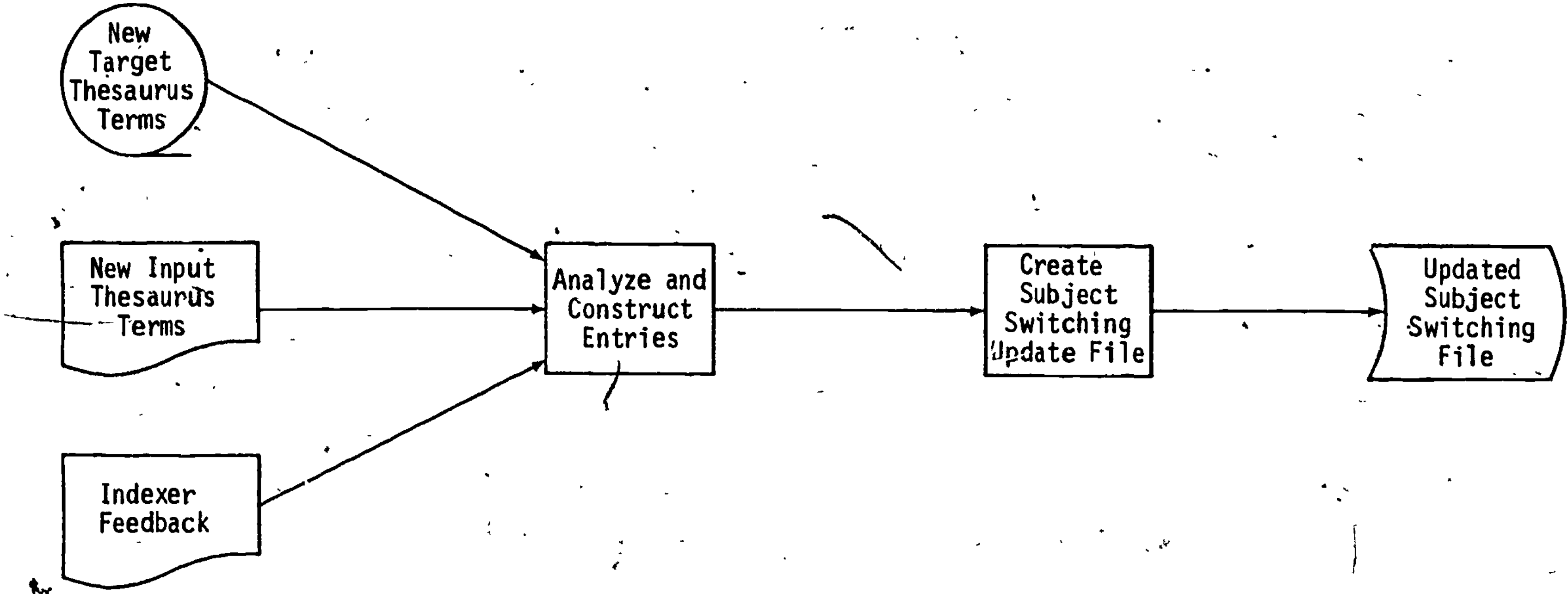
It is anticipated that indexer feedback will suggest:

- Modifications to translations based on operational experience,
- Changes of translations based on new terms in either vocabulary,
- Additions of table entries, and
- Deletions of table entries.

Feedback must be written and two-way, between the indexers and the Lexical Dictionary team. An orientation meeting prior to the implementation of full Subject Switching is essential for initiating the feedback process.

Figure 10

Phase Four: User Feedback and Maintenance


New Target Thesaurus Terms
New Input Thesaurus Terms
Indexer Feedback
Analyze and Construct Entries
Create Subject Switching Update File
Updated Subject Switching File

The NASA STI Facility has designed a form to streamline the feedback process and to encourage the inclusion of all needed information. See Figure 11.

Subject Switching Error List. In any application involving Subject Switching capability, it is useful to produce an error list of all input terms and partial coordinations which could not be matched in the Subject Switching file. Analysts should review these to determine if new entries should be added to the file.

Validation. To check for accuracy, the validation programs described for Phases One, Two, and Three are run periodically. Any errors which are detected by these validation programs are corrected using the online maintenance software.

Required Programs/Tools. Phase Four requires:

1. An operational NLD System with Access Routine, data files, and application programs.

2. Online maintenance software.

3. Phrase Matching file and Subject Switching file validation programs.

Manpower Estimates. It is estimated that each lexical dictionary data file will require approximately 5 manweeks of maintenance each year after an initial "shake-down" period.

NASA LEXICAL DICTIONARY FEEDBACK

Date(s) ______________

Analyst ______________

AD or DE Report No.	Input Term	NLD Translation	Recommended NLD Translation	Comments

Figure 11

DATA FILE MAINTENANCE

Sources of Change to Data Files

Four sources of change to the NLD data files have been identified.

- Changes in the NASA Thesaurus. Because the Phrase Matching file contains an entry for every NASA posting term and thesaurus Use reference, this file must be updated every time the NASA Thesaurus is changed. New NASA terms may replace old translations in the Subject Switching file or otherwise change translations already recorded.

- Changes in the Input (DTIC) Thesaurus. The Subject Switching file contains an entry for every posting term in the input vocabulary; therefore, every new input term must be translated and this translation added to the file.

- Changes Recommended by Indexer Feedback. These may be for either data file and are entered following approval by the abstracting/indexing supervisor and the NLD project director:

  - Phrase Matching file Use references.
    There is an ongoing effort to increase the number of Use references constructed specifically for the NLD. These consist of synonyms, variant spellings, and different word forms of NASA posting terms. The match capability of this file, designed for general purpose phrase matching, increases with the number of Use references in the file.

  - DTIC/NASA Subject Switching file.
    Indexers provide recommendations for improved translations based on actual documents in hand. Most of these suggestions initiate changes in the Subject Switching file.

- Changes Derived from Lists. Lists of input terms that find either no match or only a partial match in the NLD are printed out each time that a DTIC tape is run through the NLD Access Routine. These lists are called exception listings. See Figure 12. The first column on the printout shows how many times the term or combination of terms was encountered on this tape. The second column gives the DTIC accession number of the first occurrence. The third column indicates the DTIC field from which the DTIC posting term came; Field 23 for descriptors, DTIC's controlled vocabulary - Field 25 (unmarked) for DTIC's identifiers or open-ended indexing. The final column shows the partial matches--combinations of terms that are part of a longer coordinated entry-- or unmatched DTIC terms. The ones in this example did not translate because of an input error. One added and the other omitted an "s". New terms would appear here, too, if they had not been added to the NLD.

```
04/26/84 DTIC TERMS NOT FOUND IN NASA LEXICAL DICTIONARY       1
    1   C033784 23 ALTITUDE;GUIDED MISSILES
    2   B080915 23 ALUMINUM;COMPOSITE MATERIALS
    1   A139253 23 BRIDGES;CIRCUITS
    1   B080692 23 CIRCUITS;CONTROL
    1   B080706 23 COMMUNICATIONS NETWORKS;GLOBAL COMMUNICATIONS
    3   B080729 23 COMPOSITE MATERIALS;MATRIX MATERIALS
    1   A139438 23 DATA PROCESSING;DATA STORAGE SYSTEMS
    1   C033792 23 DETECTION;HIGH ALTITUDE
    1   A139261 23 ESTIMATES;ORBITS
    1   B080888 23 FIBER REINFORCEMENT;GLASS FIBERS
    1   B080682 23 FLIGHT;SPACE FLIGHT
    1   A139485 23 HAZARDS;SAFETY
    2   A139476 23 HIGH RATE;INTENSITY
    1   B080693 23 LIMITATICNS;POWER
    1   B080799 23 MEASUREMENT;PARTICLES
    1   A139271 23 SLOPE
    2   B080719 23 TEST METHODS;THERMAL PROPERTIES
    1   A139216 23 VAPOR
    1   A139227    A EXCITCNS
    1   B080621    A/A37U-15 TOWING REELS
    1   B080823    ABCS AIRBORNE BEAM CONTROL SYSTEM
    1   A139155    ACB AIR CUSHION BARGES
    1   A139337    ACES AIRDROP CONTROLLED EXIT SYSTEM
    1   B080779    ACOUSTIC HOLOGRAPHY
    1   A139482    ACOUSTIC IMAGES
    1   B080916    ACOUSTOOPTIC CELLS
    1   C033856    ACTIVE MASS INJECTION
    3   B080720    ADAPTIVE ANTENNAS
```

Figure 12
Exception Listing.

Some of the Field 25 terms may be the same as authorized NASA terms except for an acronym preceding the DTIC term or for some variations in spelling. Any such DTIC terms now initiate new entries into the Phrase Matching file. The exception listing also may suggest new coordinations of DTIC terms that could be translated to a NASA term.

To summarize, the files changed by various sources of input are as follows:

Input Material	Phrase Matching	Subject Switching
NASA Thesaurus update	X	X
Input thesaurus update		X
Indexer feedback	X	X
Exception listings	X	X

The NLD maintenance procedures triggered by each input source will be described in the following sections.

Record Coding

In each instance, the correct logic code, key, and posting term(s) for the record will be determined and written out for data entry. For online update, the entry is coded as follows:

Logic code$Key$Posting term

When needed, the posting term will be followed by a symbol, as previously described. Elements in the keys of the Phrase Matching file are words; in the Subject Switching file, they are terms from the vocabulary of the contributing organization.

Elements in the key are separated by semicolons, and single element keys are followed by ";00". Multiple posting terms are separated by commas. The "$" separates the fields.

Here are some examples of entries coded for online updating:

For Phrase Matching

```
E$Analyzing;00$Analyzing@
E$Controllability;00$Controllability
L$Geoastrophysics;00Astrophysics,Geophysics
T$Geological;Survey$Geological surveys
T$Hinged;Rotor;Blades$Hinges,Rotary wings
```

For Subject Switching

```
E$Acids;00$Acids
C$Amino plastics;00$Thermosetting Resins>
```

For Subject Switching (continued)

```
L$Animal diseases;00$Diseases,Veterinary medicine
T$Blood Circulation;Brain$Brain circulation
T$Blood Circulation;00$Blood circulation+
I$Abiotic processes;00$Abiogenesis?
O$Acne;00$NIS
O$Aerial pickup system;00$00
```

When the entries are loaded into the file, the "$"s which are used as field delimiters are dropped, and the online maintenance software automatically places the logic code in the correct column. The fields are entered in the records as follows:

- the logic code in column 1 for the Phrase Matching file and in column 2 for the Subject Switching file,
- the key in columns 4 through 127, and
- the posting term in columns 130 through 400 (variable length).

Maintenance Functions

The functions or capabilities provided by the NLD's online maintenance system are executed through series of commands. These allow maintenance personnel to process input from any of the maintenance sources described above. A separate set of commands is provided for each of the NLD data files. The chart below indicates the capabilities or functions provided by the maintenance system, along with the command used to carry out each function for each of the NLD data files.

Maintenance Functions	Maintenance System Commands	
	Phrase Matching File	DTIC/NASA Subject Switching File
Creating Authority Files	VALSETUP	DTICVSAM
Data File Validation	NASAVAL	DTICVAL
Entering Update Transactions	NASAUPDT	DTICUPDT
Loading Transaction Files	NASALOAD	DTICLOAD
Printing Maintenance Tool	NASAPRNT	DTICPRNT
Printing Maintenance Tool	NASANVRT	DTICNVRT
Creating Backup Tapes	NASABKUP	DTICBKUP

The commands listed above, in addition to several miscellaneous maintenance commands, are explained in more detail in the section on "Maintenance Commands".

Additions of New Records. To add a record to any file, use the appropriate update command as indicated in the table of online maintenance commands above. The form of the entry is:

Logic Code$Key$Posting term either with or without a symbol.

Deletion of Existing Records. To delete a record, from any file, use the appropriate update command as indicated in the table of online maintenance commands above. Enter DEL dollar sign and the key of the unwanted record. For example:

Existing record:
T$DISTRIBUTION;PARAMETERS$DISTRIBUTED PARAMETER SYSTEMS proves to be a poor choice of coordinated terms for translation. To prevent the coordination of these terms in future translations, the record must be deleted.

Enter: DEL$DISTRIBUTION;PARAMETERS

Changes to an Existing Record. To change the key, use the appropriate update command as indicated in the table of online maintenance commands above. Delete the existing record and add the record in its correct form.

Existing record: E$ERUOPE;OO$EUROPE must be deleted as the key is misspelled.

Enter: DEL$ERUOPE;OO to erase the error and
Enter: E$EUROPE;OO$EUROPE to add the correct record.

Changes to Logic Code Field of a Record. To change a logic code, use the appropriate update command as indicated in the table of online maintenance commands above. Re-enter the record in its correct form.

For example:

Existing record: E$PHOTOGRAPHIC;EMULSIONS$PHOTOGRAPHIC EMULSIONS should have a logic code of T.

Enter: T$PHOTOGRAPHIC;EMULSIONS$PHOTOGRAPHIC EMULSIONS

Any logic code entered will replace any previously entered logic code for that same key.

Changes to Posting Term Field of a Record. To change the posting term field in any way, use the appropriate update command as indicated in the table of online maintenance commands above. Re-enter the record in its correct form.

Existing record: E$MEDICINE;OO$MEDICINE should have an array term symbol following the posting term.

Enter: E$MEDICINE;OO$MEDICINE@

Any posting term(s) entered will replace any previously entered posting term(s) for that same key.

Symbols

Symbols should be used as needed. These have been described at some length in the subsections on Symbols under Phase One and Phase Two.

Logging On

Additions, deletions, or changes to any NLD record are done online. One user ID has been designated for NLD file maintenance; a second one is available for data entry only. Follow the log on procedure for whatever database management system used, and when the system prompts that it is ready, type in the desired command. NASAUPDT is used to correct the Phrase Matching file or DTICUPDT is used to maintain the DTIC/NASA Subject Switching file. The use of either of these commands creates a dataset of entries which will be used to update the master NLD file. Errors in this dataset are corrected online also.

Maintenance Commands

The NLD maintenance system provides a series of commands that are used to accomplish file maintenance activities. For each type of online activity, there are normally parallel commands for each data file. The corresponding command for the NASA Phrase Matching file usually begins with the letters "NASA". The command for the DTIC/NASA Subject Switching file begins with the letters "DTIC".

	Commands	
Functions	Phrase Matching File	DTIC/NASA Subject Switching Files
Provides NLD translations online	DTICACC	DTICACC
Creates backup tapes	NASABKUP	DTICBKUP
Creates continuation entries	NASACONT	DTICCONT
Displays file entries online	NASAFIND	DTICFIND
Loads transaction files	NASALOAD	DTICLOAD
Prints file, alpha by postings	NASANVRT	DTICNVRT
Prints file, alpha by key	NASAPRNT	DTICPRNT
Counts entries, sorted by logic code	NASATOT	DTICTOT
Unloads file for large-scale editing	NASAUNLD	DTICUNLD
Enters update transactions	NASAUPDT	DTICUPDT
Validates file entries	NASAVAL	DTICVAL
Creates authority files	VALSETUP	DTICVSAM
Displays records online	PRINT IDS	PRINT IDS

These commands are described more fully in the pages that follow.

DTICACC This command processes an input word or phrase through the Access

Routine. It provides, on the terminal screen, the full or partial translation of the input material, if any translation into NASA terms is available through the NLD. Otherwise, the program returns the message:

UNABLE TO IDENTIFY

The command can be used to see how the NLD will translate phrases or groups of terms that do not appear on a tape.

DTICBKUP
NASABKUP

These commands create backup tapes for the VSAM master files:

'NLD.SSDTIC.MASTER'and
'NLD.NASA.MASTER' respectively.

A backup is run after every file update so that the most current backup tape always reflects the current status of the VSAM file. Three successive backup tapes are retained in the tape library for each file. When a new backup tape is created, it replaces the oldest existing backup. An entry is recorded in the File Backup Log (shown in Figure 13) for each successful run of a backup command. The job printouts for the last three backup jobs are also kept for reference.

DTICCONT
NASACONT

These commands initiate jobs that read every entry in the data file, generate all required continuation entries and add them to the file, and when a new continuation entry has a key identical to an existing posted entry, adds a ;00 to the end of the key of the posted entry. DTICCONT and NASACONT are used only when an update is so large that coding and entering continuation entries individually is too time consuming to be economically feasible. The commands are executed after the update and at the end of the work day so that the programs can be run overnight.

DTICFIND
NASAFIND

These commands search the data files for a specified key, and print at the terminal ten sequential Lexical Dictionary records, beginning with the key requested, if it exists. If the requested key is not found, the program will locate the sequential position in which the key should occur and print the next ten records.

DTICLOAD
NASALOAD

These commands load additions and corrections from the dataset created by the UPDT command, that is LEX.DTIC.MOD or LEX.NASA.MOD, into the appropriate master file in order to update it. For DTIC the master file is 'NLD.SSDTIC.MASTER' and for NASA it is 'NLD.NASA.MASTER'. The Load command performs a number of edit checks on the transactions. Transactions passing the edit checks (good transactions) are loaded into the master file and are deleted from the work dataset. Transactions rejected by the edit checks are not loaded, but are rewritten to the appropriate LEX._____.MOD dataset for correction. Rejected transactions are listed on the printout with a notation of the error which caused

Figure 13

FILE BACKUP LOG

NLD.NASA.MASTER
BACKUPS

Date	Time	Job#	Initials

NLD.SSDTIC.MASTER
BACKUPS

Date	Time	Job#	Initials

the entry's rejection. The person doing the NLD maintenance corrects rejected transactions in the LEX.___.MOD dataset and then re-executes the Load command.

DTICNVRT
NASANVRT These commands print the master files sorted alphabetically by posting terms. In order to readily locate a particular posting term in the NLD, it is necessary to have a print of the file sorted alphabetically by posting term. Entries with multiple posting terms are listed once for each posting term. The Invert Print commands above sort and print the following files, respectively:

'NLD.SSDTIC.MASTER'
'NLD.NASA.MASTER'

Sample pages of NASANVRT and DTICNVRT are shown in Figures 14 and 15.

DTICPRNT
NASAPRNT These commands generate prints of the master files sorted alphabetically by keys. The files are, respectively:

'NLD.SSDTIC.MASTER'
'NLD.NASA.MASTER'

Sample pages of NASAPRNT and DTICPRNT are shown in Figures 16 and 17.

DTICTOT
NASATOT The Total command provides a count of the number of entries in the appropriate data file, broken down by logic code. Error messages are written for entries that do not have a valid logic code.

DTICUNLD
NASAUNLD The Unload command copies the entries in a VSAM data file into a series of smaller sequential files that can be edited online. These sequential files contain 3,000 entries each and have extra space allocated for additions. The job creates as many sequential files as are needed to hold all of the VSAM file entries. The files are named in this pattern:

LEX.SEQ1.DTIC	or	LEX.SEQ1.NASA
LEX.SEQ2.DTIC	or	LEX.SEQ2.NASA
LEX.SEQ3.DTIC	or	LEX.SEQ3.NASA
LEX.SEQ4.DTIC, etc.	or	LEX.SEQ4.NASA, etc.

The entries in these sequential files are in the following format:

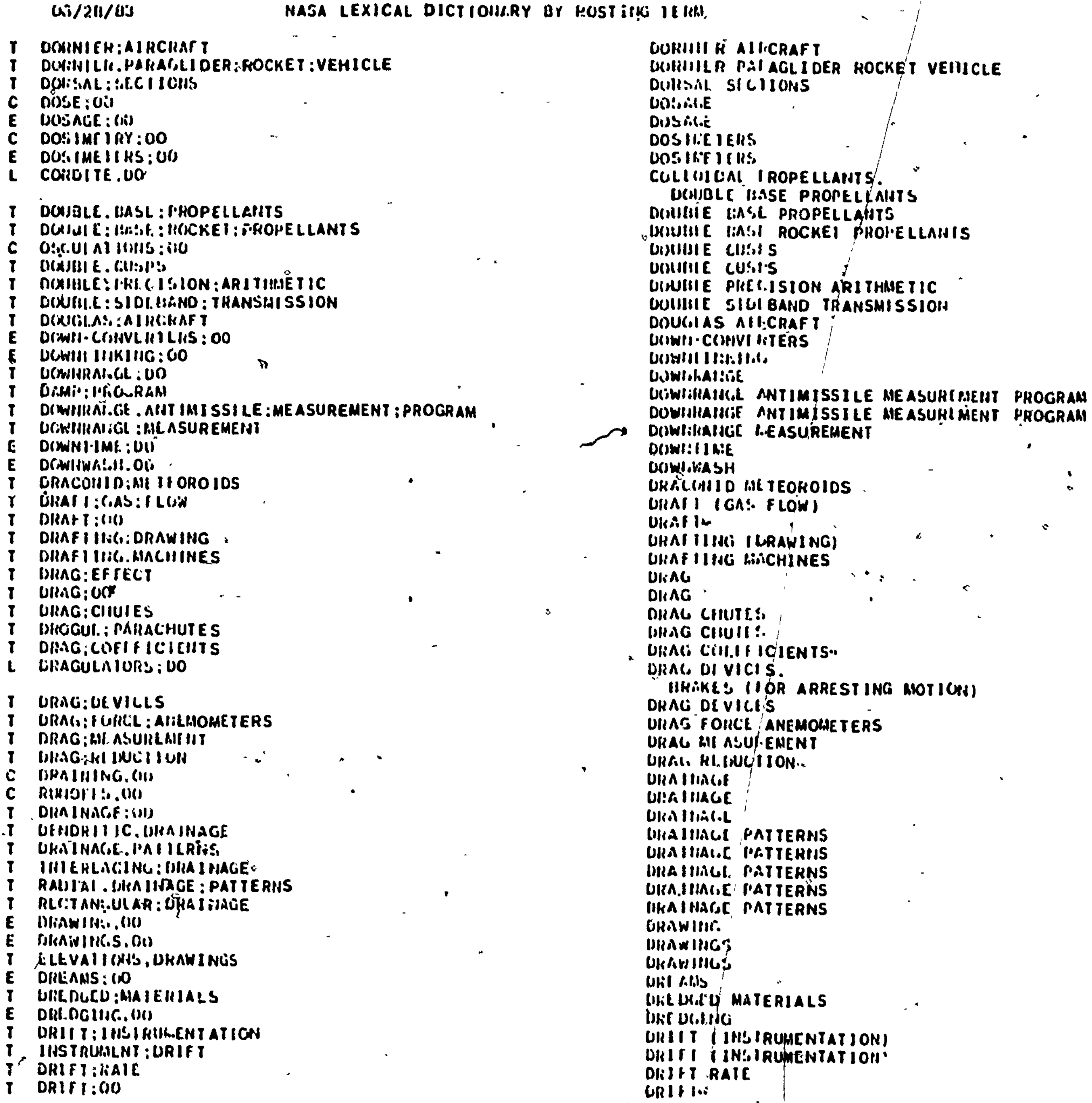
Columns 1 - 3	Logic Code
Columns 4 - 127	Key
Columns 130 - 400	Posting term

When editing is completed, the corrected files are loaded into the appropriate data file. Programmer assistance is required for this reload, so maintenance personnel are cautioned not to attempt this reload themselves.

```
06/28/83                  DTIC LEXICAL DICTIONARY BY POSTING TERM

EE REFINING:00                                  REFINING+
 C REFLECTIVITY:00                              REFLECTANCE
 T COEFFICIENTS:REFLECTION                      REFLECTANCE
EE REFLECTANCE:00                               REFLECTANCE
 T RADIATION:REFLECTION                         REFLECTED WAVES
 T REFLECTION:WAVES                             REFLECTED WAVES?.
                                                  WAVE REFLECTION?
 T REFLECTION:TELESCOPES                        REFLECTING TELESCOPES
 C INTERNAL REFLECTION:00                       REFLECTION
 L DIFFUSE REFLECTION:00                        REFLECTION.
                                                  DIFFUSE RADIATION
TE REFLECTION:00                                REFLECTION
 T NEBULAE:REFLECTION                           REFLECTION NEBULAE
EE REFLECTOMETERS:00                            REFLECTOMETERS
 L REACTOR REFLECTORS:00                        REFLECTORS.
                                                  NUCLEAR REACTORS
EE REFLECTORS:00                                REFLECTORS+
 L VASOMOTOR REFLEXES:00                        REFLEXES.
                                                  NERVOUS SYSTEM
EE REFLEXES:00                                  REFLEXES+
EE REFORESTATION:00                             REFORESTATION
 L ACOUSTIC REFRACTION:00                       ACOUSTIC ATTENUATION.
                                                  REFRACTED WAVES
 T REFRACTION:TELESCOPES                        REFRACTING TELESCOPES
EE REFRACTION:00                                REFRACTION
 C REFRACTIVE INDEX:00                          REFRACTIVITY
EE REFRACTOMETERS:00                            REFRACTOMETERS
 E REFRACTORY COATINGS:00                       REFRACTORY COATINGS
 L HEAT RESISTANT MATERIALS:00                  REFRACTORY MATERIALS.
                                                  THERMAL RESISTANCE
 T HIGH TEMPERATURE:MATERIALS                   REFRACTORY MATERIALS
 E REFRACTORY MATERIALS:00                      REFRACTORY MATERIALS+
 E REFRACTORY METAL ALLOYS:00                   REFRACTORY METAL ALLOYS
 E REFRACTORY METALS:00                         REFRACTORY METALS
 L HEAT RESISTANT METALS:00                     REFRACTORY METALS.
                                                  THERMAL RESISTANCE
EE REFRIGERANTS:00                              REFRIGERANTS
 L REFRIGERANT COMPRESSORS:00                   COMPRESSORS.
                                                  REFRIGERATING MACHINERY
 L REFRIGERANT CONDENSERS:00                    CONDENSERS (LIQUEFIERS).
                                                  REFRIGERATING MACHINERY
 T MACHINES:REFRIGERATION SYSTEMS               REFRIGERATING MACHINERY
 C REFRIGERATION SYSTEMS:00                     REFRIGERATORS
 I COLD STORAGE:CO                              ENERGY STORAGE?.
                                                  REFRIGERATORS?
 L CLOSED CIRCUIT REFUELING:00                  REFUELING.
                                                  AUTOMATIC CONTROL
 L REFUELING PUMPS:00                           REFUELING.
                                                  FUEL PUMPS
EE REFUELING:00                                 REFUELING
 C REGENERATION ELECTRONICS:00                  REGENERATION (ENGINEERING)
 E REGENERATION ENGINEERING:CO                  REGENERATION (ENGINEERING)
 T CYCLES:REGENERATION ENGINEERING              REGENERATION (ENGINEERING)
 C SUPERREGENERATION:00                         REGENERATION (ENGINEERING)>
 E REGENERATION PHYSIOLOGY:00                   REGENERATION (PHYSIOLOGY)
 E REGENERATIVE COOLING:00                      REGENERATIVE COOLING
 T FUEL CELLS:REGENERATION ENGINEERING          REGENERATIVE FUEL CELLS
 L GAS TURBINE REGENERATORS:00                  REGENERATORS.
```

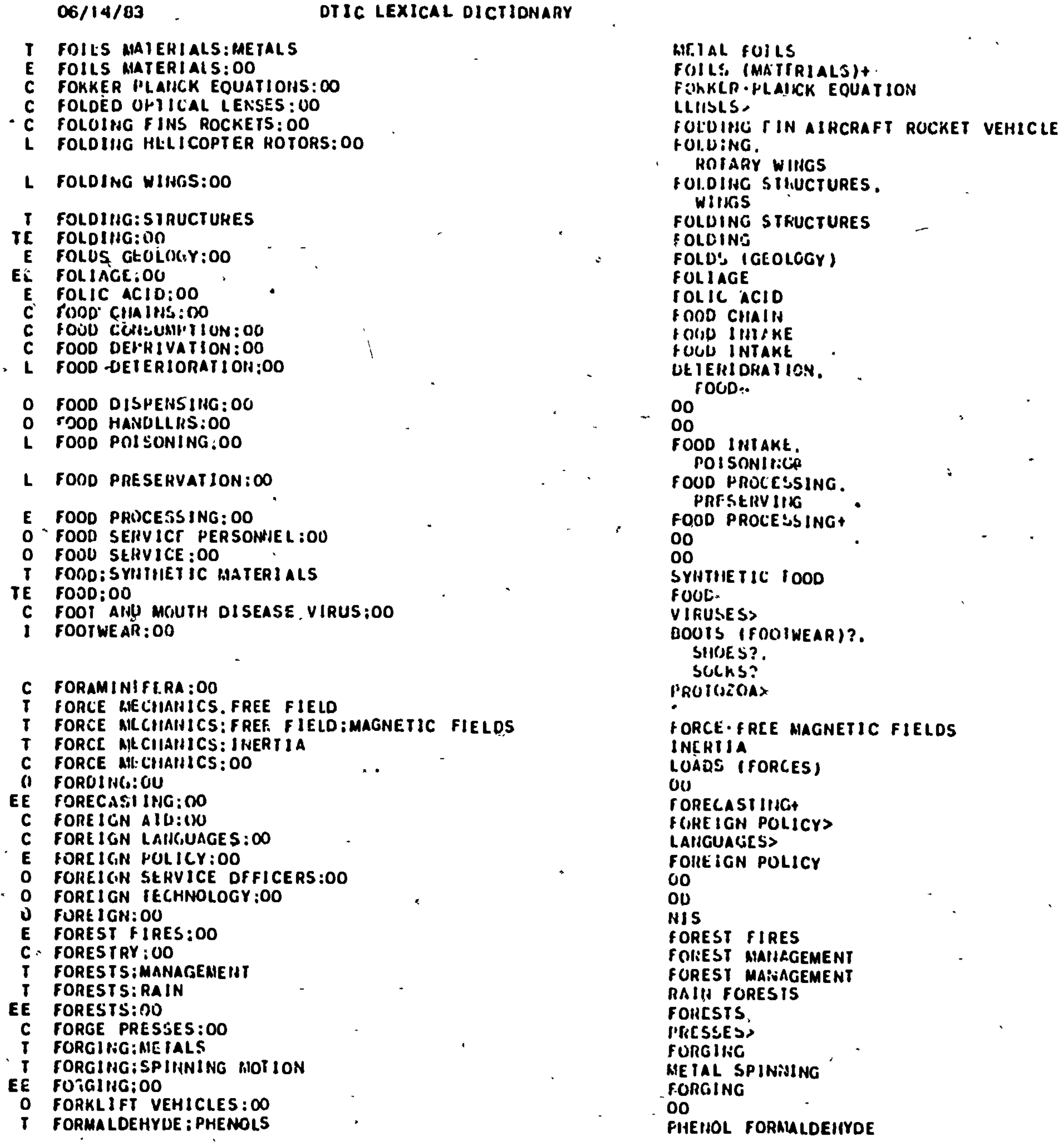
Figure 14 DTICNVRT Sample Output



```
   05/28/83            NASA LEXICAL DICTIONARY BY POSTING TERM

T  DORNIER:AIRCRAFT                                   DORNIER AIRCRAFT
T  DORNIER.PARAGLIDER:ROCKET:VEHICLE                  DORNIER PARAGLIDER ROCKET VEHICLE
T  DORSAL:SECTIONS                                    DORSAL SECTIONS
C  DOSE;00                                            DOSAGE
E  DOSAGE:00                                          DOSAGE
C  DOSIMETRY:00                                       DOSIMETERS
E  DOSIMETERS;00                                      DOSIMETERS
L  CORDITE.00                                         COLLOIDAL PROPELLANTS.
                                                        DOUBLE BASE PROPELLANTS
T  DOUBLE.BASE:PROPELLANTS                            DOUBLE BASE PROPELLANTS
T  DOUBLE:BASE:ROCKET:PROPELLANTS                     DOUBLE BASE ROCKET PROPELLANTS
C  OSCULATIONS:00                                     DOUBLE CUSPS
T  DOUBLE.CUSPS                                       DOUBLE CUSPS
T  DOUBLE:PRECISION:ARITHMETIC                        DOUBLE PRECISION ARITHMETIC
T  DOUBLE:SIDEBAND:TRANSMISSION                       DOUBLE SIDEBAND TRANSMISSION
T  DOUGLAS:AIRCRAFT                                   DOUGLAS AIRCRAFT
E  DOWN-CONVERTERS:00                                 DOWN-CONVERTERS
E  DOWNLINKING:00                                     DOWNLINKING
T  DOWNRANGE:00                                       DOWNRANGE
T  DAMP;PROGRAM                                       DOWNRANGE ANTIMISSILE MEASUREMENT PROGRAM
T  DOWNRANGE.ANTIMISSILE:MEASUREMENT;PROGRAM          DOWNRANGE ANTIMISSILE MEASUREMENT PROGRAM
T  DOWNRANGE:MEASUREMENT                              DOWNRANGE MEASUREMENT
E  DOWNTIME;00                                        DOWNTIME
E  DOWNWASH.00                                        DOWNWASH
T  DRACONID;METEOROIDS                                DRACONID METEOROIDS
T  DRAFT:GAS:FLOW                                     DRAFT (GAS FLOW)
T  DRAFT;00                                           DRAFT
T  DRAFTING:DRAWING                                   DRAFTING (DRAWING)
T  DRAFTING.MACHINES                                  DRAFTING MACHINES
T  DRAG:EFFECT                                        DRAG
T  DRAG;00                                            DRAG
T  DRAG;CHUTES                                        DRAG CHUTES
T  DROGUE;PARACHUTES                                  DRAG CHUTES
T  DRAG;COEFFICIENTS                                  DRAG COEFFICIENTS
L  DRAGULATORS;00                                     DRAG DEVICES.
                                                        BRAKES (FOR ARRESTING MOTION)
T  DRAG:DEVICES                                       DRAG DEVICES
T  DRAG;FORCE:ANEMOMETERS                             DRAG FORCE ANEMOMETERS
T  DRAG;MEASUREMENT                                   DRAG MEASUREMENT
T  DRAG;REDUCTION                                     DRAG REDUCTION
C  DRAINING.00                                        DRAINAGE
C  RUNOFFS.00                                         DRAINAGE
T  DRAINAGE:00                                        DRAINAGE
T  DENDRITIC.DRAINAGE                                 DRAINAGE PATTERNS
T  DRAINAGE.PATTERNS                                  DRAINAGE PATTERNS
T  INTERLACING:DRAINAGE                               DRAINAGE PATTERNS
T  RADIAL.DRAINAGE:PATTERNS                           DRAINAGE PATTERNS
T  RECTANGULAR:DRAINAGE                               DRAINAGE PATTERNS
E  DRAWING.00                                         DRAWING
E  DRAWINGS.00                                        DRAWINGS
T  ELEVATIONS,DRAWINGS                                DRAWINGS
E  DREAMS:00                                          DREAMS
T  DREDGED:MATERIALS                                  DREDGED MATERIALS
E  DREDGING.00                                        DREDGING
T  DRIFT;INSTRUMENTATION                              DRIFT (INSTRUMENTATION)
T  INSTRUMENT;DRIFT                                   DRIFT (INSTRUMENTATION)
T  DRIFT;RATE                                         DRIFT RATE
T  DRIFT:00                                           DRIFT
```

Figure 15 NASANVRT Sample Output

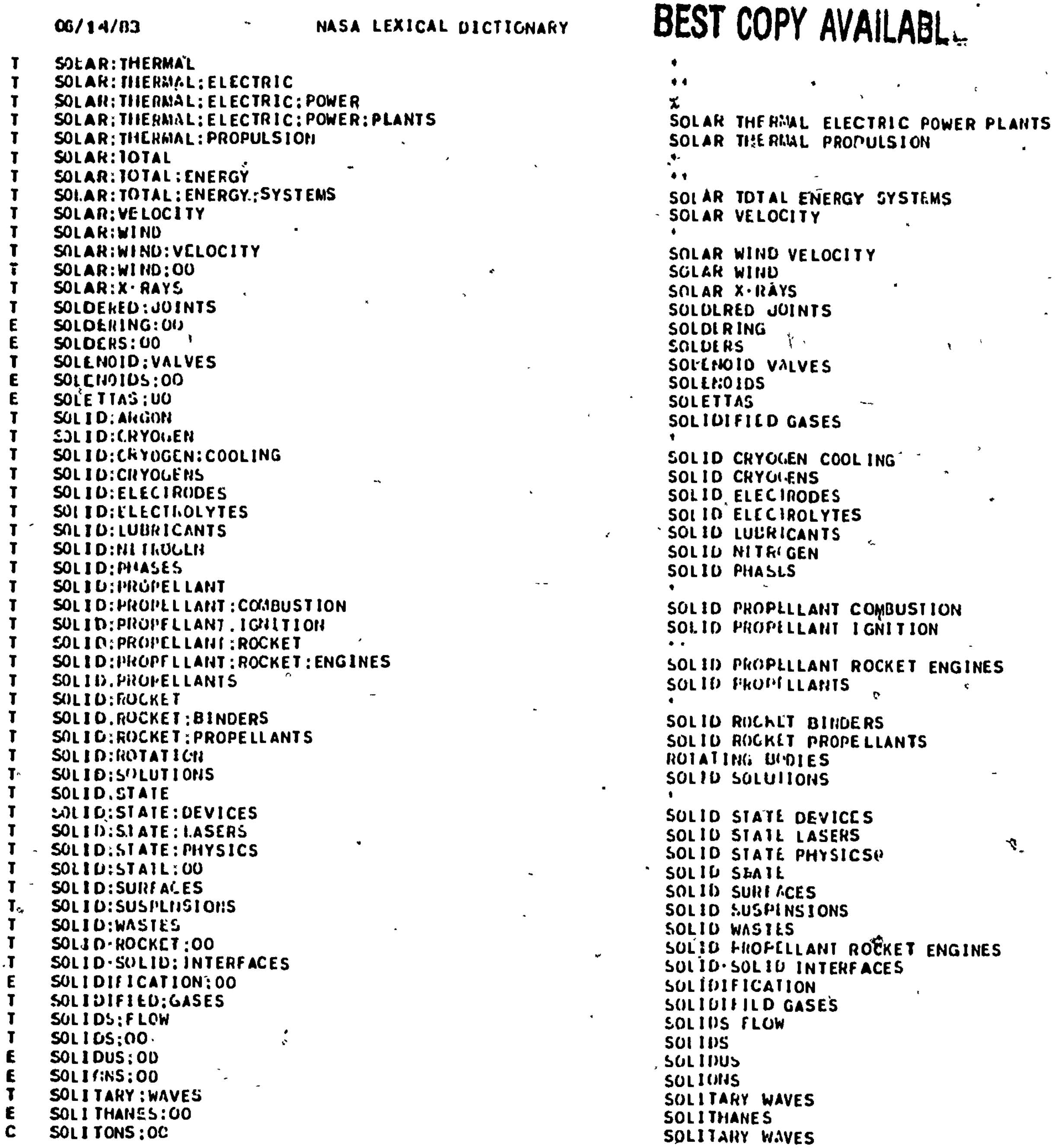


```
   06/14/83                DTIC LEXICAL DICTIONARY

T  FOILS MATERIALS:METALS                              METAL FOILS
E  FOILS MATERIALS:00                                  FOILS (MATERIALS)+
C  FOKKER PLANCK EQUATIONS:00                          FOKKER-PLANCK EQUATION
C  FOLDED OPTICAL LENSES:00                            LENSES>
C  FOLDING FINS ROCKETS:00                             FOLDING FIN AIRCRAFT ROCKET VEHICLE
L  FOLDING HELICOPTER ROTORS:00                        FOLDING,
                                                         ROTARY WINGS
L  FOLDING WINGS:00                                    FOLDING STRUCTURES,
                                                         WINGS
T  FOLDING:STRUCTURES                                  FOLDING STRUCTURES
TE FOLDING:00                                          FOLDING
E  FOLDS GEOLOGY:00                                    FOLDS (GEOLOGY)
EE FOLIAGE:00                                          FOLIAGE
E  FOLIC ACID:00                                       FOLIC ACID
C  FOOD CHAINS:00                                      FOOD CHAIN
C  FOOD CONSUMPTION:00                                 FOOD INTAKE
C  FOOD DEPRIVATION:00                                 FOOD INTAKE
L  FOOD DETERIORATION:00                               DETERIORATION,
                                                         FOOD<
O  FOOD DISPENSING:00                                  00
O  FOOD HANDLERS:00                                    00
L  FOOD POISONING:00                                   FOOD INTAKE,
                                                         POISONING@
L  FOOD PRESERVATION:00                                FOOD PROCESSING,
                                                         PRESERVING
E  FOOD PROCESSING:00                                  FOOD PROCESSING+
O  FOOD SERVICE PERSONNEL:00                           00
O  FOOD SERVICE:00                                     00
T  FOOD:SYNTHETIC MATERIALS                            SYNTHETIC FOOD
TE FOOD:00                                             FOOD<
C  FOOT AND MOUTH DISEASE VIRUS:00                     VIRUSES>
I  FOOTWEAR:00                                         BOOTS (FOOTWEAR)?,
                                                         SHOES?,
                                                         SOCKS?
C  FORAMINIFERA:00                                     PROTOZOA>
T  FORCE MECHANICS.FREE FIELD                          
T  FORCE MECHANICS:FREE FIELD:MAGNETIC FIELDS          FORCE-FREE MAGNETIC FIELDS
T  FORCE MECHANICS:INERTIA                             INERTIA
C  FORCE MECHANICS:00                                  LOADS (FORCES)
O  FORDING:00                                          00
EE FORECASTING:00                                      FORECASTING+
C  FOREIGN AID:00                                      FOREIGN POLICY>
C  FOREIGN LANGUAGES:00                                LANGUAGES>
E  FOREIGN POLICY:00                                   FOREIGN POLICY
O  FOREIGN SERVICE OFFICERS:00                         00
O  FOREIGN TECHNOLOGY:00                               00
O  FOREIGN:00                                          NIS
E  FOREST FIRES:00                                     FOREST FIRES
C  FORESTRY:00                                         FOREST MANAGEMENT
T  FORESTS:MANAGEMENT                                  FOREST MANAGEMENT
T  FORESTS:RAIN                                        RAIN FORESTS
EE FORESTS:00                                          FORESTS,
C  FORGE PRESSES:00                                    PRESSES>
T  FORGING:METALS                                      FORGING
T  FORGING:SPINNING MOTION                             METAL SPINNING
EE FORGING:00                                          FORGING
O  FORKLIFT VEHICLES:00                                00
T  FORMALDEHYDE:PHENOLS                                PHENOL FORMALDEHYDE
```

Figure 16 DTICPRNT Sample Output

BEST COPY AVAILABLE



```
      06/14/83                NASA LEXICAL DICTIONARY

T     SOLAR:THERMAL                                        *
T     SOLAR:THERMAL;ELECTRIC                               **
T     SOLAR:THERMAL;ELECTRIC;POWER                         X
T     SOLAR;THERMAL;ELECTRIC;POWER;PLANTS                  SOLAR THERMAL ELECTRIC POWER PLANTS
T     SOLAR;THERMAL:PROPULSION                             SOLAR THERMAL PROPULSION
T     SOLAR:TOTAL                                          *
T     SOLAR:TOTAL;ENERGY                                   **
T     SOLAR:TOTAL;ENERGY.;SYSTEMS                          SOLAR TOTAL ENERGY SYSTEMS
T     SOLAR;VELOCITY                                       SOLAR VELOCITY
T     SOLAR:WIND                                           *
T     SOLAR:WIND:VELOCITY                                  SOLAR WIND VELOCITY
T     SOLAR:WIND:00                                        SOLAR WIND
T     SOLAR:X-RAYS                                         SOLAR X-RAYS
T     SOLDERED:JOINTS                                      SOLDERED JOINTS
E     SOLDERING:00                                         SOLDERING
E     SOLDERS:00                                           SOLDERS
T     SOLENOID;VALVES                                      SOLENOID VALVES
E     SOLENOIDS:00                                         SOLENOIDS
E     SOLETTAS;00                                          SOLETTAS
T     SOLID:ARGON                                          SOLIDIFIED GASES
T     SOLID:CRYOGEN                                        *
T     SOLID:CRYOGEN:COOLING                                SOLID CRYOGEN COOLING
T     SOLID:CRYOGENS                                       SOLID CRYOGENS
T     SOLID:ELECIRODES                                     SOLID ELECIRODES
T     SOLID:ELECTROLYTES                                   SOLID ELECIROLYTES
T     SOLID:LUBRICANTS                                     SOLID LUBRICANTS
T     SOLID:NITROGEN                                       SOLID NITROGEN
T     SOLID:PHASES                                         SOLID PHASES
T     SOLID:PROPELLANT                                     *
T     SOLID:PROPELLANT:COMBUSTION                          SOLID PROPELLANT COMBUSTION
T     SOLID:PROPELLANT.IGNITION                            SOLID PROPELLANT IGNITION
T     SOLID:PROPELLANT;ROCKET                              **
T     SOLID:PROPELLANT;ROCKET:ENGINES                      SOLID PROPELLANT ROCKET ENGINES
T     SOLID.PROPELLANTS                                    SOLID PROPELLANTS
T     SOLID:ROCKET                                         *
T     SOLID.ROCKET;BINDERS                                 SOLID ROCKET BINDERS
T     SOLID:ROCKET;PROPELLANTS                             SOLID ROCKET PROPELLANTS
T     SOLID:ROTATION                                       ROTATING BODIES
T     SOLID;SOLUTIONS                                      SOLID SOLUTIONS
T     SOLID.STATE                                          *
T     SOLID:STATE:DEVICES                                  SOLID STATE DEVICES
T     SOLID:STATE:LASERS                                   SOLID STATE LASERS
T     SOLID:STATE:PHYSICS                                  SOLID STATE PHYSICS
T     SOLID:STATE;00                                       SOLID STATE
T     SOLID:SURFACES                                       SOLID SURFACES
T     SOLID:SUSPENSIONS                                    SOLID SUSPENSIONS
T     SOLID:WASTES                                         SOLID WASTES
T     SOLID-ROCKET;00                                      SOLID PROPELLANT ROCKET ENGINES
T     SOLID-SOLID;INTERFACES                               SOLID-SOLID INTERFACES
E     SOLIDIFICATION;00                                    SOLIDIFICATION
T     SOLIDIFIED;GASES                                     SOLIDIFIED GASES
T     SOLIDS;FLOW                                          SOLIDS FLOW
T     SOLIDS;00                                            SOLIDS
E     SOLIDUS;00                                           SOLIDUS
E     SOLIONS;00                                           SOLIONS
T     SOLITARY;WAVES                                       SOLITARY WAVES
E     SOLITHANES:00                                        SOLITHANES
C     SOLITONS;00                                          SOLITARY WAVES
```

Figure 17 NASAPRNT Sample Output

DTICUPDT
NASAUPDT

These commands utilize work datasets for compiling changes including additions or deletions, which are intended for the master files. These work datasets, respectively, LEX.DTIC.MOD and LEX.NASA.MOD can be edited online.

As a transaction is entered at the terminal, a series of edit checks take place. Transactions passing the edit checks (good transactions) are loaded into a temporary work dataset.

Rejected transactions generate error messages online.

The error messages that are returned interactively by the system follow:

INVALID CHARACTERS
The transaction contains characters other than the following valid set: A-Z, 0-9, +, ?, > , &, ', $, (,), ;, ., %, *, /, @, -, or blank.

INVALID CHARACTER IN LOGIC
The transaction contains characters other than the following valid set in the logic code position (or before the first dollar sign):

DEL, C, E, I, L, T, 0 (zero), or blank.

LOGIC CODE TOO LONG
More than three characters or blanks appear before the first $ in the transaction.

LOGIC CODE ALL BLANKS
Three blanks appear before the first $ in the transaction.

NO POSTING TERM
Nothing appears following the second $ in the transaction.

TOO MANY $'s
More than two $'s appear in the transaction.

INVALID FORMAT
The transaction does not conform to one of the formats:

```
     Logic code$Element;Element$Posting term
     Logic code$Element;00$Posting term
     DEL$(Key of record to be deleted)
```

COMMAND NOT FOUND
If this error message appears following any command, check to make sure that you are logged on under the maintenance ID and that you have spelled the command correctly. If the problem persists or if other error messages appear, check with the application programmer assigned to the Lexical Dictionary.

An error that generates one of these messages must be corrected in the manner indicated before the system will accept the entry.

When the session is ended by entering /*, the system performs a second series of edit checks on the transactions held in the temporary file, and loads the transactions into the LEX.____.MOD file appropriate to the command. These files are respectively, LEX.DTIC.MOD and LEX.NASA.MOD. Rejected entries are listed on the printout under the heading TRANSACTIONS IN ERROR and must be researched, reformatted if necessary, and entered correctly using the online edit capability of the data base management system.

DTICVAL
NASAVAL These commands initiate comparisons between data files and authority files. DTICVAL compares the DTIC/NASA Subject Switching file entries with the NASA and DTIC thesauri authority files. DTICVAL checks:

- Every DTIC term appearing in the key field against the DTIC Thesaurus authority file. If a term in the NLD key does not appear in the authority file, an error message is generated.

- Every NASA term appearing in the posting term field against the NASA Thesaurus authority file. If an NLD posting term does not appear in the authority file, an error message is generated.

- Every posting term in the DTIC Thesaurus authority file against the NLD keys. If there is no key in the NLD for the Thesaurus posting term, an error message is generated.

These error messages highlight the additions, modifications, and deletions required in the DTIC/NASA Subject Switching file.

NASAVAL initiates a set of comparisons between the entries in the NASA Phrase Matching file and the NASA Thesaurus authority files. NASAVAL checks:

- Every term appearing in the posting term field against the NASA Thesaurus authority file. If an NLD posting term does not appear in the Thesaurus authority file, an error message is generated.

- Every posting term and Use reference appearing in the NASA Thesaurus authority file against the NLD file keys. Each of these terms should appear as a key in the NLD file, and an error message is generated if it does not.

- Every posting term in the NASA Thesaurus authority file against the NLD file posting terms. If a Thesaurus posting term does not also appear as an NLD posting term, an error message is generated.

These error messages highlight the additions, modifications, and deletions required in the Phrase Matching file.

DTICVSAM This command is used to create a DTIC Thesaurus authority file from LEX.POSTTERM.DTIC, a list of posting terms from DTIC. LEX.POSTTERM.DTIC is a sequential file created from DTIC's Thesaurus tape and so far updated manually by NLD maintenance personnel using online editing capabilities. Each time LEX.POSTTERM.DTIC is updated, a new VSAM authority file must be created with the DTICVSAM command.

VALSETUP This command creates two authority files for NASA Thesaurus terms from the online Thesaurus files:

- A sequential file of NASA posting terms and Use references. This file is used by the validation routine to check that there is an entry in the Phrase Matching file for every NASA Posting term and Use reference.

- A VSAM file of NASA posting terms only: 'NLD.THES.TERMS'

The VSAM file is used by NASAVAL to verify that all posting terms appearing in the posting term field of existing entries in the Phrase Matching file are valid NASA posting terms, and by DTICVAL for the same purpose in the Subject Switching file. NASAUPDT, NASALOAD, DTICUPDT and DTICLOAD use the NASA VSAM authority file for validating new transactions being added to the data files.

As the VSAM file is being created, each term is checked against the Phrase Matching file ('NLD.NASA.MASTER') to determine if it should be marked as an array term and to add the @ to the term if required. To look at a term in the NASA VSAM authority file 'NLD.THES.TERMS', use the PRINT IDS command.

PRINT IDS This command allows an online look at any VSAM file record and at a user-specified number of additional sequential records. DTICFIND and NASAFIND are shortcuts for displaying records in master files 'NLD.SSDTIC.MASTER' and 'NLD.NASA.MASTER', respectively. However, to see a record in other VSAM files, for example the NASA file 'NLD.THES.TERMS', one must use PRINT IDS.

Printout Review

Following execution of a command, any printout that has been generated is examined by NLD personnel. This is to see:

- Whether or not the job has run satisfactorily.

- Whether or not there are any errors that must be corrected.

The Facility NLD Maintenance Manual lists step-by-step instructions for recognizing and correcting errors from each listing.

In general, errors listed in printouts generated by any of the commands are listed under a heading that implies what the problem is and how to fix it.

A sample of error messages listed on printouts generated by NLD maintenance commands follows:

KEY (unmatched element) OF (key of rejected transaction) IS NOT FOUND

The transaction has been rejected because the specified element of the key does not match any entry in the input posting term authority file. The non-match may be the result of:

1. misspelling in the transaction,
2. failure to separate multiple elements of the key with semicolons, or
3. an error in the input posting term authority file.

If the error is of types 1 or 2, correct the error in the appropriate LEX.___.MOD. file before re-executing the load. If the error is of type 3, use the online edit capability of the data base management system to correct the error in the corresponding LEX.POSTTERM.___ file, and execute the ___VSAM command to recreate a corrected input posting term authority file. Then the Load command may be re-executed.

POSTING TERM (unmatched posting term) OF (entire posting term field of rejected transaction) IS NOT FOUND

The transaction has been rejected because the specified element of the transaction's posting term does not match any entry in the NASA posting term authority file. The non-match may be the result of:

1. misspelling in the transaction,
2. leaving out a required "@",
3. including an incorrect "@",
4. failure to separate multiple posting terms with commas, or
5. an error in the NASA posting term authority file.

If the error is of types 1 through 4, correct the error in the appropriate LEX.___.MOD file before re-executing the load. If the error is in the authority file, (type 5), notify the lexicographer. The error must be corrected in the RECON online Thesaurus and a new NASA posting term authority file created by executing VALSETUP before the Load command can be re-executed.

INVALID LOGIC CODE IN RECORD (rejected transaction) or NO LOGIC CODE IN RECORD (rejected transaction).
The transaction has been rejected because the logic code was missing or was incorrect. Correct the logic code in the appropriate LEX.___.MOD file and re-execute the Load command.

ELEMENTS IN KEY NOT IN ALPHA ORDER (key of rejected transaction)
In a Subject Switching file, the elements of the key must be in alphabetical order. Correct the key in the appropriate LEX.___.MOD file before re-executing the Load command.

Errors in a LEX.___.MOD file are corrected using the online edit capabilities of the data base management system. When the file is corrected, it is loaded into the appropriate data file by entering DTICLOAD, or NASALOAD.

In a DTICVAL printout, there may be a page headed UNMATCHED KEYS. This contains error message for all Lexical Dictionary entries that contain in the key a term that is not matched on the DTIC Thesaurus authority file. The job prints out both the erroneous term and the entire record containing the erroneous term, in the following format:

```
KEY NOT FOUND    =    CLARK
OF RECORD        =    T CLARK;DUKE SUPER PROGRAMMER
```

If the Lexical Dictionary key is erroneous, delete the record and add a corrected entry to the file if necessary. In some cases, these errors are due to errors in the DTIC Thesaurus authority file. If an authority file error is located, fix the error in the LEX.POSTTERM.DTIC file using the online edit capabilities of the data base management system and then execute DTICVSAM to create a new DTIC Thesaurus authority file.

The page headed UNMATCHED POSTING TERMS contains an error message for each Lexical Dictionary entry containing a posting term which is not matched in the NASA Thesaurus authority file. The job prints out the errors in the following format:

Logic Code	Key	Posting Term	Error Message

For example:

T	Clark;Duke	Super Programmer	Key not found-Super Programmer Next record is Supercavitating Flow

In this example, the program could not match the posting term in the authority file at all. (This is a test record added to the file by said programmer.) As an aid to the analyst, the programmer prints out the record appearing in the authority file following the place where the unmatched posting term should have appeared. When the error is a typographical error in the Lexical Dictionary key, the "next record" is frequently the correct spelling of the NASA Posting term.

```
C   Power Equipment;00     Electric Equipment@     Key not found-
                                                   Electric Equipment@
                                                   Found Electric
                                                   Equipment without @
```

In this example, the program could not exactly match the Lexical Dictionary posting term Electric Equipment@ in the NASA Thesaurus authority file, but did find Electric Equipment (the same term without the "@"). Check the term in question in the NASA Thesaurus.

If the term is not an array term, correct the entry in the Lexical Dictionary by deleting the @ from the posting term in the entry using the ___UPDT procedure. See Changes to an Existing Record subsection of "Maintenance Functions."

If the term is an array term, the error (a missing "@") is in the NASA Thesaurus authority file. The VALSETUP program which creates the NASA Thesaurus authority file adds the "@" to terms if an "@" appears following that term in the NASA Phrase Matching file. Therefore, if this type of error is located in the NASA Thesaurus authority file, it means that there is an error in the NASA Phrase Matching file for that term. The Phrase Matching entry is located in the 'NLD.NASA.MASTER' file using NASAFIND.

Using the NASAUPDT procedure, the @ is added to the posting term, and VALSETUP is executed to create a correct NASA Thesaurus authority file.

```
T   Flight;Stresses    Flight Stress     Key not found-Flight stress
                                         Found Flight Stress
                                         with @
```

In this example, the program could not exactly match the NLD key Flight Stress in the NASA Thesaurus authority file, but did find Flight Stress@. Check the term in question in the NASA Thesaurus.

If the term is not an Array term, the error is in the NASA Thesaurus authority file. See the explanation in the preceding example for how to correct this type of authority file error.

If the term is an Array term, the error is in the DTIC/NASA Subject Switching file and can be corrected by adding the "@" to the posting term in the NLD entry.

A page headed UNMATCHED DTIC TERMS lists an error message for every term which appears in the DTIC Thesaurus authority file for which no entry appears in the DTIC/NASA Subject Switching file. The errors appear in the following format:

TERM NOT FOUND = (unmatched term)

Check the unmatched term in the most recent DTIC Thesaurus supplement.

If the term is a valid DTIC term: determine the correct NASA translation, create a new Subject Switching file record for the term, and add the record to the Lexical Dictionary file using DTICUPDT.

If the unmatched term is not a valid DTIC term, the error is in the authority file. Correct or delete the incorrect entry in the manually maintained DTIC Thesaurus file LEX.POSTTERM.DTIC. Execute the DTICVSAM command to create a corrected DTIC Thesaurus authority file.

Other error messages may report on the status of:

NASA THESAURUS VALIDATION

THESAURUS TERMS-NO NLD KEY

THESAURUS TERMS - NO NLD POSTING TERM

READING FOR ATSIGN AND KEY NOT FOUND (term)

If there are problems that the NLD maintenance personnel cannot correct, the application programmer assigned to the NLD must be consulted.

RESULTS AND CONCLUSIONS

Applications

A lexical dictionary can be used for any application that requires translation of input phrases to phrases in a target vocabulary. As of January 1, 1984, NASA was using the NLD for three applications:

- building Subject Switching capabilities
- processing DTIC (TAB) tapes, and
- processing DOE (EDB) tapes.

Each of these applications uses the NLD system in a different way. Building Subject Switching capabilities uses the Phrase Matching mode and accepts all matches, complete or partial, from the NLD system. DTIC TAB tape processing uses both the Phrase Matching and Subject Switching modes, and accepts only complete matches from the NLD System. DOE EDB tape processing used only the Phrase Matching mode (until the DOE/NASA Subject Switching file became operational) and accepted only complete matches from the NLD system.

As this is being written, two other applications are in the programming stage:

- Using the Phrase Matching file to process Library of Congress MARC records and accepting complete or partial translations.
- Using the Phrase Matching file to process natural language phrases automatically extracted from abstracts or other text and accepting complete or partial translations.

Benefits

Benefits obtained from the use of the NLD were measured with the least possible disruption to the indexing process. The hypothesis upon which the NLD was authorized was that the NLD would increase the indexers' productivity and reuse the indexing already done by DTIC. It was intended that the quality of the indexing would remain high. The following analyses shows that our samples were adequate, that the results are significant, and that we have proven our hypothesis.

Evaluation Methods. The evaluation of the NLD was based on a comparison of the preliminary subject analysis study done for December 1982 through March 1983 with a post-implementation study done for December 1983 through March 1984.

Study number 1 included confidential interviews with each indexer; all were conducted by the same interviewer to ensure consistency. In addition, a sample of 100 documents was selected from a single DTIC TAB tape, and the results of Subject Switching for these documents were analyzed. These documents were taken from all categories represented on the tape. Since there were fewer than 100 categories, multiple selections were made from some categories in approximate proportion to the number of documents assigned to the more populous categories.

Study number 2 utilized a questionnaire because there was concern that observed time studies would be intrusive and slow production. Indexers, without consulting with one another, filled out their questionnaires simultaneously. In addition, a representative sample of 150 DTIC documents, drawn over a three-month period, was analyzed.

Comparisons. See Figure 18. Although study 1 had a sample of 100 documents, two of the DTIC posting term values were discarded as being too deviant leaving a sample size for DTIC of 98. The following table shows some of the comparisons made.

		Study 1 (Pre)		Study 2 (Post)	
		DTIC	NASA	DTIC	NASA
Documents in sample	N	98	100	250	250
Mean of term assigned	$\overline{X}$	14.32	9.59	13.09	12.60
Standard deviation where Σ = the sum and X = the deviation from the mean	$S = \sqrt{\frac{\Sigma X^2}{N-1}}$	4.88	2.03	4.73	2.05
Variance	$S^2 = \frac{\Sigma X^2}{N-1}$	23.81	4.11	22.35	4.19
Standard error of mean	$S_{\overline{X}} = \frac{S}{\sqrt{N}}$	.49	.20	.30	.13

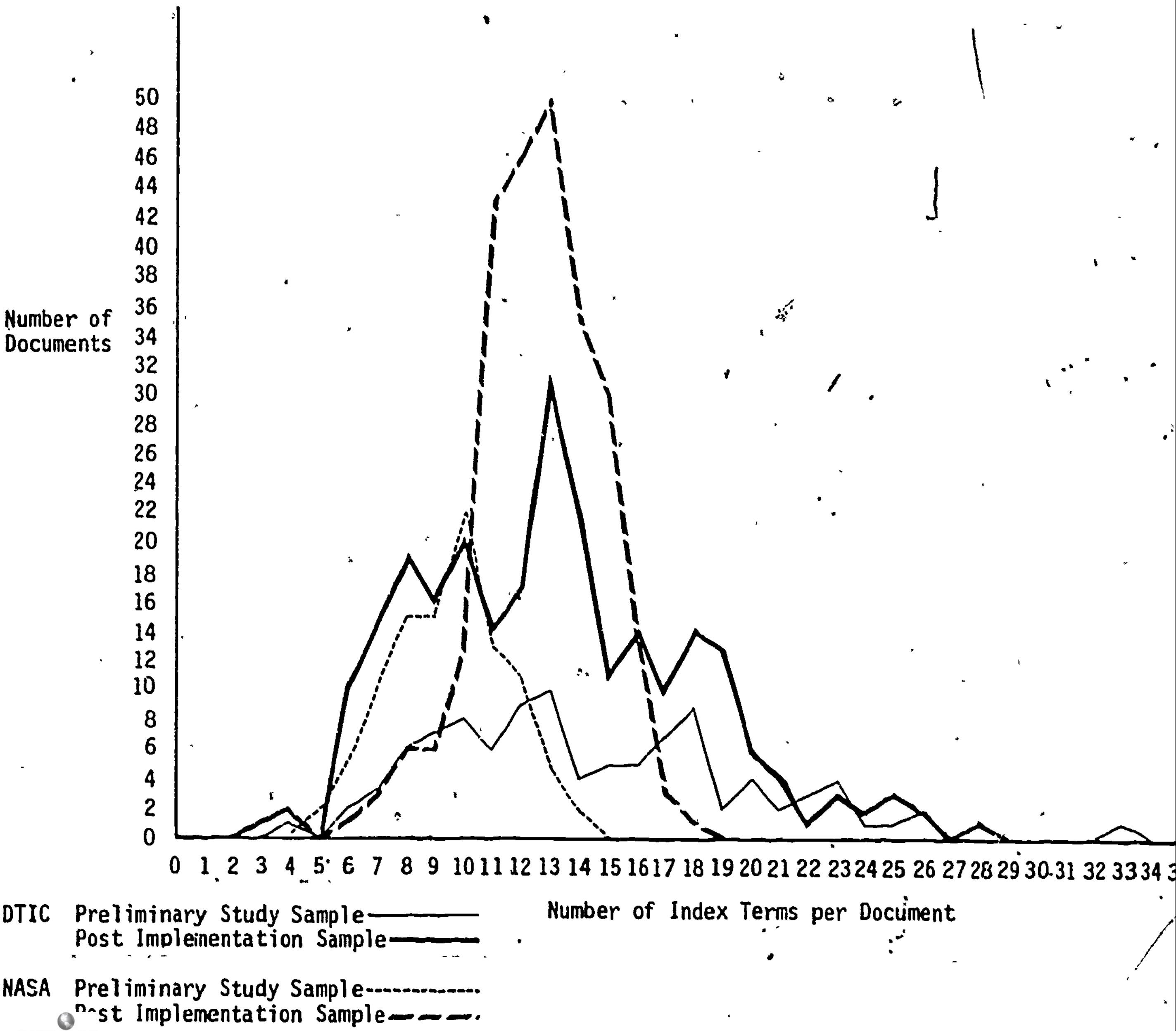
The standard error is small. Since the "t" test (used to ascertain the deviation of the estimated mean from the mean of the population) gives us a value which is off the t chart but indicates a better than 99% confidence level, we conclude that our samples and the results of our comparative study (shown above) are valid for the entire population.

It is interesting to note that before using the NLD, there was considerable difference in the average number of index terms assigned by the two agencies: 14.32 to 9.59. Study 2 shows that the averages are now very close: 13.09 to 12.60.

Access Points and Productivity. The increase in the number of NASA index terms assigned to a document, as indicated in the above table, not only signals increased productivity, but also increases the number of access points to a document. Evaluations of document retrieval have

Figure 18

A Comparison of DTIC and NASA Indexing



indicated that these changes have not affected retrieval adversely. The pertinency level of retrieval has remained high throughout the introduction and use of the NLD.

Time Savings. Two concerns in the field of library and information science are the ever-growing amounts of material to be classified, stored, and disseminated, and a constant need to do more work for less or the same amount of money. Information scientists are looking for ways to get information to the user more quickly. We feel that the NLD is making a contribution in this area.

82% of the indexers reported that having index terms provided by the NLD makes indexing DTIC documents faster. The remaining indexers indicated that having the suggested NLD terms has no effect on their speed.

Indexers were asked to estimate the time saved by having NLD terms. The average was 5.4 minutes per document. See Figure 19. Indexers then were asked to estimate the time required to index a DTIC document with NLD terms provided. The average of these estimates was 10 minutes. When this figure was compared with the study 1 (pre-NLD) average indexing time of 13 minutes, a 3 minute difference was noted. The predicted savings per document was 2 to 3 minutes. Based on the indexers' estimates, the intended goal has been reached and may have been exceeded. This time savings obviously speeds up the document turnaround time and can increase the timeliness of the product.

Changes in Work Emphasis. As an indexer tool, the NLD has relieved the indexers of having to look up many terms in the thesaurus. The correct form is presented for use or for deletion. In place of this rather mechanical task, indexers are asked to watch for coordinations of DTIC terms that should be translated to single NASA terms. This process should result naturally from the review of the indexing terms presented by the NLD printout, and the change in emphasis can provide more challenge to the indexer's job than just looking up correct forms of terms.

Shared Resources. Reindexing work that already has been done at taxpayer expense is wasteful of government resources. The original and primary purpose of the NLD was to utilize indexing done by other agencies. The sharing of indexing with DTIC also has brought about sharing of some programming and improved quality in the thesauri and lexical dictionaries of both DTIC and NASA.

Stepping Stone. The Lexical Dictionary has been found to be a stepping stone to other endeavors. Its Phrase Matching capabilities are being expanded and will soon be used to add NASA terms to MARC records. The NLD is also a way of approach to machine-aided indexing of abstracts or other text.

Problems

Different Indexing Philosophies. Indexing philosophies differ from agency to agency. This difference must be addressed in translating

Estimate of Time Saved
(7 indexers replied)

3 minutes saved per document
2
10
4
3
1
15
38 ÷ 7 = 5.4 minutes average

Estimate without the outliers

3
2
4
3
1
13 ÷ 5 = 2.6 minutes

Pooled estimate of time saved

5.4
2.6
8.0 ÷ 2 = 4 minutes per document

Estimate of time to index
(10 indexers replied)
6 minutes per document
8
7.5
20
4
15
7
10
12.5
10
100 ÷ 10 = 10 minutes per document

Figure 19

concepts. For example, one agency may index to "Ablative nose cones" and another to "Ablation" and "Nose cones". The first is precoordinated, i.e. the two concepts are joined in the index term. The second is post-coordinated, i.e. coordinated at the time of retrieval.

Some agencies include in their indexing any broader terms that appear in the hierarchy of the term used. For example, if the term used 's "Hafnium-alpha" the indexer or the system would assign all of the following terms: Hafnium, Refractory metals, Metals, and Elements. Another system would assign only the most specific term that applies. If the document were on the subject of Hafnium-alpha, they would index to that only, or if that term were not available, to Hafnium.

These differences must be taken into account in setting up a lexical dictionary system.

Chemical Compounds, Complexes, Metal Alloys, etc.. If you can't match a chemical compound or complex or an alloy term exactly, you may expect trouble with translations and retrieval. Coordinations of terms in this area of knowledge are likely to produce unwanted citations in retrieval. It is important to know how an indexer is instructed to handle these concepts.

Semantics and Scope. The term "Performance tests" in DTIC's environment has no restrictions in meaning. In NASA's environment, this term applies only to operating equipment. Another agency has the term Blowouts and defines it as "high pressure...ejection of water, gas, or oil from a borehole"; in NASA's Thesaurus, Blowouts is related to Tires and Fatigue life. These are homonyms, two terms that match character for character, but convey different concepts. Every term must be examined as to its scope and meaning in both the input and target environments.

Recommendations

Automation. Automate the initial entries, the continuation entries, and use online editing. It will keep the manhours needed for data entry and for error correction to a minimum.

Indexing Policies and Vocabularies. Become familiar with the indexing policies and vocabularies of both organizations: the one contributing, and your own, the target.

SUMMARY

Problem and Proposed Solution

Because the NASA STI Facility and the Defense Technical Information Center have overlapping interests, they share information. Twenty percent of the NASA data base was previously indexed by DTIC. Most of the documents received from DTIC are on microfiche accompanied by a magnetic tape that provides DTIC's cataloging, abstracting, and indexing in machine-readable form. Management proposed that the Facility automatically translate DTIC's posting terms to NASA's terms in order:

- To avoid the reindexing that was necessary to adapt the information to the NASA system and
- To save indexing time.

Construction of the NLD

The NASA Lexical Dictionary was constructed in four phases.

Phrase Matching. Phase One centered on constructing a file consisting of entries for every posting term and Use reference in the NASA Thesaurus, as well as additional Use references constructed specifically for the NLD System. Most of the programs were written in this phase also. The Phrase Matching mode attempts to find word-by-word matches between any input phrases and NASA terms or Use references. Matches may be complete or partial. For example:

Input-Any Terms	Output-NASA Terms
Gold	Gold
Gold plate	Gold coatings
Gold plated	Gold coatings
Gold plated chassis	Gold coatings, chassis
Gold-plated chassis	Chassis (because gold-plated, with a hyphen, has not been added to the file yet)

Subject Switching Individual Terms. Phase Two consisted of the construction of a translation table between the DTIC Thesaurus terms and NASA's. Entries in the file pair each DTIC term with one or more NASA terms that best express the same concept. For example:

Input-DTIC Terms	Output-NASA Terms
Anti Fogging Agents	Fog Dispersal

Input-DTIC Terms	Output-NASA Terms
Antioxidants	Antioxidants
Apogee	Apogees
Approach	Approach+
Architects	Architecture, Personnel
Area Bombing	NIS (meaning Not In Scope)
Area Coverage	OO (meaning no equivalent concept)

Subject Switching Coordinates. Phase Three completed the Subject Switching file by adding entries of coordinated DTIC terms translated to one or more NASA terms that express the same concept. For example:

Input-DTIC Terms	Output-NASA Terms
Angles; resolution	Angular resolution
Antennas; Gravity waves	Gravity wave antennas

Feedback and Maintenance. Phase Four was concerned with user feedback and file maintenance. New terms added to either the DTIC Thesaurus or the NASA Thesaurus require additions and modifications to entries in the data files. In addition, users can supply feedback as to translations that should be added or modified.

Results

In the NLD, the NASA STI Facility has a system that translates words and phrases from input material into equivalent concepts expressed in NASA posting terms. The system was designed particularly to allow the reuse of DTIC indexing in the NASA environment. According to a study of 250 DTIC documents, 89 percent of the terms assigned to DTIC documents by NASA indexers now are suggested by the NLD. The Facility also saves an estimated 3 minutes of indexing time per document.

While translating DTIC's index terms to those of the NASA Thesaurus, the NLD has preserved the quality of NASA's indexing. Also we know of no other system that differentiates concepts that are expressed by homonyms, or that coordinates terms in one vocabulary and translates them to the same concept expressed in the different terms of another vocabulary. The NASA Lexical Dictionary system is not only operating, but it is doing so with considerable success.

GLOSSARY

Access Routine — A general purpose computer program that accesses the NLD files. The Access Routine never operates independently; it is always called by an application program.

application program — A program that passes to the Access Routine two parameters:

- a ~~code that~~ indicates whether Phrase Matching or Subject Switching mode should be employed, and
- a character string that is either a word or phrase for Phrase Matching, or a set of posting terms assigned to a citation by a contributing source for Subject Switching.

continuation entry — An entry that tells the computer to continue to look for additional key elements in order to reach the posting term.

continuation symbol — The one or more asterisks (*) or percent signs (%) used in the posting term field of continuation entries.

controlled vocabulary — Terms that are authorized by an organization for their indexers to use in listing the subject matter of, or the concepts contained in, a document; a list of posting terms acceptable to the system and available for use.

DOE — Department of Energy

DTIC — Defense Technical Information Center

element — An element is part of the key to a record. In the Phrase Matching file, each word of the input phrase is an individual element. In the Subject Switching file, each contributing source posting term (which may be single or multiple words) is an element.

exception listing — A list of input posting terms which cannot be matched and translated by the NLD.

gloss — A parenthetical expression used to clarify the meaning of a posting term or Use reference that otherwise might be ambiguous. For example:

LOX (oxygen)
Pitch (inclination)
Pitch (material)

key — The subject of and a unique field in an NLD record. The key consists of terms that may be encountered in the input material. The key can consist of a single element, followed by a semicolon and two zeros (;00) or multiple elements separated by a semicolon (;).

LDICT Lexical Dictionary.

LEX.POSTTERM.-DTIC A sequential file of DTIC posting terms.

logic code A one character code which indicates how the key is to be processed.

major terms Posting terms that, in the judgment of the indexer, express the major concepts and research areas of a document. Major terms may be used for online searching and retrieval or to generate printed indexes. See also Minor terms.

minor terms Posting terms that, in the judgement of the indexer, indicate minor concepts and areas of interest in the information presented in a document. Minor terms encompass such aspects as properties, characteristics, action determined, relevant conditions of the investigation, measurement techniques, and instruments or calculations used when these aspects are not of primary importance. Minor terms do not appear in published subject indexes but may be used for online searching and retrieval. See also Major terms.

NASA National Aeronautics and Space Administration.

NLD NASA Lexical Dictionary.

posting terms Controlled vocabulary terms that are used by an organization's indexers to index documents for the use of that organization.

Phrase Matching file A file of NASA terms and Use references which are posted to valid NASA Thesaurus terms. The file is used as a general purpose translation table. This translation table accepts as input all of the posting terms and Use reference from the NASA Thesaurus as well as additional Use references constructed especially for the NLD.

Phrase Matching mode A general-purpose matching routine which attempts to find word-by-word matches between any input phrases and NASA posting terms or Use references.

The Phrase Matching mode can be used to process any type of phrase input. This input can consist of document titles, freely assigned keywords, or posting terms from a contributing source Thesaurus for which Subject Switching entries have not yet been created.

PMF — Phrase Matching file.

SSF — Subject Switching file.

STI — Scientific and Technical Information, as in the NASA STI Facility.

Subject Switching file — A file of a contributing organization's authorized vocabulary which provides, in the posting term field, one or more NASA posting terms that express the same concept. An entry is created for every contributed posting term, but in some cases the translation may indicate that the term is out of scope or not able to be translated. Additional entries are created for combinations of input posting terms posted to one or more different NASA terms.

Subject Switching mode — A special purpose routine that translates the posting terms assigned to a document by a particular contributing source (such as DTIC) into NASA posting terms. Subject Switching translates the concepts represented by the posting terms, in contrast to Phrase Matching which looks only for word matches. A unique translation table, called a Subject Switching file, is built for the posting terms of each contributing source.

Use reference — A reference from a posting term that is not in the controlled vocabulary to one that is. For example:

Condensation trails use Contrails

Use for reference — A posting term that is "used for" a term that is not in the controlled vocabulary. For example:

Contrails use for Condensation trails

VSAM — Virtual Storage Access Method. VSAM records, stored on direct access devices, may have fields of fixed or variable length and may be processed directly or sequentially.

APPENDIX A

PROCESSING OF INPUT WORDS AND PHRASES BY THE NLD ACCESS ROUTINE

The Access Routine determines the logic code associated with each element in the input array from the appropriate data file. If an element is not located in the file, a logic code of "?" is assigned for the use of Access Routine processing. The logic code controls the way the element will be processed by the Access Routine. A "T" logic code indicates that the Access Routine must look for combinations between that input element and other elements from the input array. Using "T" logic, a search key is created by adding to the "T" element a ";" followed by the next element in the input array. The Access Routine then tries to match this search key with a key in the data file.

- If the search key matches a file key which translates to a posting term, that posting term is returned.
- If the search key matches a file key which contains a continuation character (*, **, etc.) in the posting term field, then the next element from the input array is added to the end of the search key. A match is again attempted with the file.
- If no match is found for the search key, then the final element of the search key is replaced by the next element of the input array and a match is again attempted with the file.
- If all of the elements from the input array are tried without finding a match and the lead element has not been used in any other successful match, then a ";00" is added as the final element of the key for a final search attempt.

The "T" logic processing is repeated with each "T" logic element from the input array as the first element of the search key.

When the logic code of an element is anything other than a "T", the following logic is followed:

- If the element has already been used in combination with a "T" logic element in a successfully matched search key, then that element is skipped.
- If the element has not already been used in a successfully matched search key, the ";00" is added to the end of the element to create a search key that is matched against the file.

The Access Routine returns all of the matches made on the input character string to the application program.

The following example illustrates a simple case of Access Routine processing for the Phrase Matching example given in the section on SYSTEM DESCRIPTION under Lexical Dictionary Access Routine. Subject Switching processing is basically the same, but the Subject Switching input array would consist of contributing source posting terms sorted alphabetically.

Input Phrase: Engine endurance testing research laboratories

Input Array	Logic Code
Engine	T
Endurance	E
Testing	T
Research	T
Laboratories	E

Phrase Matching File Entries for the Words in the Input Array:

Logic Code	Key	Posting Term
E	Endurance;00	Endurance
T	Engine;Control	Engine Control
T	Engine;Design	Engine Design
T	Engine;Testing	*
T	Engine;Testing;Laboratories	Engine Testing Laboratories
T	Engine;Tests	Engine Tests
E	Laboratories;00	Laboratories
T	Research;And	*
T	Research;And;Development	Research and Development
T	Research;Facilities	Research Facilities
T	Research;00	Research
T	Testing;Machines	Testing Machines
T	Testing;Time	Testing Time
T	Testing;00	Tests

Access Routine Processing:

(References are made to the Input Array and Lexical Dictionary Entries shown above.)

Processing Description	Outcome
Logic code of first element is T. Create search key from first two elements. Look search key "Engine;Endurance" up in file.	Key not found.
Replace final element of search key with next element in array. Lock search key "Engine; Testing" up in file.	Key found. Continuation symbol returned.
Add next element in the array to the search key. Look search key "Engine;Testing;Laboratories" up in file.	Key found. Posting term "Engine testing laboratories" returned
Move on to next element in array. Logic code of second element is E. Second element has not been used in any T combination. Create search key from second element. Look search key "Endurance;00" up in file.	Key found. Posting term "Endurance" returned

Processing Description	Outcome
Move on to next element in array. Logic code of third element is T. Create search key from third and fourth element. Look search key "Testing;Research" up in file.	Key not found.
Replace final element of search key with next element in array. Look search key "Testing;Laboratories" up in file.	Key not found.
No more elements remain in array to be tried as a final element of the search key. "Testing" has already been used in a previous successful match. End processing for "Testing".	
Move onto next element in array. Logic code of fourth element is T. Create search key from the fourth and fifth elements. Look search key "Research;Laboratories" up in file.	Key not found.
No more elements remain in array to be tried as final element of search key. "Research" has not yet been used. Create search key be adding ";00" to "Research" Look search key "Research;00" up in file	Search key found. Posting term "Research" returned.
Move on to next element in array. Logic code of fifth element is E. "Laboratories" has already been used in a successful match. End processing for "Laboratories". No more elements in array. End processing.	

The final outcome of the processing is that the input phrase "Engine endurance testing research laboratories" is translated into the NASA posting terms "Engine testing laboratories", "Endurance", and "Research".

APPENDIX B

PROCEDURES FOR DETERMINING NASA TRANSLATION FOR DTIC TERMS FOR THE SUBJECT SWITCHING DATA FILE

1.0 INTRODUCTION

The DTIC/NASA Subject Switching capability requires that a file be created containing translations from DTIC posting terms to NASA posting terms. Because the creation of this file is a significant effort, automated methods have been employed to generate aids for the analysts. Lists of potential translations are obtained in the following way: DTIC terms are run through the NLD Phrase Matching file. The output is sorted alphabetically by DTIC terms in one printout and by NASA terms in another. NASA terms are run through the DTIC Lexical Dictionary (LDICT). This output is also sorted alphabetically by DTIC and by NASA terms.

Frequent references to these printouts required new nomenclature for ease of communication. Since the printouts are divided into four separate volumes, they are referred to as Books and numbered in alphabetical order, as follows:

DTIC/NASA sorted alphabetically by DTIC is Book 1.
DTIC/NASA sorted alphabetically by NASA is Book 2.
NASA/DTIC sorted alphabetically by DTIC is Book 3.
NASA/DTIC sorted alphabetically by NASA is Book 4.

In order to facilitate the use of the information contained in these four Books, additional programs re-sort portions of them and generate files of properly formatted candidate entries for the Subject Switching file. From the DTIC/NASA translations, separate files are created for:

- Exact Matches - DTIC posting terms for which there are exactly matching NASA posting terms and
- Partial Matches - DTIC posting terms for which there are one or more partial phrase matches with NASA posting terms or Use references.

and a printout is created of all:

- No Matches - DTIC terms for which there were no matches with NASA posting terms or Use references.

From the NASA/DTIC translations a file is created for:

- Tables - Two or more DTIC posting terms which, used in combination, translate to a NASA posting term or terms.

The record for an entry contains a tentative logic code, a key consisting of one or more DTIC terms and a posting term field which consisted of one or more NASA posting terms. In the case of the exact matches and partial matches, the analysts edit the tentative NASA translations, which were generated by the NLD. In the suggested table entries, the analysts edit the Keys which are tentative DTIC translations which were generated by the DTIC LDICT. The no match printout is a listing, rather than a file, of DTIC terms for which the NLD had no matching entry, either complete or partial.

Analysts study these potential entries and handle them according to the following set of guidelines covering both general procedures and procedures which are unique for each type of entry.

2.0 GENERAL PROCEDURES

2.1 RECORD FORMAT

The record format for the computer generated candidate entries for the DTIC/NASA Subject Switching file is as follows:

Logic code

A one character code which is entered in the second column of the record. The logic code provides information for Access Routine processing and describes the relationship between the key and the posting terms.

Logic codes may be one of the following:

E	The single term key and the posting term are equal.
C	There is a change between the single term key and the single term posting term.
L	The single term key translates to a list of posting terms.
I	The translation of the single term key is indeterminate and a choice is offered to the indexer.
O	There is no NASA translation for the single term key.
T	There are multiple terms in the key.

Key

The key begins in column 4 and consists of one or more valid DTIC posting terms. If there is only one DTIC posting term in the key, it is followed by ";00." If there are multiple posting terms in the key they are separated by semicolons. The key is followed by a "$" which separates the key from the posting term. Parentheses are removed from any term in the key.

Posting Term

The posting term begins following the key and $. The posting term consists of one or more NASA posting terms in exactly the same format in which they appear in the NASA Thesaurus. If there are multiple NASA posting terms, they are separated by commas.

Note: Entries which are manually coded for data entry using the online maintenance software follow the format given above with two exceptions.

1. A dollar sign is placed between the logic code and the key, as well as between the key and the posting term. No spaces are left between the logic code and the key. For example:

   E$Radar;00$Radar

2. No space is left before the logic code, because the online maintenance software will automatically place the logic code in the correct column.

2.2 USE OF SYMBOLS

Four symbols may be used with the NASA Thesaurus terms in the posting term field, to indicate the following conditions.

+	NASA has narrower terms to the suggested term, which are not covered in the DTIC Thesaurus. The indexer should look at the narrower terms to see if they are appropriate for use with the document in hand.
@	The suggested NASA term is an Array term.
>	The suggested NASA term is a broader concept than the DTIC term.
?	The indexer must select the NASA term or terms that are appropriate from the suggested terms.

Special rules for symbols:

1. When the "?" is used, all suggested NASA terms should be followed by a "?". No more than three terms should be suggested in an indexer choice entry.

2. The ">" should be used when the suggested NASA term is broader than the DTIC term. However, the ">" is not used when switching form DTIC terms incorporating broad concepts such as "methods", "systems", "equipment", etc. to a NASA term for the general subject. For example:

DTIC	NASA
Fire Control Equipment	Fire Control
Adaptive Control Systems	Adaptive Control

The ">" is also not used when a coordinated list of NASA terms is suggested as a translation, even if the coordinated terms are broader then the DTIC terms.

2.3 OTHER GENERAL GUIDELINES

1. If the NASA Thesaurus does not contain an exact match, then the name of a discipline, the product of the discipline, or the instrument used in the discipline may be used interchangeably as translation. No ">" or "?" is added. For example:

   Holography, Holograms

2. If the NASA Thesaurus does not contain an exact match, then the noun form and the gerund form of a term may be used interchangeably. No ">" or "?" is added. For example:

   Couplers, Coupling

3. If the source vocabulary has only one form of a term, and NASA has more than one, then all of the NASA variants should be presented followed by "?", as an indexer choice. For example:

   ØI Estimates;OO$Estimates?,Estimating?

4. Consistency should be maintained between similar switches. Check other entries that have been coded and try to follow the same pattern when a similar switch is encountered. For example:

   Brigade Level Organizations, Platoon Level Organization, etc. NIS

5. If the NASA Thesaurus does not contain an exact match, but does have the opposite, then the opposite is used as the translation. No">" or "?" is added. For example:

   Antijamming, Jamming

6. Avoid the use of Array terms when possible, but use Array terms rather than translating a term to "OO". If three or fewer "?" terms can be substituted for an Array term, do so. For example:

   NOT ØE Ballast;OO$Ballast@
   BUT ØI Ballast;OO$Ballast(Mass)?,Ballasts(Impedances)?

7. Geographical terms should be translated according to the following rules. The rules are listed in order of priority:

   Rivers

   a. Use the specific "River" term if it is available.
   b. If a "basin" term is available for the river, list both the "basin" term and "Rivers" followed by "?" as an indexer choice.
   c. If the river belongs in only one country or state, list "Rivers" and the country or state term.
   d. If the river is in the United States and belongs in more than one state, list "Rivers" and "United States."

e. If the river belongs in more than one country, list "Rivers" and the continent name.

Islands

a. Use the specific island term if it is available.
b. If the island is part of a larger island group for which there is a term, use the broader group term.
c. Use the "Islands" and the body of water in which the islands are located, unless a narrower combined term, such as "Pacific Islands" is available.
d. Use the term "Islands" and the name of a country only if the island is both owned by and adjacent to the country.

Cities, Towns, etc.

a. Use the specific city term if it is available.
b. Use the term "Cities" and the state or country in which the city is located.

Seas

a. Use the specific sea term if it is available.
b. Use the term "Seas" and the country or continent in which the sea is located.
c. If there is no single appropriate country or continent, use the term "Seas" with a >.

8. Limit a list of posting terms to three terms - two, if possible.

9. Post any term to 00 (zero zero) in preference. to a poor translation or one of questionable accuracy.

3.0 PROCEDURES FOR NO MATCHES

No computer generated entries are available for No Matches. These entries are manually coded and entered using the online maintenance software. Because of this different entry method, an additional "$" is placed between the logic code and the key for these entries.

1. Look for the DTIC term in the NASA Thesaurus, Volume 2: Access Vocabulary. This will locate any NASA term that is a format variation of the DTIC term, such as singular or plural, term inversion, or hyphenated form. It also will locate any NASA terms of which the DTIC term is part. If a variant form of the DTIC term is found, assign a logic code of C and enter the NASA term in the posting field. For example:

```
C$Abandonment;00$Escape (Abandonment)
C$Acetones;00$Acetone
C$Acoustooptics;00$Acousto-optics
```

2. If no helpful information is found in the Access Vocabulary, then look up the DTIC term in the DTIC Thesaurus to determine if any broader terms are listed for that term. If there is a broader DTIC term, look up the broader term in Volume 1 of the NASA Thesaurus to see if that term exists. Examine the hierarchy of that term evaluating all narrower and related terms, for a possible translation of the original DTIC term. If the best translation of the DTIC term is a broader NASA term, assign a logic code of C and enter the NASA Broader Term followed by a greater than sign (>) in the posting term field. For example:

   ```
   C$Acetonitrile;00$Nitriles>
   ```

3. If approaches 1 and 2 produce no translation for the DTIC term, check dictionaries for synonyms or related terms to the DTIC term, and look up these terms in Volume 1 of the NASA Thesaurus. If it is determined that two or more NASA terms are required to express the meaning of the DTIC term, assign a logic code of L and enter the NASA terms in the posting term field, separated by commas. For example:

   ```
   L$Adamantanes;00$Agents@,Curing
   ```

4. If the DTIC term has two or more equally valid NASA term translations, assign a logic code of I and enter the possible NASA terms in the posting term field following each NASA term with a question mark and separating terms with a comma. For example:

   ```
   I$Automata;00$Automatic control?,Automata theory?
   ```

5. If no appropriate NASA translation can be found for the DTIC term, assign a logic code of 0 (zero) and enter NIS in the posting term field if the concept is Not In Scope for NASA, or enter 00 (zero zero) in the posting term field if the concept is in scope for NASA but no term is available to express the concept. For example:

   ```
   0$Attorneys;00$NIS
   0$Autumn;00$00
   ```

4.0 PROCEDURES FOR PARTIAL MATCHES

Procedures for Partial Matches are the same as for No Matches, except that computer generated candidate entries suggest one or more NASA terms as a possible translation. These suggested NASA translations serve as a starting point for research in the NASA Thesaurus. Edit the posting term field of the computer generated entry by changing, adding, or deleting NASA posting terms based on research in the Thesaurus. Change the suggested logic code if it is required by changes made to the posting term field.

5.0 PROCEDURES FOR EXACT MATCHES

1. Look up the Exact Match term in both the DTIC Thesaurus and the NASA Thesaurus, Volume 1. Check to see whether the NASA and DTIC terms appear to express the same concept by checking broader, narrower, and related terms as well as the scope notes. If the terms express the same concept, go on to procedure 2. If the terms do not express the same concept, determine whether or not there is another NASA term that does translate the DTIC term. If there is no better translation, go on to procedure 2. If a better translation is available, assign a logic code of ØC and post to the different NASA term.

2. Check both Thesauri to see if the NASA term has narrower terms that are not in the DTIC Vocabulary. If so, then place a plus sign (+) after the NASA term. This indicates that the indexer should consider the NASA narrower terms.

   ØE Bolts;00$Bolts+

3. Look up the DTIC term in Book 3, NASA/DTIC sorted by DTIC, in order to see if multiple NASA terms are translated into the same DTIC term. If there are multiple, valid translations, then the following type of entry will be created:

   ØI DTIC term$NASA term 1?,NASA term 2?

   to indicate that the given DTIC term could have been used to express either of the concepts represented by the NASA terms listed with a question mark. The indexer must select the appropriate NASA term.

4. If the NASA term sometimes expresses the same concept as the DTIC term, and sometimes it does not, and if no better translation has been found, then assign a logic code of ØI and enter the NASA term in the posting term field, followed by a question mark to indicate that the indexer must decide whether or not this term is appropriate for the document being indexed.

   ØI Performance tests;00$Performance tests?

   (The NASA term applies only to operating equipment.)

5. If the DTIC and NASA terms express the same meaning, and if there is only one entry for the DTIC term in Book 3, then the logic code remains ØE and the posting term will be the NASA term listed.

   ØE Europe;00$Europe

6.0 PROCEDURES FOR DTIC TERM COORDINATIONS (TABLES)

1. DTIC terms in the key MUST be in alphabetical order. This is a requirement of the Access Routine. The term order within keys in

the computer generated entries will be correct.

2. If the translation presented appears reasonable, it should be accepted exactly as it is. For example:

   ƀT Abrasion;Resistance$Abrasion resistance

3. If other combinations of DTIC terms come to mind which provide equally good or better translations of the same posting term, code it or them for entry. Remember that the DTIC terms in the key MUST be in alphabetical order

   ƀT Abrasion;Wear resistance$Abrasion resistance

4. If multiple generic levels of the same concept appear in a suggested table, two entries should be made into the NLD. One will include the multiple levels of the concept, as suggested. The other will include only the most specific term.

   ƀT Acids;Ascorbic acid;Metabolism$Ascorb~~ic~~ acid metabolism
   add: ƀT Ascorbic acid;Metabolism$Ascorbic acid metabolism

   Note: Do not create tables with multilevel terms; only use multiple terms in the same hierarchy in a table if they appear as suggested entries.

5. If the broader generic term is the final term or if any table is imbedded in another table, the key in the shorter entry must have ;00 as the final segment.

   ƀT Acoustic waves;Excitation;Waves$Acoustic excitation
   ƀT Acoustic waves;Excitation;00$Acoustic excitation

6. If more than two DTIC terms are indicated in a suggested table, examine the terms for the most pertinent expression of the concept in the fewest possible terms.

   ƀT Aircraft rockets;Fins;Folding fins;Rockets;Vehicles$Folding fin aircraft rocket vehicle

   Fins and rockets as individual terms in the key are unnecessary and should be eliminated.

7. Check all DTIC terms in the suggested table for possible narrower terms that would be appropriate substitutions. If an appropriate substitution is found, code an additional table using that term. Remember: DTIC terms in the key MUST be in alphabetical order.

   ƀT Atmospheres;Spacecraft cabins$Spacecraft cabin atmospheres
   add: ƀT Controlled atmospheres;Spacecraft cabins$Spacecraft cabin atmospheres

8. If a translation appears to be inappropriate, look for a substitute term or terms to convey the concept of the NASA term.

   ƀT Aircraft cabins;Similators;Spacecraft$Spacecraft cabin
      simulators
   Change to: ƀT Simulators;Spacecraft cabins$Spacecraft cabin
      simulators

9. A ¢ in the computer generated entry indicates that the entry came from a NASA Use reference. Look up ¢ references in the NASA Thesaurus, Volume 1, to find the pertinent "Use For" (UF) reference. If the DTIC terms suggested are a reasonable translation of the UF reference, use the suggested table. If the terms are not a reasonable translation for that UF reference, delete the entry.

   ƀT Aircraft;Stars;Warning systems$EC-121 aircraft¢

   The UF reference applicable is "Warning Star aircraft". Since this suggested table does not translate the concept of "Star aircraft" by "stars" and "aircraft", this entry should be deleted.

10. Delete suggested tables posted to specific NASA terms for space vehicles, programs, projects, etc. which may be considered identifiers by DTIC, or which have no valid equivalent in DTIC's vocabulary. For example, delete:

    ƀT Antisubmarine ammunition;Engines;Underwater rockets$
       ASROC engine
    ƀT Artificial satellites;Biomedicine$BESS (satellite)

11. When a suggested table indicates that DTIC expresses a NASA term for a chemical compound as two terms, consisting of a chemical element and a complex, add the word "compounds" to the element if DTIC has the element-compound term.

    Computer generated entries: ƀT Acetates;Lead$Lead acetates
                                ƀT Aluminum;Hydrides$Aluminum
                                   hydrides

    Code as: ƀT Acetates;Lead compounds$Lead acetates
             ƀT Aluminum compounds;Hydrides$Aluminum hydrides

12. A single key may be found going to different posting terms in nonadjacent entries on the computer printout. Watch for these and examine them carefully. If possible, choose one posting term and delete the other. If both posting terms seem to be reasonable translations, list both NASA terms in one table, following each term with a question mark and separating them with a comma; then delete the other table.

    The automatically generated entries will appear as follows:

ƀT Acoustics;Stability$Acoustic instability
ƀT Acoustics;Nozzles$Acoustic nozzles
ƀT Acoustics;Stability$Frequency stability
ƀT Acoustics;Nozzles$Sonic nozzles

The analyst's revision results in the following two entries to replace the four automatically generated entries:

ƀT Acoustics;Stability$Acoustic instability?,Frequency
    stability?
ƀT Acoustics;Nozzles$Acoustic nozzles?,Sonic nozzles?

13. When a printed table suggests another table which you wish to add, be sure that the DTIC and NASA concepts are the same. If necessary, add a second NASA term to achieve this.

Computer generated entry:

ƀT Additives;Propellants$Propellant additives

add entries:

ƀT Additives;Rocket propellants$Propellant additives,
    rocket propellants
ƀT Additives;Solid propellants$Propellant additives,
    Solid propellants

14. Don't construct a table going to a list of NASA terms if the separate elements of the key already are or should be posted individually to the same terms you intend to list.

Examples of tables which should not be constructed:
ƀT Machinery;Performance tests$Machinery,Performance
    tests

because DTIC's term "Machinery" goes to NASA's term "Machinery" and DTIC's term "Performance tests" goes to NASA's term "Performance tests" (followed by a question mark). Another table which should not be constructed is:

ƀT Aircraft;Composite structures;Construction$Aircraft
    structures,Composite structures

because DTIC terms "Aircraft" and "Construction" translate to the NASA term "Aircraft structures" while "Composite structures" translates to "Composite structures".

REFERENCES

1. Klingbiel, Paul H., "Phrase Structure Rewrite Systems in Information Retrieval." Information Processing and Management (to be published).

2. Gevrter, William B., An Overview of Artificial intelligence and Robotics. Volume 1: Artificial Intelligence. Part B: Applications. NASA-TM-85836, National Aeronautics and Space Administration, October 1983.

1. Report No. NASA CR-3838	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle
An Operational System for Subject Switching Between Controlled Vocabularies: A Computational Linguistics Approach

5. Report Date
October 1984

6. Performing Organization Code

7. Author(s)
June P. Silvester, Roxanne Newton, and Paul H. Klingbiel

8. Performing Organization Report No.

10. Work Unit No.

9. Performing Organization Name and Address
Planning Research Corporation
Government Information Systems
1500 Planning Research Drive
McLean, VA 22102

11. Contract or Grant No.
NASW-3330

13. Type of Report and Period Covered
Contractor Report
Nov. 2, 1981- Dec. 31, 1983

12. Sponsoring Agency Name and Address
National Aeronautics and Space Administration
Washington, D.C. 20546

14. Sponsoring Agency Code
NIT-2

15. Supplementary Notes

Reference: Technical Directive 83-130

Abstract

The NASA Lexical Dictionary (NLD), a system that automatically translates input subject terms to those of NASA, was developed in four phases. Phase One provided Phrase Matching, a context sensitive word-matching process that matches input phrase words with any NASA Thesaurus posting (i.e. index) term or Use reference. Other Use references have been added to enable the matching of synonyms, variant spellings, and some words with the same root. Phase Two provided the capability of translating any individual DTIC term to one or more NASA terms having the same meaning. Phase Three provided NASA terms having equivalent concepts for two or more DTIC terms, i.e. coordinations of DTIC terms. Phase Four was concerned with indexer feedback and maintenance. Although the original NLD construction involved much manual data entry, ways were found to automate nearly all but the intellectual decision-making processes. In addition to finding improved ways to construct a lexical dictionary, new applications for the NLD have been found and are being developed.

17. Key Words (Suggested by Author(s))
Translating; Machine translation; Information systems; Information theory; Linguistics; Words (language); Computer programming; Information retrieval Semantics; Computer techniques; Terminology Nomenclature

18. Distribution Statement
Unclassified - Unlimited

Subject Category 82

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 96	22. Price* A05

*For sale by the National Technical Information Service, Springfield, Virginia 22161