ONR-25

Department of Defense

AD633279

MANUAL FOR BUILDING A TECHNICAL THESAURUS

PROJECT LEX
Office of Naval Research

A Task Force for the Preparation of a DoD Technical Thesaurus

April 1966

CLEARINGHOUSE
FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION

Hardcopy	Microfiche		
$2.00	$.50	28 pp	as

ARCHIVE COPY

Best Available Copy

Document is Unlimited

DOD MANUAL FOR BUILDING A TECHNICAL THESAURUS

Prepared By

Project LEX of the Office of Naval Research in coordination with DoD Representatives, other Government Representatives and the Engineers Joint Council.

APRIL 1966

Best Available Co

Distribution of this document is Unlimited

REPRODUCTION QUALITY NOTICE

This document is the best quality available. The copy furnished to DTIC contained pages that may have the following quality problems:

- Pages smaller or larger than normal.
- Pages with background color or light colored printing.
- Pages with small type or poor printing; and or
- Pages with continuous tone material or color photographs.

Due to various output media available these conditions may or may not cause poor legibility in the microfiche or hardcopy output you receive.

☑ If this block is checked, the copy furnished to DTIC contained pages with color printing, that when reproduced in Black and White, may change detail of the original copy.

DoD MANUAL FOR BUILDING A TECHNICAL THESAURUS

CONTENTS

SECTION	PAGE
FOREWORD	ii
PARTICIPANTS	iii
KEY TO ABBREVIATIONS	iv
I. INTRODUCTION AND SPECIAL DOD REQUIREMENTS	1
A. THESAURUS FORMAT	2
B. STANDARD DATA ELEMENTS	4
C. COMPUTER UTILIZATION	5
D. REFERENCES	6
II. THESAURUS RULES AND CONVENTIONS	7

Best Available Copy

FOREWORD

On October 12, 1965, the Director of Defense Research and Engineering assigned to the Office of Naval Research the responsibility for preparing a DoD-Wide Technical Thesaurus. ONR established Project LEX to undertake this mission.

One of the assigned requirements was to prepare a manual setting forth DoD conventions for thesaurus building. The manual is presented herewith. It has been developed through a coordinated effort that has included each of the Military Services and pertinent Defense Agencies as well as many other government and non-government activities. Members of the COSATI (Committee on Scientific and Technical Information of the Federal Council on Science and Technology) Sub-Panel on Cataloging and Indexing also participated in deliberations. This effort was terminated with a meeting on 29 March 1966 in which remaining issues were resolved.

The conventions set forth in this manual will be followed in the development of the DoD-Wide Technical Thesaurus. It is important to note that the same basic rules and conventions recorded here have been concurrently adopted by the Engineers Joint Council and will also be used in the revision of the Thesaurus of Engineering Terms.

Those who have attended meetings at Project LEX for development of these conventions are listed on the following page. It is not possible to list all those throughout the country who have attended gatherings at distant points to discuss this development or who collaborated in comments forwarded to the Project. This cooperation has been most gratifying and is heartily appreciated.

J.K. LEYDON
Rear Admiral, USN
Chief of Naval Research

Best Available Copy

PARTICIPANTS

Persons, and their affiliations, who have participated in deliberations resulting in the development of this manual:

Mr. J. Heston Heald	ONR, Director, Project LEX
Miss Mildred Bailey	OSD(Comp)-LEX Focal Point
Mrs. Ruth D. Camp	NARDIS, EJC, LEX Task Force
Dr. Joseph Caponio	National Institute of Neurological Diseases and Blindness
Mr. Parmely Daniels	ARO, Army-LEX Focal Point
Mr. John A. Dovel, Jr.	FTD, SDC, DIA, LEX Task Force
Mr. James Eller	OE, EJC
Miss Mary L. Engel	DoA, EJC
Mrs. Alma S. Evans	DIA, LEX Task Force
Mr. Terry L. Gillum	COSATI, DDC, EJC, OE, LEX Task Force
Mr. Edward Groff	NSA-LEX Focal Point
Mr. Charles W. Hargrave	COSATI, NASA
Mr. Robert Hays	ONR, Navy-LEX Focal Point
Mrs. Margaret S. Hicks	CFSTI, COSATI, EJC, LEX Task Force
Dr. Jack W. Hilf	OWRR, EJC
Mr. Paul C. Janaske	CFSTI, COSATI
Mr. Paul Klingbiel	DDC-LEX Focal Point
Mr. David M. Liston, Jr.	BMI, EJC
Dr. Aiden B. McNamara	MIT, EJC
Mr. Mark Newmark	EI, EJC
Lt. Colonel Davis Potter	DIA-LEX Focal Point
Major John Preston	DASA-LEX Focal Point
Mrs. Alpha G. Rose	DoA, EJC
Major Clayton Schlemm	AFSC, Air Force-LEX Focal Point
Dr. Matthew Schrenk	ONR, LEX Task Force
Mr. Frank Speight	EJC, Director of Information Sciences
Mr. H. Edmond Stiles	COSATI
Miss Grace Swift	NSA, LEX Task Force
Mr. Seymour I. Taine	NASA
Mrs. Carol A. Tippett	BMI
Mr. Eugene Wall	DDC, EJC, IDC, LEX Task Force
Mr. David S. Weaver	EJC

Key to abbreviations shown on following page

KEY TO ABBREVIATIONS

AFSC - Air Force Systems Command
ARO - Army Research Office
BMI - Battelle Memorial Institute
CFSTI - Clearinghouse for Federal Scientific and Technical Information
COSATI - Committee on Scientific and Technical Information, Federal Council for Science and Technology
DASA - Defense Atomic Support Agency
DDC - Defense Documentation Center
DIA - Defense Intelligence Agency
DoA - Department of Agriculture (Agricultural Vocabulary Project)
EI - Engineering Index
EJC - Engineers Joint Council
FTD - Foreign Technology Division, Air Force
IDC - Information Dynamics Corporation
MIT - Massachusetts Institute of Technology
NASA - National Aeronautics and Space Administration
NARDIS - Navy Automated Research and Development Information System, David Taylor Model Basin
NIH - National Institutes of Health
NSA - National Security Agency
OE - Office of Education, Department of Health, Education, and Welfare
ONR - Office of Naval Research
OSD - Office of the Secretary of Defense
OWRR - Office of Water Resources Research, Department of the Interior
SDC - System Development Corporation

Best Available Copy

SECTION I

INTRODUCTION AND SPECIAL DOD REQUIREMENTS

In accordance with a memorandum from the Director of Defense Research and Engineering (Ref. 1, page 6), the Office of Naval Research has instituted a project for the development of a DoD-wide technical thesaurus. This project, designated Project LEX, has as its mission the compilation of a comprehensive interdisciplinary reference authority for the terminology to be used in describing, communicating, and documenting the scientific and technical subject matter associated with such DoD activities as requirements studies, intelligence estimates, program planning, budget analysis, research and development, operations, supply, maintenance, and data element standardization.

The thesaurus will be a word-association list generically structured to permit the description of the subject content of a document to the desired level of generality or specificity at input and to permit description of the information required at output in equally precise terms. It will be a flexible authority list for vocabulary control that provides for the use of terms in combination for concept coordination, using manual, semiautomatic, or computer methods. A short list of general references is provided for those who would like more background on coordinate indexing systems. See Ref. 5-8, page 6.

Though the thesaurus will be designed primarily for coordinate indexing systems, it may be used in traditional subject heading systems, such as library card catalogs or printed subject indexes. The essential difference between subject heading and coordinate indexing systems is that combinations of terms in subject headings are fixed at the time of indexing or cataloging, whereas terms in coordinate indexing systems are manipulable at the time of searching, enabling the person looking for information to coordinate several terms simultaneously in any desired combination. With a few exceptions, which will be apparent, any authorized term in the thesaurus may be used either as a main subject heading or as a subordinate heading in the card catalog or printed index, depending upon the emphasis of the subject matter being described. The abstract terms described in Section I-A-3f are especially adaptable as subordinate headings.

To insure the comprehensiveness and technical competence of the DoD thesaurus, Project LEX will include a systematic review of existing thesauri, subject heading lists, glossaries, dictionaries, and other terminology authority lists now being used by DoD organizations, other government agencies, DoD contractors, and professional societies. Panels of subject specialists, in conjunction with Project

Best Available Copy

LEX personnel, will select descriptors (See Introduction to Section II) from these sources and develop the thesaurus structure of descriptor interrelationships.

Guidelines and conventions for the selection of descriptors and the compilation of the thesaurus are presented in Section VI, "Thesaurus Rules and Conventions." The guidelines are based on procedures developed in coordination with various DoD and other government activities and the Engineers Joint Council.

A. THESAURUS FORMAT

The thesaurus will consist of six sections: (1) introduction, (2) rules and conventions, (3) an alphabetical listing of all descriptors (See introduction to Section II) and cross references, (4) descriptors arranged by subject categories based on the COSATI Subject Category List (Ref. 3); (5) a display of hierarchical relationships of the descriptors, and (6) a permuted word display of all terms. The content to be included in each section of the thesaurus is described below.

1. Introduction. This section will present the thesaurus philosophy and explain how the thesaurus is to be used in indexing and searching.

2. Thesaurus Rules and Conventions. The rules and conventions used in constructing the thesaurus will be presented in Section II. Examples of the notations used and of the ways terms are treated will be given in the descriptor displays.

3. Alphabetical Section. The alphabetical section is the core display of the thesaurus. It will be an alphabetical list of all terms in the thesaurus.

a. USE references will be interfiled with the descriptors. Other appropriate cross references will be indented under the descriptors to which they refer. (See Section II, Rules C-1 to C-8).

b. Scope notes will be entered for descriptors whose meanings need to be clarified. (See Rule T-5c). A scope note will be used also to identify descriptors that have been established or proposed as standard data elements. (See Section I-B).

c. Each descriptor entry will include the COSATI Subject Category List notation for the subject group, or groups, to which a descriptor has been assigned. (See Section 1-A-4, below).

Best Available Copy

d. Filing rules for the alphabetical section have been set up to promote consistent interpretation by interfacing organizations and to simplify procedures for adding new terms to the initial vocabulary. Terms will be alphabetized letter-by-letter, according to Rule A-1, Section II.

e. The various elements of the alphabetical section are illustrated by the hypothetical entries in Figure 1, page 21.

f. To enhance the retrieval capabilities of the technical vocabulary, a limited number of terms representing general or abstract concepts will be established as descriptors. These descriptors will be chosen for their utility in delimiting, explaining, or modifying the treatment of the subject matter described by the technical terminology. Examples are: airborne, calibration, feasibility studies, preservation, vulnerability. To help develop consistent and appropriate use of these types of descriptors, special indexing instructions will be provided to accompany most abstract terms. A primary consideration in selecting abstract terms will be their applicability as points of subdivision in manual systems. Each abstract term will be designated by a special COSATI field and group notation. (See Section I-A-4, below).

4. Subject Categories. This section will be a display of all descriptors categorized according to the fields and groups of the COSATI Subject Category List. Subject category assignment will be based on the scope notes of the COSATI groups and the meanings of the individual descriptors. It will sometimes be necessary to assign a descriptor to more than one group. If necessary, additional groups will be provided (and recommended to COSATI for adoption) to accommodate descriptors for which no appropriate group exists in the COSATI list. Special groups will be established for abstract terms.

5. Hierarchical Display. This section of the thesaurus will display descriptors in hierarchical arrays, arranged alphabetically by the most generic descriptor in each array, showing by indentations the relationships of the narrower members of each class, as shown on the next page.

Best Available Copy

thermodynamics
- enthalpy
- entropy
- equations of state
- free energy
- heat flux
- thermodynamic cycles
  - brayton cycle
  - rankine cycle
  - stirling cycle

transmission lines
- coaxial cables
  - liquid filled coaxial cables
- field wire
- pulse cables
- radiofrequency cables
- telephone lines
- waveguides
  - step transmission lines
  - waveguide beads
  - waveguide circulators
  - waveguide couplers
  - waveguide filters
  - waveguide irises
  - waveguide slots
  - waveguide switches
  - waveguide windows

6. Permuted Term Display. This section will be a display of all terms in the thesaurus, ordered according to each meaningful word in the single and multi-word terms. All terms containing the same word will be grouped together, with the single word term first (where one exists), all other terms following in an alphabetical order:

```
                flow
           axial flow
boundary layer flow
                flow angle
                flow control
                flow deflection
          fluid flow
     large mass flow pumps
                fluid density
                fluid density measurement
                fluid flow
```

B. STANDARD DATA ELEMENTS

A memorandum from the Deputy Assistant Secretary of Defense (Comptroller) (Ref. 4) has directed that the development of the DoD Thesaurus be coordinated with the Data Elements and Data Codes Standardization Program. Procedures will be taken to establish candidate descriptors (See introduction to Section II) as standard data elements in accordance with that program.

Best Available Copy

When established standard data elements are included as descriptors, each will be identified and referenced to the standard data publication by means of a scope note showing the standardized abbreviation. (See Rule T-5c).

When proposed standard data elements are included in the thesaurus, each will be identified with a scope note as being in the process of standardization. Close liaison will be maintained with the Data Standards Division, ASD (Comptroller) to assure the proper identification of proposed standard data elements in the thesaurus.

C. COMPUTER UTILIZATION

To facilitate the editing of the thesaurus and to provide flexibility in developing displays of the vocabulary, a magnetic tape record of the thesaurus will be created. Computer support programs will be provided to manipulate these records on tape to permit (1) automatic coding of each descriptor; (2) various cross reference editing capabilities; (3) printing the thesaurus, or selected portions of it, in a format suitable for bulk reproduction; (4) additions or revisions to the vocabulary as required; and (5) the automatic compilation of the descriptor displays.

1. Coding. Once the descriptors and cross references have been recorded, a standard-length alphanumeric code will be assigned to each descriptor by the computer. The codes are being provided to assist Defense organizations which have access to computers, punched card equipment, or electronic accounting machines. The same coding will also be used internally during generation, processing, and updating operations of the thesaurus.

2. Editing. To avoid laborious manual editing of the cross reference structure, programs will be written to verify that terms are spelled consistently wherever they appear, that cross references are reciprocal as specified in the cross reference rules, that each term used as a cross reference is a valid descriptor when so indicated, and that each descriptor has been assigned to a valid descriptor group.

3. Displays. Using appropriate data from the alphabetical section, special programs will create the hierarchical and permuted word displays (See Section I-A-6), as well as the subject category display (See Section I-A-4).

Best Available Copy

D. REFERENCES

1. Memorandum from the Director of Defense Research and Engineering, dated 12 October 1965, subject: DoD-Wide Technical Thesaurus.

2. Engineers Joint Council, "Rules for Preparing and Updating Engineering Thesauri," June 1965.

3. COSATI Subject Category List, Committee on Scientific and Technical Information of the Federal Council for Science and Technology, Washington, D. C., Dec. 1964.

4. Memorandum from the Deputy Assistant Secretary of Defense (Comptroller), dated 2 December 1965, subject: Assignment of Data Elements and Data Codes Standardization Responsibility—DoD-Wide Technical Thesaurus.

5. Becker, Joseph and R. M. Hayes. Information Storage and Retrieval, John Wiley & Sons, New York, N. Y., 1963.

6. Bourne, Charles P., Methods of Information Handling, John Wiley & Sons, New York, N. Y., 1963.

7. Jaster, Josephine J., Barbara R. Murray, and Mortimer Taube, The State of the Art of Coordinate Indexing, Documentation, Inc., Washington, D. C., 1962. Contract No. NSF-C-147. AD 275 397.

8. Kent, Allen, Textbook on Mechanized Information Retrieval, 2d ed., John Wiley & Sons, New York, N. Y., 1962.

Best Available Copy

SECTION II

THESAURUS RULES AND CONVENTIONS

TERMS

The description of technical documents (indexing) for information storage and retrieval requires the use of two types of terms: (a) those that describe the information and data contained in the document, and (b) bibliographic terms--terms that describe the document itself, not the information in the document. Bibliographic terms, examples of which are personal authors, corporate authors, and publication dates, will not be included in the thesaurus. Terms that describe the information contained in the documents may include (1) project names, (2) military nomenclature, (3) identification symbols or numbers, (4) nicknames or jargon, (5) contract numbers, (6) geopolitical names, (7) tradenames, (8) other proper names, (9) terms of an analytical nature, such as aerosols, boundary layer, cardiovascular system, density, energy conversion, heat resistant alloys, spectroscopy; and (10) terms of an abstract nature, such as tests, measurement, and calibration. The index terms of primary concern to the thesaurus are types 9 and 10. For convenient reference, terms of these two types will be called "descriptors".

The rules that will be followed in constructing the thesaurus are of three types: (1) fundamental term rules (T-1 to T-10); (2) cross reference rules (C-1 to C-3); (3) the alphabetization rule (A-1).

(T-1) Term Selection: Candidate descriptors, i.e., terms in the raw vocabulary collected from a variety of sources, will be selected for inclusion in the thesaurus on the basis of their estimated usefulness in communication, indexing, and retrieval. In general, utility of terms can be estimated by considering (a) the relative frequency of appearance among the various contributed vocabularies, (b) the relative frequency of use within an operating system, (c) relationships to descriptors that have been selected previously, and (d) scientific or technical preciseness and acceptability. These factors are very much interdependent and will be considered together in the selection of descriptors.

a. The contribution by several sources of an identical term to represent a given concept suggests a usage consensus. However, the subject specializations of the contributors and the way in which terms are cross referenced will be taken into account in verifying that identical terms are identical in meaning.

Best Available Copy

b. Relative frequency of prior use of a term in indexing and searching within a particular vocabulary gives a rough quantitative indication of its possible usefulness. The frequency of use of a term is relative to the usage of other terms, to the relative age of the term, and to the age and scope of the collection indexed by the terms. Extremes of usage, high or low, are often caused by the term's being less than ideal. Terms that have been used relatively often within a given vocabulary may represent concepts that are poorly defined or too general to be useful in describing subject matter, whereas those that have been used very infrequently may represent concepts that are obscure or overly specific. Low frequency of use will not necessarily cause the rejection of a term that represents a novel concept and is a recent addition to the original vocabulary. Generally, it is best to establish descriptors that convey specifically the subject matter indexed and periodically review the frequency of their use to determine their utility.

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

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

(T-2) Noun Form: In keeping with established practice, noun forms will be used wherever possible: heat rather than hot; roughness rather than rough. In a limited number of instances, needed retrieval concepts can be represented only by adjectives or equivalent expressions. These usually take the form of words or phrases that describe in some manner the operation of equipment or systems: airborne, mobile, portable. Never use verbs; use catalysis rather than catalyze; the gerund pouring rather than the verb pour.

(T-3) Singular vs Plural: In choosing between singular and plural noun forms, the precedent long established by major indexing and subject cataloging operations will be followed. Generally, a useful rule of thumb may be applied as follows: use the plural form when the proposed term is a count noun, that is, a noun which may be used

Best Available Copy

to answer the question "how many?" (e.g., devices such as gages, nozzles, fuzes); use the singular form for mass nouns, those that express "how much?" (e.g., iron, wood, charcoal); use the singular for specific processes, properties, or conditions. Table 1 provides a useful summary of the recommended procedure. Common usage should be followed for term types not covered in the above general rule or in the table. Where the plural form of a word represents a distinctly different concept from that of the singular (e.g., gear, gears) both forms may be required.

(T-4) Direct Entry: Descriptors consisting of two or more words will be listed in their natural word order, i.e., the order normally used in English sentences: radar antennas rather than antennas, radar; refractory materials rather than materials, refractory.

(T-5) Term Definition: Terms that have more than one accepted meaning, that are intended to be used in a somewhat different way than ordinarily defined, or for which distinctions from other terms must be drawn, should be accompanied by an explanation. The meanings of terms will be clarified or made more specific in the following ways:

a. Modifying terms may be used to preface a given term, as in metal tubing to make the meaning of the word tubing more specific. This is the method underlying the construction of direct entries (See Rule T-4) and is subject to the limitations of the rule on combined terms (See Rule T-17).

b. A parenthetical qualifying expression may be appended to a term to clarify meaning, e.g., to distinguish among the meanings of homographs, as in mercury (metal) and mercury (planet). Such a qualifying expression becomes a part of the descriptor and should be entered with one space between the left parenthesis and the preceding character.

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

water cooling
(cooling by water)

d. When a tradename is listed in the thesaurus, the qualifying expression (tradename) should be appended to the tradename. When the meaning of a tradename is not self-evident, a scope note may also be added to clarify its meaning.

Best Available Copy

(T-6) Synonyms: When two or more candidate terms are true synonyms, one term will be selected as the descriptor, the other(s) entered as a cross reference. (See Rule C-2a).

(T-7) Quasi-Synonyms: To prevent scattering of like information in indexing and to obviate multiple searches for effective retrieval of information, it is both practical and desirable to consider terms having certain special relationships to be synonyms for indexing and retrieval purposes, i.e., quasi-synonyms.

Terms that represent different viewpoints of the same property continuum may be considered quasi-synonyms, e.g., smoothness or roughness. The preferred term will be entered as a descriptor, the other as a cross reference.

Terms representing concepts that overlap significantly may be treated as quasi-synonyms: lighting and illumination; duration and time; genetics and heredity. The preferred term will be indicated by a cross reference. (See Rule C-2b)

(T-8) Punctuation: Punctuation marks in descriptors will be restricted. Highly specific systematic names that require elaborate punctuation will be treated specially when they fall within the scope of the thesaurus. (See Rule T-10). Parentheses will be used to enclose qualifying expressions which are included in descriptors to prevent ambiguity. (See Rule T-5). Commas, periods, and apostrophes will be excluded. Hyphens will be used only in terms whose intended meaning would be altered by omission of the hyphen. In omitting a normally occurring hyphen, the space occupied by the hyphen will be handled according to these criteria: (1) retain the space for compound adjectives, noun-noun combinations, and letter-word combinations; (2) drop the space in attaching prefixing syllables to the base words. Examples:

(1)	high temperature testing	(2)	countermeasures
	man machine systems		microanalysis
	n body problem		ultrahigh frequency

(T-9) Abbreviated Forms: In general, abbreviated terms will be avoided as descriptors, since their understanding may not be universal, their meaning may be dependent on context, or their recognition may be dependent on capitalization and periods, which are constraints in computer operations. Abbreviated terms will be considered when meanings are well established and when significant gains in convenience can be demonstrated. Examples are: ACTH for adrenocorticotropic hormone; PETN for pentaerythritol tetranitrate; VTOL aircraft for vertical take off and landing aircraft. Abbreviated and unabbreviated forms of a given term

Best Available Copy

will be treated as synonyms and cross referenced accordingly. (See Rule C-2c). Some well established acronyms will be adopted as descriptors; for example, sherman, radar, laser.

(T-10) Specialized Vocabularies: Effective indexing and retrieval of information in certain specialized subject fields will require descriptor vocabularies that differ in some ways from the natural language approach of the thesaurus as a whole. For example:

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

Names of specific compounds that are commonly used materials or that are discussed frequently in a non-chemical sense and names of certain biologically significant compounds of complex structure may be entered as descriptors, e.g., sulfuric acid, carbon tetrachloride, morphine, progesterone.

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

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

(T-11) Types of Terms: As illustrated in Table 1, there are five basic types of terms: (1) materials, (2) properties, conditions or characteristics; (3) equipment or devices, (4) classes of use, and (5) processes. Cross references of the types USE, UF, BT, and NT will be made only between terms of the same type. (See Rules C-2 to C-5). However, RT references may be shown between terms of varying types. (See Rule C-8).

Employ the "-ing" suffix for processes and the "-ion" suffix or other appropriate suffixes for materials, characteristics, etc., when necessary to distinguish clearly between them. Examples are: altering and alterations; calculating and calculations. Where term construction prohibits differentiating by the use of suffixes, a parenthetical

Best Available Copy

qualifying expression should be used; for example, cladding (process) and lining (process).

(T-12) Combined Terms: The means by which to represent specific concepts that are, or appear to be, combinations of other individual terms must be considered from several points of view. The use of specific, i.e., precombined, terms to represent concepts that can be represented adequately by using individual terms in combination can needlessly increase the size and complexity of the indexing vocabulary. On the other hand, the improper use of term combinations often leads to inefficient retrieval.

As a rule, precombined terms should be established instead of using individual terms in combination when:

a. the meaning of one of the terms would be changed as a result of combination, for example: the term landing lights cannot properly be represented by a combination of the terms landing and lights because to do so would result in an improper use of the term landing and lessen its effectiveness in retrieval.

b. each term of the combination falls into a generic class which differs from that of the specific precombined term. For example, heat conduction is a logical member of the class heat transmission, but neither heat nor conduction is a member of that class.

c. the combined term represents a specific physical entity, for example: digital computers, or a specific material: sodium chloride.

d. the specific precombined term represents a concept that is encountered so frequently in indexing and searching that the ability to index and search directly is both expeditious and economical.

e. one or both of the individual terms in the combination is so heavily posted as to make searches awkward or inaccurate.

f. reasonable doubt remains after examining the foregoing criteria; if a specific term is established and later proves to be superfluous it can easily be reduced to a combination of terms, whereas, if a combination proves unworkable, great expense is incurred in establishing the proper specific term and performing the required reindexing.

Combination of individual terms should be used instead of establishing specific precombined terms when:

g. the concept conveyed by the specific term is duplicated exactly by the combination, for example: arc oxygen cutting can be

Best Available Copy

represented by the combination arc cutting and oxygen cutting.

h. the specific precombined term is based on the combination of terms representing materials and the form of materials, for example: brass rods can be indexed by the combination brass and rods.

i. the specific precombined term includes an adjective which can be converted to a noun, for example: metallic substrates can be indexed by the combination metals and substrates. (See Rule T-2).

When combinations of individual terms are employed, it is advisable to indicate this by a USE reference, for example: arc oxygen cutting; USE arc cutting and oxygen cutting; brass rods USE brass and rods; and metallic substrates USE metals and substrates. (See Rule C-2d).

Best Available Copy

GUIDELINES TO SINGULAR-PLURAL USAGE

TYPE OF TERM	USE SINGULAR FORM	USE PLURAL FORM
Material terms, such as: chemical compounds mixtures materials	When term is specific, as: urea cellophane beeswax	When term is generic, as: amines solvents plastics
Terms representing properties, conditions, characteristics	When term is specific, as: viscosity temperature purity opacity	When term is generic, as: physical properties process conditions
Equipment terms, such as: devices apparatus	Never use singular	Always use plural, as pulverizers reactors
Class of use terms	Never use singular	Always use plural, as: adhesives catalysts
Process terms	Always use singular, as: constructing installing modulating	Never use plural

Table 1

Best Available Copy

CROSS REFERENCES

(C-1) Cross References: Relationships among terms will be shown by cross references, which will aid users in selecting descriptors from the thesaurus. Types of cross references that will be used as required are:

USE
USED FOR
BROADER TERM
NARROWER TERM
RELATED TERM

These are described in the following rules.

(C-2) Use References: The USE reference is intended to lead users of the thesaurus to appropriate descriptors and will be employed to refer from a term that is not an authorized descriptor to the term which has been chosen as the descriptor, as follows:

a. to indicate a preferred synonym, e.g., lusozyme USE muramidase; secondary batteries USE storage batteries;

b. to refer from a specific term to a more general term which has been selected to represent (i.e., subsume) the specific concept, e.g., plant waxes USE waxes; sand blasting USE abrasive blasting;

c. to indicate a preference between spelling variations, or to expand or explain abbreviations, e.g., al mosaic USE alums; inflammability USE flammability; pentaerythritol tetranitrate USE PETN; IFF USE identification systems;

d. to prescribe the use of two or more descriptors to express a concept, e.g., ferromagnetic films USE ferromagnetic materials and films; antitank rockets USE antitank ammunition and rockets; optical illusions USE illusions and vision; compressor stall USE engine compressor systems and flow discontinuity;

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

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

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

Best Available Copy

h. to reflect current terminology, e.g., electrical condensers USE capacitors;

i. to eliminate jargon, e.g., whirly bird USE helicopters.

(C-3) Used For: The USED FOR reference (UF) is the mandatory reciprocal of the USE reference and accompanies the descriptor to which the USE reference refers. Accordingly, the USE references in two of the examples given in Rule C-2b and C-2c would generate the following USED FOR references:

abrasive blasting
UF sand blasting

pions
UF pi mesons

When a USE reference has prescribed two or more descriptors to represent a concept, e.g., ferromagnetic films USE ferromagnetic materials and films (See Rule C-2a), a number sign (#) will be placed after the unauthorized term in the USED FOR reference:

ferromagnetic materials
UF ferromagnetic films#

films
UF ferromagnetic films#

(C-4) Broader Terms: The BROADER TERM reference (BT) is employed to refer from a term representing a member of a class (or classes) of concepts to any term(s) in the thesaurus representing that class or classes, for example: steels BROADER TERM iron alloys. For each BROADER TERM reference there must also be provided a corresponding NARROWER TERM reference. (See Rule C-5). The part-whole relationship is usually not a broader-narrower relationship, for example: gear teeth BROADER TERM gears is incorrect. However, in certain specific areas, part-whole generics can be usefully employed; examples are anatomic names and geographic locations. Also specifically excluded for the broader-narrower term category are relationships based on the possible applications or uses of an entity, for example: platinum is not considered to be a member of the generic family catalysts because, although it is sometimes used as a catalyst, it has too many other applications to list all as broader terms. Platinum is, however, always a member of the class metals, so that the reference platinum BROADER TERM metals should be entered.

Best Available Copy

(C-5) Narrower Terms: The NARROWER TERM reference (NT) is the reciprocal of the BROADER TERM reference. (See Rule C-4) and is employed to refer from a term symbolizing a concept class to all terms symbolizing concepts that are members of that class, for example iron alloys NARROWER TERMS gray iron, mottled iron, steels. For each NARROWER TERM reference, there must be provided a corresponding BROADER TERM reference.

(C-6) Hierarchy: BROADER TERM references and NARROWER TERM references are hierarchical references. If there exist more than two levels in such hierarchies, the cross references for all levels must be completed for each term. This is done to enable the thesaurus user to ascertain the appropriate level of specificity in a family of generically related concepts and to promote editorial consistency during thesaurus revision, or in cases where portions of the thesaurus are extracted as specialized indexing vocabularies.

In a few instances, terms will be so broad in meaning that their utility as indexing terms will be doubtful, yet they must be retained for use in disciplines of peripheral interest or merely as a guide to more specific terminology. Under these circumstances, append the scope note "This term generally should not be used--use a related term." For example, the term materials is of little use in indexing documents that deal with materials in any but the most general way, but in a vocabulary in which many specific materials types are represented by indexing terms, the term is a useful point at which to display certain more specific terms (as related terms) for further study without carrying a useless BROADER TERM reference to materials on each of many specific terms. In general, only a very small proportion (< 1%) of terms in a thesaurus should require this scope note.

(C-7) Hierarchy Overlap: Terms may be members of more than one hierarchy. (See Rules C-4 and C-5). For example, consider the entries avalanche diodes BROADER TERMS diodes and semiconductor devices; diodes NARROWER TERM avalanche diodes. The term avalanche diodes represents a concept that is properly a member of the two different classes of concepts represented by the terms diodes and semiconductor devices.

(C-8) Related Terms: The RELATED TERM reference (RT) is used to refer to and from descriptors that bear a non-structured relationship to each other. (Cf. Rules C-4 to C-7). In general, any two descriptors are cross referenced RT if it is believed that the user, when examining one descriptor, might want to be reminded of the existence of the other.

Best Available Copy

RELATED TERM references may be used to identify:

a. descriptors that are closely related in meaning or concept

b. descriptors that are near synonyms

c. descriptors that have viewpoint interrelationships, such as a broader term-narrower term relationship based on usage. Example: alcohols, RT solvents, RT antifreezes

d. descriptors that have a part-whole relationship

e. descriptors that are members of different hierarchical structures and are related conceptually.

RELATED TERM references are not employed from one descriptor to other descriptors that are at different levels of the same hierarchy. Example:

abatement RT control	OR	abatement RT flood control pollution control	BUT NOT	abatement RT control flood control pollution control

In the above example, either the generic term, control, may be entered as an RT, or the two specific terms, flood control and pollution control, may be RT's to abatement, but both levels of the generic family should not be used. When many or most of the narrower terms under the broad related term are also related terms, the first alternative should be selected. Conversely, when only a few of the narrower terms are also related terms, the second alternative should be selected.

All RELATED TERM references will be reciprocal (e.g., lacquering, RT painting will require the reciprocal entry painting, RT lacquering) with the following exception: Where one of the terms involved bears the scope note, "(This term generally should not be used--use a related term)", a RELATED TERM reference to the term so scope-noted will be provided only from another term that has been scope-noted in the same way.

Best Available Copy

ALPHABETIZATION

(A-1) Alphabetize descriptors letter-by-letter, according to the following rules:

(1) Ignore all spaces between words.

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

(3) File according to the sequence:

(a) left parenthesis

(b) numerals in usual order: 0-9

(c) letters in usual order: A-Z.

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

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

Note: Metal working is sometimes spelled as one word (metalworking). In letter-by-letter alphabetization, the sequential position of metalworking is unaffected by the spelling selected for the authorized descriptor.

Best Available Copy

(Reverse blank)

ILLUSTRATION OF NOTATIONS

USE REFERENCE ——————> sonics ← NOT AN AUTHORIZED DESCRIPTOR
(Rule C-2) USE sound waves ← DESCRIPTOR

DESCRIPTOR ——————> sound waves
COSATI FIELD & GROUP ——> 20 01
SCOPE NOTE ——————>
(Rule T-5c) (Disturbances propagated through an elastic medium causing alterations in pressure or a displacement of the particles of the medium. Does not include sound sensation.)

USED FOR ——————>
(Rule C-3) UF elastic waves ← NOT AUTHORIZED DESCRIPTORS
sonics

NARROWER TERMS ——————>
(Rule C-5) NT seismic waves
shock waves
underwater sound

BROADER TERM ——————>
(Rule C-4) BT waves ← DESCRIPTORS

RELATED TERMS ——————>
(Rule C-8) RT acoustics
audio frequency
hearing
noise control
sound propagation

Figure 1

Note: The terms shown above have been selected merely as examples for this display. They will not necessarily appear in the finished thesaurus.

(Reverse blank)

Best Available Copy

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

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

ORIGINATING ACTIVITY (Corporate author) PROJECT LEX, OFFICE OF NAVAL RESEARCH	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

REPORT TITLE

DOD MANUAL FOR BUILDING A TECHNICAL THESAURUS

DESCRIPTIVE NOTES (Type of report and inclusive dates)

AUTHOR(S) (First name, middle initial, last name)

PROJECT LEX STAFF

REPORT DATE April 1966	7a. TOTAL NO. OF PAGES 24	7b. NO. OF REFS 8
a. CONTRACT OR GRANT NO b. PROJECT NO MISC-0359 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) ONR-25 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	

10. DISTRIBUTION STATEMENT

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

Guidelines and conventions are presented for the selection of appropriate technical subject indexing and retrieval terminology and the display of this terminology in a thesaurus format. These conventions were developed by Project LEX, the DoD technical thesaurus task force, in cooperation with the Engineers Joint Council. A discussion is given of (1) the criteria for determining the usefulness of prospective thesaurus terms or "descriptors," (2) means of resolving ambiguities among descriptors, and (3) methods for systematically creating a cross reference structure that will display hierarchical and conceptual interrelationships. The thesaurus format comprises five main sections, (1) an introductory statement explaining the purpose of the thesaurus and the way in which it is arranged, (2) an alphabetical listing of all descriptors and cross references, (3) descriptors arranged by subject categories based on the COSATI Subject Category List, (4) a graphic display of hierarchical relationships, and (5) a permuted display of all descriptors in the order of each meaningful word.

Best Available Copy

DD FORM 1473 (PAGE 1)

S/N 0101-807-6801

UNCLASSIFIED

Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
Information retrieval						
Vocabulary						
Thesauri						
Instruction manuals						
Descriptors						
Indexing vocabularies						
LEX project						

DD FORM 1 NOV 65 1473 (BACK)
(PAGE 2)

UNCLASSIFIED

Security Classification

Best Available Copy

END

DATE

FILMED

6-21-66