chemical dictionary.

For the CPSC project, the chemical dictionary consists of an unstructured machine-readable file of acceptable chemical nomenclature for product ingredients. The terms are listed in alphabetical order and compiled in conjunction with the chemical edit. As new chemical terms are derived from the chemical edit, they are added to the chemical dictionary without regard to synonymous or hierarchical relationships, providing they are valid and specific. Upon project completion, the chemical dictionary will be processed through the Name Match Program to standardize nomenclature and associate synonymous terms.

Computerized editing routines, which check the individual files for completeness and compatibility with each other and with the manufacturer and product input data, are an important element of the data base development system.

CONCLUSION

This work, which has been conducted for both NIOSH and CPSC, in compiling data bases containing the chemical ingredients of over 100 000 trade name products, represents a significant and perhaps unprecedented undertaking. Printed dictionaries of trade name products and ingredients have been published in the past, but these have been neither as exhaustive nor as specialized as the NIOSH and CPSC efforts, and they have not been computer searchable by either product name or ingredients. In contrast, the final result of both the NIOSH and CPSC projects is a machine-readable, standardized, and highly specific data base.

This paper represents the views of Auerbach Associates, Inc., and does not necessarily reflect the views of the government agencies concerned.

Production of a Hierarchical Chemical Thesaurus†

HERBERT B. LANDAU* and WENDY L. BYER

Auerbach Associates, Inc., Philadelphia, Pennsylvania 19107

Received March 23, 1976

Through the utilization of computer-aided thesaurus building techniques, a standard vocabulary of chemical nomenclature was established to facilitate indexing and retrieval for a data base which defines worker exposures to organic and inorganic chemical compounds. Developed for the National Institute for Occupational Safety and Health, this vocabulary control tool, in the form of an information retrieval thesaurus, contains 12 000 terms with approximately 8000 preferred descriptors and approximately 4000 chemical synonyms. The thesaurus follows the ANSI Standard as to structure and cross-referencing conventions and was constructured according to rigorous lexicographical procedures employing a methodology built around a computer system known as the Hierarchical Indented Thesaurus System (HITS). The thesaurus features a hierarchical indented term display, a term "tree structure", and automatically generated hierarchical decimal classification codes.

INTRODUCTION

Since July 1973, AUERBACH has been conducting the Trade Name Ingredient Clarification (TNIC) Project for the Hazards Surveillance Branch of the National Institute for Occupational Safety and Health (NIOSH) (Landau[1]). The purpose of this study is to identify and record the specific chemical ingredients of up to 86 000 trade name industrial products recorded as exposures to workers during the National Occupational Hazard Survey.

To ensure maximum utility and retrievability of the resulting ingredient data base, component information reported by manufacturers must be standardized. In addition, editing of manufacturers' reports to ensure that valid and specific chemical nomenclature is employed requires some form of vocabulary control. It was therefore decided early in the project to create a chemical thesaurus, with a full hierarchical and cross-reference structure, to serve as the means of converting the diverse chemical nomenclature reported by manufacturers into a single logically structured vocabulary. This chemical thesaurus, known officially as the "Exposure Dictionary of the National Occupational Hazard Survey" (EDNOHS), has grown as the project progressed. It now fills approximately 1500 pages of computer printout and consists of approximately 8000 main postable (i.e., preferred) terms, 4000 synonyms, and over 55 000 cross references and is employed as a vital working tool in this project.

† This work was performed under National Institute for Occupational Safety and Health Contract No. HSM-99-73-67. This paper was presented at the 10th Middle Atlantic Regional Meeting of the American Chemical Society, Philadelphia, Pa., Feb 24–25, 1976.

THESAURUS PURPOSE AND FUNCTION

By common definition, an information retrieval thesaurus is a "...compilation of words and phrases showing synonymous, hierarchical, and other relationships and dependencies, the function of which is to provide a standardized vocabulary for information storage and retrieval".[2]

Therefore, when the need for some form of vocabulary control tool for trade-name product ingredients became evident early in the NIOSH TNIC project, it appeared that the information retrieval thesaurus concept might be suited to our needs.

A thesaurus could provide a means of resolving the cross-industry (and in many cases intra-industry) synonymy present in the diverse ingredient nomenclature submitted by manufacturers (e.g., Burnt Lime vs. CaO vs. Calcium Oxide; EtOH vs. Ethanol; Melaniline vs. 1,3- Diphenylguanidine; Chlorophenol Sodium Salt vs. Sodium Chlorophenate, etc.). It could also ensure that those who will search the chemical ingredients data base will be speaking the same language as those who built the data base by providing a common standard vocabulary. In addition, since chemical elements and com-

pounds readily lend themselves to hierarchical and generic functional groupings, the logical technique of structuring and displaying chemical ingredient class relationships (as found in a thesaurus) could prove helpful to chemical editors and data base users. Likewise, the various notations and cross references employed by a thesaurus could also prove valuable in distinguishing between or grouping similar chemical entities and in linking chemically related ingredients for future epidemiological studies.

Therefore, through the use of a thesaurus structure, vocabulary ambiguity and redundancy could be minimized using the following techniques:

Resolution of synonyms
Definition of class membership
Standardization of nomenclature
Display of term relationships

These considerations led to the decision to create a full chemical thesaurus, patterned after document retrieval thesauri, in accord with the ANSI[2] standard guidelines as a vocabulary tool for the TNIC project. Although historically, information retrieval thesauri have been developed primarily for document retrieval applications, we saw no reason why they could not be applied with equal success to chemical data retrieval systems. However, it was recognized that in this instance, *specific chemicals*, not documents about chemicals, are being indexed. This means that, unlike document indexing applications, only the most specific term (i.e., the actual chemical compound or element name) is a valid index term. Generic names such as aromatic hydrocarbons or aldehydes are therefore used only for thesaurus structuring and not for indexing chemical responses. This is somewhat contrary to general document retrieval thesaurus-building practice, where the objective is to limit specific chemical names. For example, the Thesaurus of Engineering and Scientific Terms (TEST) rules and conventions state:

> "To avoid proliferation of terms in the field of chemistry, the names of specific chemical compounds as descriptors should be restricted. Instead, a vocabulary of descriptors representing generic compound classes, functional groups, and structural features should be devised".[3]

While this guideline might apply to a general purpose document retrieval thesaurus, it does not serve a specific chemical ingredient retrieval thesaurus such as this project required. This different slant in approach, along with the relative dearth of existing chemical thesauri, dictated that a new thesaurus should be created for the TNIC. Therefore, the EDNOHS thesaurus was designed as a special-purpose, mission-oriented tool and not an exhaustive, general-purpose vocabulary.

THESAURUS LOGICAL STRUCTURE

The basic logical framework around which the EDNOHS thesaurus is built is the lattice structured term family (Landau[4]). This thesaurus structure follows a form of an "inverted tree" logic which allows a particular thesaurus term to belong to more than one conceptual class (i.e., a term can have more than one broader term) in contrast to the more traditional "tree structure" which allows only one parent term per node. This lattice structure concept (as opposed to the conventional tree structure) is illustrated in Figure 1.

A term may be a member of more than one generic family hierarchy when it is properly a member of two different classes of concepts. For example, CHROMIUM POTASSIUM SULFATE is a subset of both INORGANIC CHROMIUM COMPOUNDS and INORGANIC POTASSIUM COMPOUNDS. Soergel,[5] who termed this structure the "polyhierarchy", has found it to be flexible and useful in

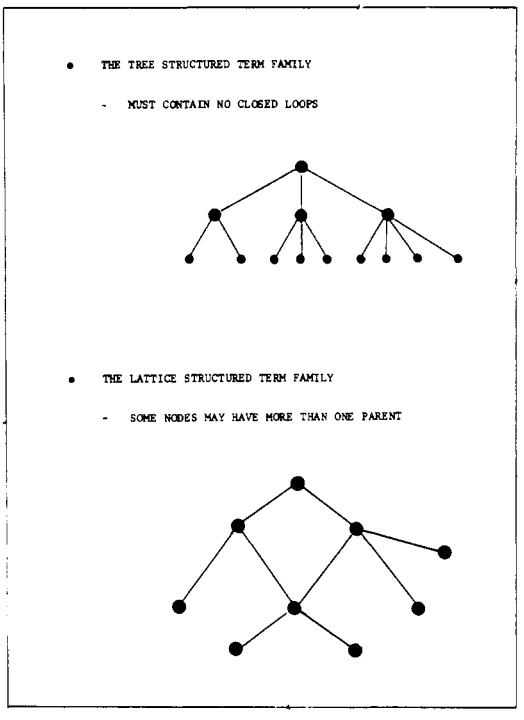


Figure 1. Thesaurus hierarchical structures.

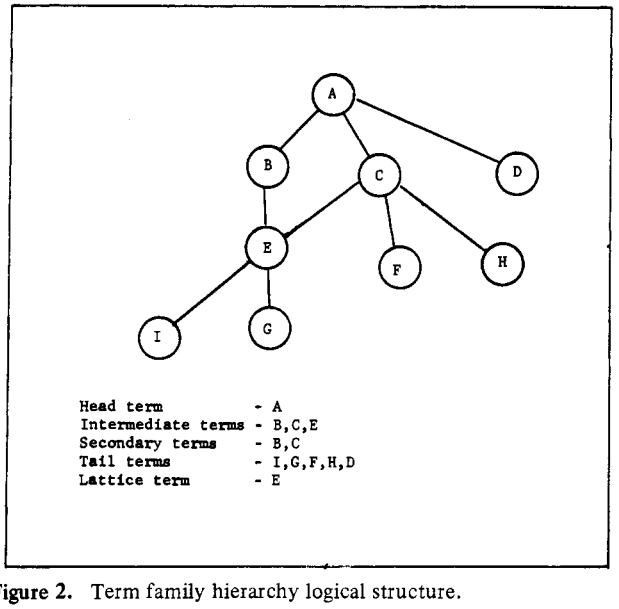


Figure 2. Term family hierarchy logical structure.

thesaurus construction, as have we at AUERBACH, in the construction of four scientific thesauri. The structure readily lends itself to chemistry.

In building the logical framework of EDNOHS according to a lattice structure, we defined a set of three generic "head" or category terms under which succeeding generations of specific chemical compounds could be grouped as the thesaurus evolved. These three main divisions are:

Organic Chemical Exposures
Inorganic Chemical Exposures
Nonchemical Exposures

Each of these divisions is further subdivided into secondary "intermediate terms" under which detailed term hierarchies or families are constructed. Intermediate terms are generic terms which will either have additional intermediate or "tail terms" narrower to them. The intermediate, or generic, terms are necessary to produce a logical and full hierarchical

Table I. Organic Head Terms

Aldehydes
Alicyclic Compounds
Alkaloids
Alkene Derivatives
Alkyne Derivatives
Carbohydrates
Chelates
Cyclic Compounds
Esters
Ethers
Halogenated Hydrocarbons
Hydrocarbons
Ketones
Nitrogen–Containing Organic Compounds
Organic Acid Anhydrides
Organic Acids
Organic Alcohols
Organic Peroxides
Organometallic Compounds
Organic Salts
Phosphorus–Containing Organic Compounds
Polymers, Organic
Silicon–Containing Organic Compounds
Steroids
Sulfur–Containing Organic Compounds
Vitamins

structure to accommodate the specific tail terms. Tail terms represent specific chemical substances or elements (generally terms at the lowest hierarchical level) which have no terms below them. Figure 2 demonstrates this logical thesaurus structure.

A typical "family tree structure" could appear in a hierarchical form such as the following:

INORGANIC CHEMICAL COMPOUNDS

- INORGANIC METAL CATION COMPOUNDS
- ALKALI METAL CATION COMPOUNDS
- INORGANIC POTASSIUM COMPOUNDS
- INORGANIC POTASSIUM COMPOUNDS CONTAINING HALOGENS
- POTASSIUM CHLORIDE

INORGANIC METAL CATION COMPOUNDS, ALKALI METAL CATION COMPOUNDS, INORGANIC POTASSIUM COMPOUNDS, and INORGANIC POTASSIUM COMPOUNDS CONTAINING HALOGENS are the intermediate terms. INORGANIC CHEMICAL COMPOUNDS is the head term of this "tree" and POTASSIUM CHLORIDE is the specific substance term.

The intermediate term which heads a term's family may be considered that family's head term. Head terms for the principal families of organic chemistry (see Table I) are selected to represent major functional groups. Functional groups are defined for this purpose to mean groups defined by principal structural features as well as defined by the common substituent chemical components. This approach is consistent with the organizational approach of most standard organic chemistry text books, chemical information systems, and scientific library classification schemes (e.g., Library of Congress). It relies less on the knowledge of chemical reactivity than on a systematized application of lexicographic rules, and it tends to reflect named parts of the specific compounds and therefore establish clear hierarchical relationships throughout the term families.

The head terms for the principal families of inorganic chemistry (see Table II) are selected on the basis of the periodic table of elements and the thermodynamic properties of their inorganic compounds. The elements are classed according to metals and nonmetals and grouped according to their traditional series and group names: alkaline earth metals, lanthanides, actinides, halogens, etc. The compounds are arranged according to metals and nonmetals and then according to their anion and cation characteristics (e.g., aluminates, aluminum cation compounds) (see Table II).

Table II. Inorganic Head Terms

Chemical Elements
Metal Elements
Nonmetal Elements
Inorganic Chemical Compounds
Inorganic Complexes
Metal Anion Compounds
Metal Cation Compounds
Nonmetal Anion Compounds
Nonmetal Cation Compounds
Inorganic Mixtures

This traditional approach to inorganic chemistry is reflected in the standard treatises and classification schemes, including Gmelin's "Handbuch der Anorganischen Chemie" and Mellor's "Comprehensive Treatise on Inorganic and Theoretical Chemistry".

Nonchemical Exposure families are developed to accommodate exposure terms that are reported in the Trade Name Ingredient Clarification Project (or were collected through prior NIOSH efforts), but which cannot be readily classified as specific chemical compounds or chemical elements (e.g., botanicals, foodstuffs, fly ash, seawater, etc.)

This terminological approach is somewhat of a compromise between the two levels of chemical nomenclature defined by Tate:[6] (a) the "convenient general language" or "everyday" language of chemistry which uses simple common names (also known as "arbitrary" or "trivial" names), and (b) the "legal" language which requires strict and systematically defined terms (such as IUPAC system).

While we have attempted to systematize and standardize the everyday language of industrial chemicals, we have not gone so far as to apply some of the rigorous and detailed IUPAC systematized nomenclature rules such as those employed by *Chemical Abstracts*[7] in its subject indexes, since these would, we believe, complicate both thesaurus construction and its eventual usage by a wide range of users, including industrial hygienists and engineers.

THESAURUS BUILDING PROCEDURE

The construction of EDNOHS involves essentially five distinct and somewhat repetitive operations:

(1) Definition of Thesaurus Rules and Conventions
(2) Candidate Term Collection and Standardization
(3) Term Classification and Cross-Referencing
(4) Computer Cross-Reference Validation and Editing (Preprocessor)
(5) Computer Generation and Display of Complete Hierarchies and By-Products

Figure 3 shows the interrelationship of these operations in flow chart form. Each of these tasks is reviewed below.

DEFINITION OF THESAURUS RULES AND CONVENTIONS

The first operation, the definition of thesaurus rules and conventions, was a prerequisite to all other steps. A set of unambiguous conventions, covering all aspects of the thesaurus construction process, was necessary in order to ensure both intra- and inter-lexicographer consistency in the building of EDNOHS and to facilitate training of new lexicographers. It was decided to base the EDNOHS conventions on the recently issued ANSI Thesaurus Guidelines[2] with relatively minor modifications as dictated by project requirements. In order to serve as the "authority" for thesaurus construction, the rules

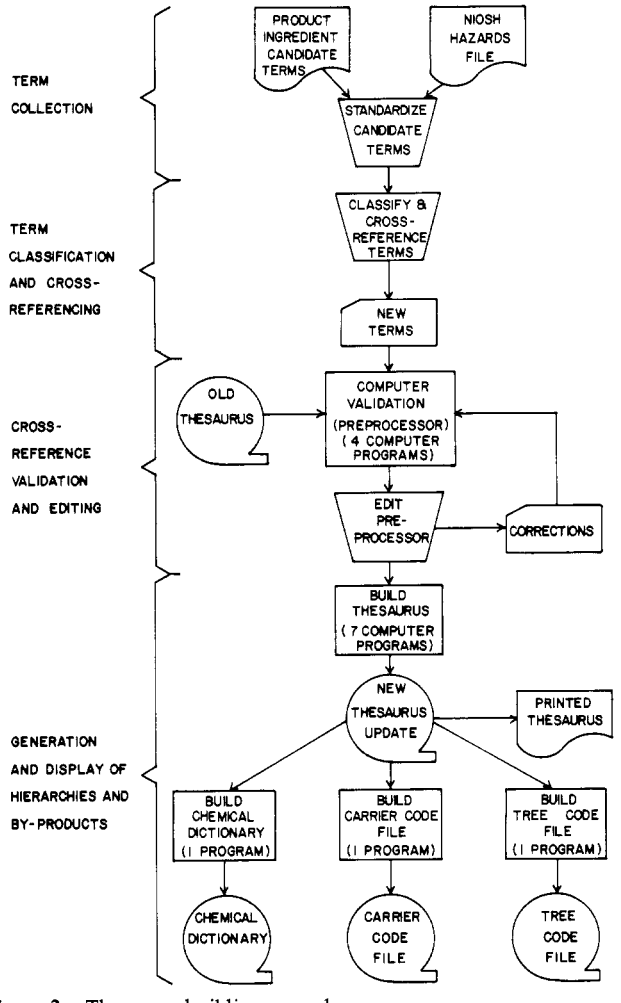


Figure 3. Thesaurus building procedure.

and conventions, along with detailed instructions for each operation, were issued as a 115-page procedure manual which was given to all lexicographers. This manual specified thesaurus rules and procedures for such areas as:

Term Selection Criteria
Assignment of Terms to Chemical Families
Grammatical Forms and Punctuation
Multiword Terms
Superscripts and Subscripts
Scope Notes and Qualifiers
Cross References
Alphabetization
Display Formats
Abbreviations

CANDIDATE TERM COLLECTION AND STANDARDIZATION

Candidate chemical terms for EDNOHS come principally from two sources: (1) product ingredients reported by manufacturers in the course of the trade name ingredients survey and (2) chemicals identified by NIOSH as potential occupational hazards. A candidate term is recorded on an index card and is first compared with the existing thesaurus to see if it is already entered. If not, it is then verified in one or more reference authorities to ensure that it is a valid chemical entity or a synonym to an existing EDNOHS term. These authorities also serve as a source of standard common names for chemical compounds. They include:

"Merck Index, An Encyclopedia of Chemicals and Drugs", 8th ed, P. G. Stecher, Ed., Merck, Rahway, N.J., 1968.

Table III

Cross reference	Symbol	Reciprocal
Broader Term	BT	NT
Narrower Term	NT	BT
Related Term	RT	RT
Use	USE	UF
Used For	UF	USE

"Handbook of Chemistry and Physics", 53rd ed, R. C. Weast and S. M. Selby, Ed., Chemical Rubber Co., Cleveland, Ohio, 1972.
G. S. Brady, "Materials Handbook", 10th ed, McGraw-Hill, New York, N.Y., 1971.
Synthetic Organic Chemical Manufacturers Association, "SOCMA Handbook: Commercial Organic Chemical Names", American Chemical Society, Washington, D.C., 1965.
"Condensed Chemical Dictionary", Van Nostrand, Princeton, N.J., 1971.
"Hackh's Chemical Dicitionary" (American and British Usage), 4th ed, J. Grant, Ed., McGraw-Hill, New York, N.Y., 1969.
"Naming and Indexing of Chemical Substances for Chemical Abstracts during the Ninth Collective Period (1972-1976)", Sect. IV, *Chem. Abstr.*, **76**, (1972).
"Gmelins Handbuch der Anorganischen Chemie", 8th ed, Verlag Chemie, Weinheim, 1924, Vol. 1–(in progress), imprint varies.

The EDNOHS rules and conventions are then applied to the verified candidate terms and term synonyms with the result being a list of verified and standardized terms and synonyms ready for categorization and cross referencing.

TERM CLASSIFICATION AND CROSS-REFERENCING

Using the classification scheme of head and intermediate terms (Tables I and II), candidate terms are sorted into their respective term families. Terms belonging to more than one family are identified as lattice points. Terms not fitting into any family are set aside for further study (which may lead to a revision of the classification scheme).

Each term is then placed into its appropriate hierarchical position in its respective term family or families. Intermediate terms necessary to complete term families or to act as bridges are established in standard fashion and added to the hierarchies along with scope notes and qualifiers.

When each individual term family hierarchy is complete, synonyms and appropriate cross references to other postable (i.e., valid) or nonpostable (i.e., invalid) terms may be added. In conference with the ANSI Standard Z39.19,[2] the five-function cross reference set shown in Table III is employed. The entire family (including all terms and cross reference postings), now in standardized form, is transferred to a set of Term Review Form coding sheets (Figure 4) for data conversion and computer validation and editing.

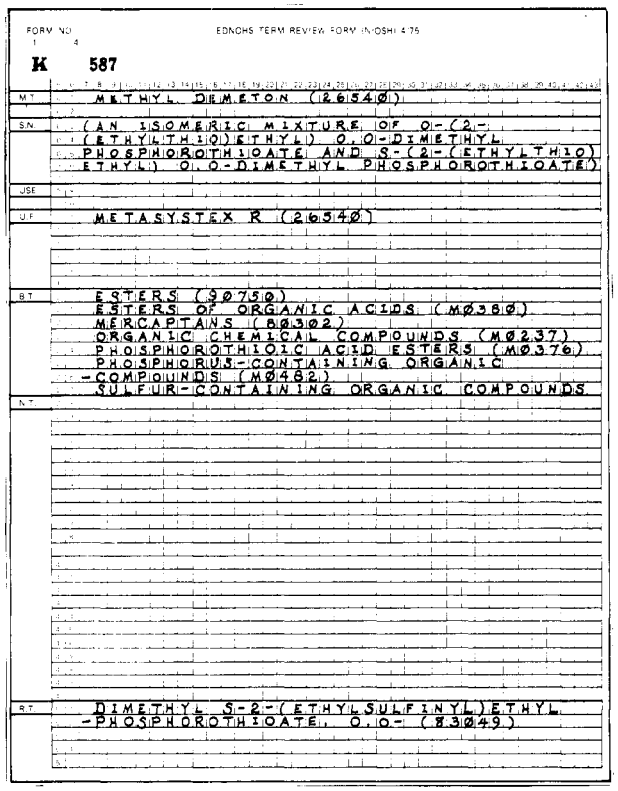
COMPUTER CROSS REFERENCE VALIDATION AND EDITING (PREPROCESSOR)

The task of ensuring cross reference consistency and generating missing cross references is handled by the preprocessor validating and editing module of HITS, AUERBACH's Hierarchical Indented Thesaurus System.[4]

HITS consists of a set of 11 COBOL programs which operate on either IBM 360 or 370 series computers. HITS has been continually employed as a thesaurus building tool since 1971, and the system exhibits the following features desirable for the production of a chemical thesaurus:

(a) It allows for chemical families to be built on a one by one basis and later joined into complete hierarchies by the system using lattice terms common to more than one family.

(b) It allows for term families to grow by the inclusion of new terms either as intermediate terms within a family, head terms, and/or tail terms.



FORM NO. EDNOHS TERM REVIEW FORM (NIOSH) 4-75

K 587

MT METHYL DEMETON (26540)

SN (AN ISOMERIC MIXTURE OF O-(2-(ETHYLTHIO)ETHYL) O,O-DIMETHYL PHOSPHOROTHIOATE AND S-(2-(ETHYLTHIO) ETHYL) O,O-DIMETHYL PHOSPHOROTHIOATE)

USE

UF METASYSTEX R (26540)

BT ESTERS (90750)
ESTERS OF ORGANIC ACIDS (M0380)
MERCAPTANS (80302)
ORGANIC CHEMICAL COMPOUNDS (M0237)
PHOSPHOROTHIOIC ACID ESTERS (M0376)
PHOSPHORUS-CONTAINING ORGANIC COMPOUNDS (M0462)
SULFUR-CONTAINING ORGANIC COMPOUNDS

NT

RT DIMETHYL S-2-(ETHYLSULFINYL)ETHYL PHOSPHOROTHIOATE, O,O- (83049)

Figure 4. EDNOHS Term Review Form.

(c) It can accommodate individual term families of up to 5000 terms, with up to a 30 level BT-NT hierarchy in each family.

(d) It allows for complete editing of cross reference logic prior to hierarchy building.

The input to this computer system consists of machine readable records (cards or tape) of one or more term families plus any existing thesaurus hierarchies developed from previous editions of EDNOHS.

The computer programs compare term cross references across all thesaurus families. If there are inconsistencies in the data, three types of error messages may be generated by HITS: (1) missing reciprocal cross references (these will be automatically generated in the course of the Preprocessor run), (2) conflicting term cross references, and (3) duplicate entries. Any logical error reports are analyzed in detail and corrections are made. The corrections, together with the original data, are again run through the HITS Preprocessor. If there are still error reports, the correction procedure is repeated until the data are acceptable.

The output of this procedural module is a set of logical and consistent term families, corrected and revised as indicated by HITS Preprocessor error reports, and in proper format to be processed by the HITS hierarchy generation program modules.

COMPUTER GENERATION AND DISPLAY OF COMPLETE HIERARCHIES AND BY-PRODUCTS

After the individual families (or groups of families) have been validated via the HITS preprocessor, they (in machine readable form) will be merged together, expanded into a complete thesaurus structure, and displayed via the seven main HITS programs.

The EDNOHS Chemical Thesaurus has its printed output displayed in two primary formats: an alphabetical, hierarchical

HIERARCHICAL THESAURUS

```
AROMATIC CHEMICAL COMPOUNDS (CONTD.)
  NT   ,,CUMENE
       ,,DIPHENYL
       ,,HYDROGENATED TERPHENYL
       ,,STYRENE
       ,,TERPHENYL
       ,,,TERPHENYL, META-
       ,,,TERPHENYL, ORTHO-
       ,,,TERPHENYL, PARA-
       ,,TOLUENE
       ,,TRIPHENYL METHANE
       ,,XYLENE
       ,,,XYLENE, META-
       ,,,XYLENE, ORTHO-
       ,,,XYLENE, PARA-
       ,POLYCYCLIC AROMATIC HYDROCARBONS
       ,,NAPHTHALENE
       ,,PHENANTHRENE
      AROMATIC KETONES
      ,ACETONAPHTHALENE, BETA-
      ,ACETOPHENONE
      ,ACETYLSALICYLIC ACID
      ,ANTHRONE
      ,BENZOIN
      ,BENZOIN CYANATE
      ,BENZOIN ETHYL ETHER
      ,BENZOIN ISOBUTYL ETHER
      ,BENZOPHENONE
      ,BENZYLIDENE ACETONE
      ,BROMPHENOL BLUE, SODIUM SALT
      ,CHLOROACETOPHENONE
      ,CHLOROBENZOPHENONE
      ,COLCHICINE
      ,DEOXYANISOIN
      ,DIHYDROXY-4-METHOXYBENZOPHENONE,  2,2'-
      ,DIOCTYLBENZOPHENONE
      ,DOXYCYCLINE
      ,HEXAMETHYL-6-ACETYLTETRAHYDRONAPHTHALENE,  1,1',2,4,4',7-
      ,HYDROXY-4-METHOXYBENZOPHENONE, 2-
      ,HYDROXY-4-OCTOXYBENZOPHENONE, 2-
      ,MERALEIN SODIUM
      ,METHOXYACETOPHENONE, PARA-
      ,METHYLACETOPHENONE
      ,NONYLPHENONE
      ,PROPIOPHENONE
      ,TETRAMETHYL-6-ETHYL-7-ACETYLTETRAHYDRONAPHTHALENE,  1,1',4,4'-
      ,THYMOLPHTHALEIN
      BROMCRESOL BLUE
      CYCLOHEXYLTHIO)PHTHALIMIDE, N-)-
```

PAGE 112 AROMAT 11/25/75

```
AROMATIC COMPOUNDS, MONOCYCLIC
 USE MONOCYCLIC AROMATIC COMPOUNDS

AROMATIC COMPOUNDS, POLYCYCLIC
 USE POLYCYCLIC AROMATIC COMPOUNDS

AROMATIC ESTERS
C,41,6,26
C,46,16
 BT  AROMATIC CHEMICAL COMPOUNDS
     CYCLIC ORGANIC COMPOUNDS
     ESTERS
     ORGANIC CHEMICAL COMPOUNDS

AROMATIC HYDROCARBONS
C .41.6.31
C .61.11.11
       (UNSATURATED CYCLIC HYDROCARBONS CONSISTING OF ONE OR
       MORE RINGS)
 UF  AH
     HYDROCARBONS, AROMATIC
 BT  AROMATIC CHEMICAL COMPOUNDS
     CYCLIC HYDROCARBONS
     CYCLIC ORGANIC COMPOUNDS
     HYDROCARBONS
     ORGANIC CHEMICAL COMPOUNDS
 NT  AROMATIC HYDROCARBONS ABOVE 371 C
     AROMATIC HYDROCARBONS 149 - 232C
     AROMATIC HYDROCARBONS 15.5 - 149C
     AROMATIC HYDROCARBONS 232 - 343C
     AROMATIC HYDROCARBONS 343 - 371C
     BICYCLIC AROMATIC HYDROCARBONS
     ,INDENE
     ,METHYLINDENES
     ,NAPHTHALENE
     C13-C20 AROMATIC HYDROCARBONS
     C20-C23 AROMATIC HYDROCARBONS
     C23 AND HIGHER AROMATIC HYDROCARBONS
     C6-C9 AROMATIC HYDROCARBONS
     C9-C13 AROMATIC HYDROCARBONS
     ISOPROPYLTOLUENE
     MONOCYCLIC AROMATIC HYDROCARBONS
     ,BENZENE
     ,CUMENE
     ,DIPHENYL
     ,HYDROGENATED TERPHENYL
     ,STYRENE
     ,TERPHENYL
     ,,TERPHENYL, META-
     ,,TERPHENYL, ORTHO-
     ,,TERPHENYL, PARA-
     ,TOLUENE
     ,TRIPHENYL METHANE
     ,XYLENE
```

Figure 5. Alphabetical hierarchical thesaurus.

THESAURUS - TREE STRUCTURES

ORGANIC CHEMICAL COMPOUNDS	C
ALDEHYDES	C.6
.ALICYCLIC ALDEHYDES	C.6.6
.ALIPHATIC ALDEHYDES	C.6.11
..ACETALDEHYDE	C.6.11.6
..ACETAMIDE	C.6.11.11
..ACROLEIN	C.6.11.16
..ALLANTOIN GALACTURONIC ACID	C.6.11.21
..BUTYRALDEHYDE	C.6.11.26
..CHLOROACETALDEHYDE	C.6.11.31
..CITRAL	C.6.11.36
..CITRONELLAL	C.6.11.41
..CROTONALDEHYDE	C.6.11.46
..DICHLOROACETALDEHYDE	C.6.11.51
..DODECANAL	C.6.11.56
..FORMALDEHYDE	C.6.11.61
...FORMALIN	C.6.11.61.6
...SUFONAMIDE-FORMALDEHYDE RESIN	C.6.11.61.11
..GLYOXAL	C.6.11.66
..HEPTANAL	C.6.11.71
..HYDROXYCITRONELLAL	C.6.11.76
..ISOBUTYL ALDEHYDE	C.6.11.81
..METHYLMERCAPTOPROPIONALDEHYDE	C.6.11.86
..NONYL ALDEHYDE, N-	C.6.11.91
..SUCCINALDEHYDE	C.6.11.96
..TRICHLOROACETALDEHYDE	C.6.11.101
..UNDECANAL	C.6.11.106
.AROMATIC ALDEHYDES	C.6.16

ORGANIC CHEMICAL COMPOUNDS (CONTD.)	
..AMINOBENZALDEHYDE, PARA-	C.6.16.6
..ANISALDEHYDE	C.6.16.11
..BENZALDEHYDE	C.6.16.16
..DIMETHYLAMINOBENZALDEHYDE, PARA-	C.6.16.21
..DIMETHYLAMINOBENZALDEHYDE, 1-	C.6.16.26
..ETHOXY-4-HYDROXYBENZALDEHYDE, 3-	C.6.16.31
..ISOETHYL PROTOCATECHUALDEHYDE	C.6.16.36
..PENTYLCINNAMALDEHYDE	C.6.16.41
..PHENYL ACETALDEHYDE	C.6.16.46
..SALICYLALDEHYDE	C.6.16.51
..VANILLIN	C.6.16.56
..BUTYL-ALPHA-METHYLHYDRO CINNAMALDEHYDE, PARA-TERT-	C.6.16.61
..CINNAMALDEHYDE	C.6.16.66
.GLUTARALDEHYDE	C.6.31
.HETEROCYCLIC ALDEHYDES	C.6.36
..BIS(2-BUTENYLENE)TETRAHYDROFURFURAL, 2,3,4,5-	C.6.36.6
..FURFURAL	C.6.36.11
..PIPERONAL	C.6.36.16
.ISOBUTYLCINNAMALDEHYDE, PARA-	C.6.41
.MELAMINE FORMALDEHYDE	C.6.46
..BUTYLATED MELAMINE FORMALDEHYDE	C.6.46.6
.THIOPHENE ALDEHYDE	C.6.51
ALICYCLIC COMPOUNDS	C.11
.ALICYCLIC ALCOHOLS	C.11.6
..CYCLOHEXANOL	C.11.6.6
..HEXAHYDROXYCYCLOHEXANE	C.11.6.11
..MENTHOL	C.11.6.16
..METHYLCYCLOHEXANOL	C.11.6.21

Figure 6. Thesaurus tree structures.

arrangement (Figure 5) and as family "tree structure" hierarchies (Figure 6). In the alphabetical section, all main terms are arranged alphabetically (letter by letter), with each term having all its cross references listed under it and the NT hierarchy indicated by progressive indentation. All nonpostable terms are interfiled alphabetically with the postable terms. In the "family tree" section, each family of terms is displayed under the broadest (head) term in the family hierarchical arrangement. Nonpostable terms do not appear in this section. An alphanumeric hierarchical decimal classification code is also automatically generated for each thesaurus term by the same computer program that builds the family tree structures.

In addition to producing the printed version of the EDNOHS thesaurus, HITS also produces a machine-readable thesaurus file which is employed to spin off several by-products which serve as index guides to the large chemical data base constructed as part of the NIOSH TNIC Project. These files include a Chemical Dictionary Listing of all chemical ingredients used to delineate products in the data base, a Carrier Code File showing the special NIOSH code numbers assigned to each thesaurus term, and a Tree Code File showing the automatically generated HITS decimal classification code for each term and comparing it to the NIOSH Carrier Code.

In the course of the three-year NIOSH Trade Name Ingredient Project, EDNOHS has been continually revised and expanded, with new editions appearing approximately five times a year. The current edition contains approximately 8000 main preferred (postable) chemical terms and about 4000 synonyms and occupies 1500 pages of computer printout in two-column format. This makes EDNOHS (to our knowledge) one of the largest fully cross-referenced hierarchical chemical thesauri ever built. It is anticipated that at the close of the TNIC project in April 1976, NIOSH will assume responsibility for the continued maintenance and growth of EDNOHS for, despite its current size, EDNOHS still has a great deal of room for growth. This is not surprising, for the UNISIST Thesaurus Guidelines state:

> "It should always be remembered that a thesaurus is never completed, its size and shape being a function of time".[8]

LITERATURE CITED

(1) H. B. Landau, W. L. Byer, and M. L. Neufeld, "Applications of Information Storage and Retrieval Techniques to the Development of Product Safety Data Bases", *Procs. ASIS Annu. Meeting*, **12**, 116–117 (1975).

(2) "American National Standard Guidelines for Thesaurus Structure Construction and Use" (Z39.19-1974), American National Standards Institute, New York, N.Y., 1974.

(3) "Thesaurus of Engineering and Scientific Terms", Engineers Joint Council, New York, N.Y., 1967 pp 673–679.

(4) H. B. Landau, M. Bullard, and A. Tamir, "The Hierarchical Indented Thesaurus System", *Procs. ASIS Annu. Meeting*, **8**, 373–380 (1971).

(5) D. Soergel, "Indexing Languages and Thesauri: Construction and Maintenance", Melville, Los Angeles, Calif., 1974, pp 78–81.

(6) F. A. Tate, "Handling Chemical Compounds in Information Systems", *Annu. Rev. Inf. Sci. Technol.*, **1**, 285–309 (1965).

(7) "Naming and Indexing of Chemical Substances for Chemical Abstracts during the Ninth Collective Period (1972–1976)", *Chem. Abstr.*, **76**, sect IV (Introduction to the Index Guide) (1972).

(8) "UNISIST Guidelines for the Establishment and Development of Monolingual Thesauri" (Document SC/WS/555), UNESCO, Paris, 1973, p 36.