Roy Rada
Computer Science Department,
University of Liverpool, England

Maintaining Thesauri and Metathesauri

Rada, R.: **Maintaining thesauri and metathesauri.**
Int. Classif. 17(1990) No. 3/4, p. 158 - 164,30 refs.

Maintaining a thesaurus is a time-consuming task which should go hand-in-hand with the indexing of information and should be supported by software. To connect different document databases their respective thesauri should be related. The most straightforward way to support this by computer is to map the terms of one thesaurus to those of another. Such a mapping creates one kind of metathesaurus. As citation systems are extended to include full-text online, a new thesaurus may be used to index individual paragraphs of a document, and a metathesaurus may apply to a universe of paragraphs. To illustrate these principles several computer systems are described which help people maintain thesauri and metathesauri. Particular success has been had by the National Library of Medicine with its Medical Subject Headings and its Unified Medical Language System.

(Author)

1. Introduction

The advent of computerized information systems has increased the importance and practicality of large-scale, well-developed thesauri. Furthermore, the communication networks in the world make connecting one thesaurus to another important. How can these thesaurus systems be maintained in a way that is supported by the computer? This paper addresses the software side of the thesaurus maintenance problem.

2. Thesauri

The effectiveness of a retrieval system is largely dependent upon the document classes in it. These classes may be based on terms from an indexing language (12). One of the most popular forms for an indexing language is a thesaurus. A thesaurus is a set of concepts in which each concept is represented with at least synonymous terms, broader concepts, narrower concepts, and related concepts (5).

2.1 Principles

In the development of a bibliographic citation system, one dilemma is whether to build a thesaurus first and then use it for indexing the documents or to index the documents by free terms and then construct the thesaurus after accumulating free terms. As is often true in life, the middle ground is particularly attractive, and that means in this case building a thesaurus and indexing documents hand-in-hand.

An institution which indexes many documents and maintains a large thesaurus may have a special staff for thesaurus maintenance. This thesaurus group consists of experts in the domain of the literature to be processed. These experts have a grasp of the terminology and semantic subtleties of the subject field. The thesaurus group is responsible for collecting index terms, and making the thesaurus as up-to-date as possible for the index group. The index group indexes the documents according to the latest version of the thesaurus and suggests new index terms to the thesaurus group.

2.2 MeSH Maintenance

At the National Library of Medicine, which indexes about 300,000 documents per year, about 100 people are devoted to indexing. About a dozen index terms are assigned to each document, but an indexer is expected to spend only 15 minutes on each document. The indexing staff is complemented by a thesaurus staff of about 5 people. The thesaurus, the Medical Subject Headings (MeSH), which they maintain includes about 15,000 main headings in an 8-level hierarchy.

While the instructions to indexers emphasize that they must base their indexing work on the whole document rather than on just the title and abstract, protocol analysis has revealed that in fact the indexers place a heavy emphasis on terms available in the title and abstract. The thesaurus staff also give great heed to the terms in the bibliographic citation (which contains the title and abstract but not the full article). Accordingly, indexing by mapping terms from the citation against the thesaurus can often produce correct index terms for the document (18).

The strategy for maintaining MeSH at the National Library of Medicine proceeds along three parallel paths:

* If an indexer does not find a needed concept in the thesaurus, then the indexer proposes a new term. The indexer uses a special paper form from the thesaurus group to indicate the details of the request and then a thesaurus specialist investigates the feasibility of this proposed change.
* The thesaurus specialists annually review the usage files for MEDLINE (the database for articles indexed with MeSH). If a MeSH concept is not used in indexing, then the concept is a candidate for removal from the thesaurus.

* The thesaurus group each year reviews one or more areas of biomedicine. For instance, recently the biotechnology sections and medical informatics sections of MeSH were reviewed and updated.

The largest number of changes which occur to the thesaurus traditionally originate from indexer suggestions. But whatever the route to a proposed change, a thesaurus specialist responds to the proposal and the response is reviewed both by the entire thesaurus group and by a Library-wide committee.

As recently as two years ago, the computer tools for MeSH maintenance at NLM were negligible. Every specialist had a computer terminal on his or her desk and would routinely access MEDLINE to discover the patterns of usage of a term in the literature, but the computer system did not directly cater to decisions about thesaurus updating. In order that the yearly update to MeSH occurs, a very difficult set of programs had to be invoked. Two members of staff would spend several months meticulously recording the approved changes to MeSH into the database and then preparing a publishable record of those changes for the publication of MeSH in paper form.

About 4 years ago a requirements document was developed for an improved thesaurus maintenance system. This document specified that the system should

* support the recording of information not unlike that on the set of forms already used,
* be either easily usable by all the people who normally were involved with MeSH forms or there should be a parallel paper system,
* produce data that made its way into the thesaurus files of the MEDLINE system, and
* produce the publications of MeSH.

A prototype system was first developed on dBASE III+ because it gave flexibility in database and interface options. To communicate with MEDLINE was, however, a problem and conversion routines were needed to get data to and from MEDLINE.

The Library's computer division was planning extensive use of the Model 204 database management system from Computer Corporation of America. The thesaurus group had hired temporary programmers to implement the thesaurus maintenance prototype, and the computer division was not in a position to support a dBASE III+ development. To reduce compatibility problems the decision was made to develop the final thesaurus maintenance system in Model 204.

The 'Model 204' thesaurus maintenance system has facilitated record keeping and has removed the months of difficult labor at the end of the MeSH year. The introductory material to the publications – those parts which summarize the changes to MeSH of the past year – are relatively easily produced now, whereas they were difficult to produce before the new system was used. But the new system has not replaced the use of paper forms in proposing changes nor in circulating information that is to be discussed at meetings.

The requirements document noted possibilities which have not been feasible to implement. It would be possible to circulate various requests for changes to MeSH on the computer and to record on the computer comments. The voting of individuals could be recorded as well so that decisions would become automatic.

Critiquing strategies about the structure of MeSH could be automatically invoked. At the moment MeSH has at the top-level over a dozen sections and some contain thousands of terms while others contain only a few hundred terms. When a very large thesaurus is accessed via a computer screen, the user may need to change the contents of the screen many times to find the terms for a query (27). By balancing the thesaurus the number of levels required to get to any given concept is less. The computer could suggest ways of improving balance.

Finally, additional relations might be added to the thesaurus. The MeSH system has been refined so that instances of hierarchical relations are distinguished. But a variety of non-hierarchical relations are still not part of the thesaurus. Theories of analogical inheritance could be applied by the computer to the thesaurus, if these non-hierarchical relations were present (15,22).

2.3 Other Thesaurus Strategies

The strategies for maintenance of thesauri at other institutions do not take wide advantage of the computer. At Elsevier Scientific Publishers, the Excerpta Medica indexers submit proposals to a senior editor who may then order thesaurus changes. At Derwent Scientific Publishers, procedures for changing the thesaurus are not formalized nor are there professional staff whose principal task is to maintain the thesaurus.

The market place for electronic directories has witnessed a wide variety of products over the past few years (25). The Yellow Pages provides product and service information to an enormous audience. Each advertisement may be viewed as a small document. In a typical scenario, a salesperson has a booklet which lists the index terms available for classifying Yellow Pages entries. The thesaurus only has two levels. When a salesperson needs a new thesaurus term to characterize a product or service, the salesperson sends a request to the editors of the Yellow Pages. An Electronic Yellow Pages allows access to information from the Yellow Pages in a more flexible way (4). To take adequate advantage of such an interface, a stronger thesaurus and thesaurus maintenance tools are needed.

When small, academic thesauri are made for particular projects, the methods are often less indexing-based and more committee-oriented (21). A small committee chooses terms based on a variety of unstated principles and produces a thesaurus by hand. If some principles of good thesaurus structure were more widely recognized, software tools to guide people in this kind of structuring might be more popular. Principles of balance and inheritance are not appreciated, and more work is required to demonstrate the value of such principles.

The major document database systems provide basic thesaurus update commands (2) but do not guide maintenance. BASIS, BRS/Search, and STATUS use the ANSI standard structure for a thesaurus. In STATUS one cannot relate one narrower term to two broader terms without repeating the narrower term. Maintenance of the BASIS thesaurus is done via the Basis Menu Language as are other manipulations of the database. BRS/Search provides menu-driven software for routine thesaurus maintenance, and STATUS provides a thesaurus handler for adding or deleting terms. With all three systems it is possible to load a thesaurus in batch mode. BRS/Search has a verification procedure to validate recipro cating relationships of terms in the thesaurus.

Several new systems have been recently developed which include sophisticated thesaurus maintenance procedures. The THSRS system can present thesauri graphically and advices users about changes to the thesaurus so that the number of criss-crossing lines in the graphical depiction is reduced (6). The PRESS system is intended to support the reuse of software documents and generates thesauri from a collection of indexing forms (1). Some systems avoid the traditional thesaurus but use linguistic and word-frequency tools to dynamically generate synonym classes that are comparable to those which one would want in a thesaurus (26).

3. Metathesauri

A metathesaurus transcends a set of thesauri. If people want to query several bibliographic citation systems simultaneously, a metathesaurus can help. Historically, there are many examples of metathesauri. In the 1960's, a connection between the Universal Decimal Classification and other classification systems was established (7). The Armed Services Technical Information Agency and the Atomic Energy Commission have linked their indexing languages (28). The National Library of Medicine has developed the Unified Medical Language System (UMLS) which connects many of the indexing languages of biomedicine. The problem is to formulate translation schemes that preserve the original meaning as much as possible, while allowing different groups to create or change their languages with as much autonomy as possible (13).

In order for all groups to be able to communicate using a common language, each group must either use the common language itself or be able to translate between the common language and the group language (see Figure 1). When no common language is shared by groups that wish to communicate (for example, English and German), each pair of groups must be able to have pairwise translations made into and out of each other's languages. When the common language is the same as all group languages (for example, Dewey Decimal Classification Scheme), the need for consensus is large. Since all groups have to agree on changes, the only changes that are easy for a group to make will be those that do not affect other groups or those that are valuable to all groups. With an external common language (for example, Esperanto), a group communicates with another by first translating its language into the common language, and then having the receiving group translate the common language into its own group language.

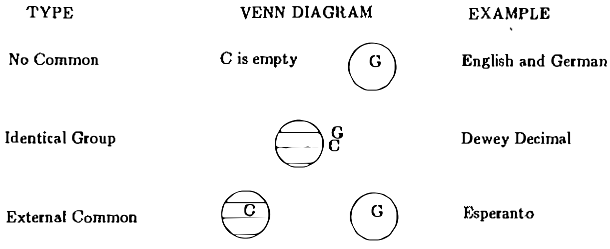

Figure 1. Group versus Common Language: G is one of the group languages. C is the common language.

3.1. Mapping

One way to create an internal common language is by mapping or matching. A variety of rearrangements of terms can be performed to determine whether a kind of syntactic match exists between two terms (17). For example, a computer program might determine that 'shortness of breath' is equivalent to 'breath shortness' based on a rule that removes the preposition and inverts the order (11).

Mapping is easily done by computer, and through this method a metathesaurus can be maintained. In the first instance a set of thesauri are connected at a central site by an automatic mapping. On an annual basis changes to the thesauri can be collected, and the mapping redone. While such automated mapping is clearly inferior in the quality of its product to what could be achieved with more knowledge-based methods, it is perhaps the only practical method to regularly updating a metathesaurus.

The International Classification of Diseases (ICD) is the World Health Organization's official classification scheme for world-wide morbidity and mortality. The Systematized Nomenclature of Medicine (SNOMED) is a thesaurus for patient records (30). EMTREE is the thesaurus used for indexing Excerpta Medica. In a project of the European Economic Community to connect patient records with the medical literature ICD and SNOMED have been mapped into EMTREE (23).

There are fifteen different facets (A through N and Q) in EMTREE and 17 different classification areas in ICD which are numbered 1 through 17. A randomly selected set of ICD terms came from thirteen different ICD classification areas. They mapped into three EMTREE facets: C, D, and G. Figure 2 shows the mapping of ICD terms to EMTREE facets. It is not surprising that an index matching of ICD terms to EMTREE would result in the vast majority of EMTREE terms being from the C facet, since the C facet includes Physical Diseases, Disorders and Abnormalities and the entire ICD tree relates to the classification of diseases.

Over 50% of SNOMED terms have been directly mapped to equivalent EMTREE terms with simple matching rules. First exact matching between SNOMED and EMTREE terms yielded EMTREE equivalents for 32% of the SNOMED terms. Replacing SNOMED terms with preferred term equivalents from EMTREE as defined in the online EMTREE synonym list yielded another 16% of the SNOMED terms. Reversing word order of unmatched SNOMED terms yielded another 4% of the SNOMED terms. If SNOMED and EMTREE terms were morphosemantically analyzed further matches would be found (19).

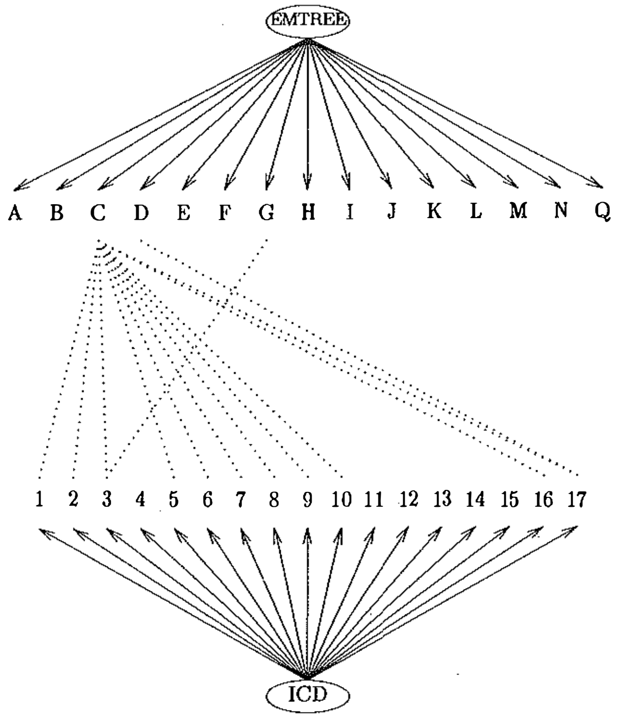

Figure 2: The dotted lines show the mapping of ICD to EMTREE at the top levels of the two tree structures.

An extension to mapping is merging. One type of merging is illustrated by the Vocabulary Switching System (VSS). The VSS contains the subject descriptors from 15 indexing languages in the areas of physical science, life science, social science, and business. With it, a user is able to automatically generate document requests in 15 indexing languages based on a request in just one (17). VSS utilizes five file types: term file, word file, stem file, stem phrase file, and concept file. The first four are inverted files, and these provide access to the fifth file. VSS has been shown to reduce search preparation time, improve search strategies and retrieval, and increase usage of existing databases.

3.2. UMLS

The Unified Medical Language System (UMLS) was formally initiated by Dr. Donald Lindberg of the National Library of Medicine in 1986. The first two years of the effort were devoted to development of a requirements document for the UMLS. The basic assumption is that (10) if the UMLS is to be improved, it must be used. Accordingly, the UMLS will be widely available. To increase availability, UMLS develop ments do not depend on any particular hardware or software systems.

UMLS has three fundamental components: an Information Sources Directory, a Metathesaurus, and a Semantic Net:

* The Information Sources Directory contains information about publicly available biomedical information resources. For each source the Directory contains the source's scope, vocabulary, syntax, rules, and access conditions.
* The Metathesaurus contains concepts from a variety of biomedical vocabularies. Each concept has a canonical representation that includes a semantic type. Mappings between different vocabularies are handled by this canonical representation.
* The Semantic Net represents relationships among the semantic types. For instance, 'Virus' and 'Organism' are semantic types, and they are related by the 'is-a' relationship. The semantic type 'Disease' is connected to 'Organism' via the 'caused-by' relation.

In the early stages of the Metathesaurus development generic frames were tried as a knowledge representation scheme (3) (8). The fundamentals of medical knowledge were to be manually encoded by experts into this Metathesaurus. After some trials, however, the inability to create a substantial, useful knowledge base meant another strategy was needed.

Lexical matching techniques offered a viable approach to an early prototype for the Metathesaurus. In January 1988, a UMLS research group at the University of California, San Francisco proposed a step-by-step approach to building a Metathesaurus by beginning with mapping (29). This first proposal became the plan for the first UMLS Metathesaurus prototype and is called Meta-1.

The purpose of Meta-1 is to assist computer information retrieval applications and to provide a browsable, machine-readable reference tool. Meta-1 contains terms from one or more of MeSH, DSM-III (a thesaurus of mental disorders), SNOMED, ICD, and CPT (a thesaurus of medical procedures). Where possible, the official name of a concept is the name used in MeSH. Meta-1 contains all MeSH and DSM-III terms and any from SNOMED, ICD, or CPT that lexically map to MeSH or DSM-III. The link between any two terms from different thesauri is classified as either an identical, a synonym, or a related link. Human intervention is required to properly determine whether a link is of the type 'related' or 'synonym'. After lexically based mappings were ascertained, subject experts reviewed and revised the mapping results.

Each Meta-1 record contains three types of information: 1) basic facts, 2) relationships, and 3) usage data. The basic facts include the name of the concept, the names of the thesauri in which the concept is located, the semantic type (such as Organism), and syntactic category (such as noun). The relationships which will be represented in a record include broader than, narrower than, synonym to, and lexical variant. For each term, the hierarchy from the root of a thesaurus to that term is explicitly recorded in the record for that term. For instance, the term 'coronary vessels' has a record whose relationship fields include two paths from anatomy in MeSH (see Figure 3). The explicit encoding of this information in each record allows for rapid retrieval and display of this hierarchical information.

Two Paths	
Anatomy	Anatomy
Cardiovascular System	Cardiovascular System
Arteries	Veins
Coronary Vessels	Coronary Vessels

Figure 3: Position of coronary vessels in MeSH hierarchies.

The usage data for MeSH terms is extensive and includes a list of the frequency with which a qualifier is used to modify the term in MEDLINE and a list of MeSH terms for other concepts that frequently occur in the indexing of the same articles. For instance, if the term 'coronary vessel' occurs 200 times with 'arteriosclerosis' and 300 times with 'aortocoronary bypass' in MEDLINE, then this information would be available in the 'coronary vessel' record.

Meta-1 is a massive information source that occupies about 100 megabytes. Much of this information occurs in the form of the explicit hierarchies recorded with each concept and the frequencies of occurrence that are recorded with each concept. An application of Meta-1 could represent the hierarchical information more succinctly but then require more computation time to display the hierarchy around any given concept. Due to its large size Meta-1 is distributed on a CD-ROM.

A HyperCard browser for Meta-1 has been developed and is distributed by the National Library of Medicine. To use that interface one must have a MacIntosh computer, but one can use Meta-1 without the HyperCard interface. The interface is not intended as the tool to connect Meta-1 to other applications but rather as a way to better see the information in Meta-1. The Meta-1 information can be accessed from the CD-ROM and utilized without the HyperCard browser. With the HyperCard browser one can point to a concept's relation to another concept and then see the new concept on the screen. Selecting a button labeled 'co-occurring terms' causes another window to open which displays the cooccurring terms in MEDLINE indexing.

CMIT is a knowledge base of medical diseases with several attributes for each disease (9). MeSH, SNOMED, and CMIT were mapped into a relational data model in one UMLS-related experiment. The overall algorithm for merging MeSH, CMIT, and SNOMED is:

1. Attempt lexical matches between SNOMED, MeSH, and CMIT disease names, including all combinations of synonyms and eponyms for a specific disease name.
2. Extract a dictionary of disease attributes, called semantic primitives, for each disease in CMIT using various parsing algorithms and use lexical matches between CMIT semantic primitives and MeSH and SNOMED synonyms to add synonyms for the semantic primitives to the dictionary.
3. Find all root-to-leaf paths in MeSH relation which contain two or more MeSH diseases mapped to CMIT diseases, then compute the intersection of the set of semantic primitives for the CMIT diseases in each root-to-leaf path. This intersection set describes the node inheritance properties of each root-to-leaf path. Use this set to decide what CMIT diseases to add to the MeSH hierarchy.
4. Repeat step 3 till no more diseases from CMIT are added to MeSH.

This algorithm is able to iteratively suggest a small set of diseases which may be connected into the metathesaurus via a narrower-than relation, based on the inheritance properties of nodes in a root-to-leaf path (20).

4. Individual Documents

Document reuse involves classifying existing document components and combining these components into a new document. A system called Many Using and Creating Hypertext (MUCH) has been developed to support this document reuse via a metathesaurus. Each generated document has an outline which has important relations to a thesaurus. The system supports collaborative work and indexing and gives guidance on the structure of the thesaurus. This approach to paragraph indexing and document reuse illustrates the superset approach to a common language. From a global common language users can acquire local languages.

The thesaurus model may be seen as a set of link objects. Each link object specifies a source node, link type, target node, and pointers to paragraphs. A group of link objects with the same source node form a frame (see Figure 4) (16). A node-link-node triple is a richer index for a paragraph than just a node. By placing paragraphs on the links, one can say that a paragraph is about a certain relation between two nodes – not just about a node. The range of allowed link types goes beyond the ANSI standard types.

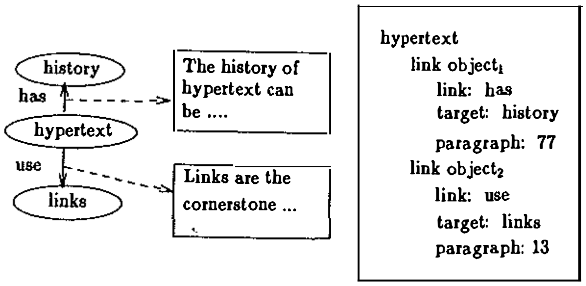

Figure 4: Graphical and frame representation are given side-by-side to show the frame representation.

Programs to help authors in dynamically generating various outlines from the thesaurus have proven invaluable in guiding the reuse of text and the generating of a cohesive, new document. The program traverses the thesaurus and prints a document from the paragraphs encountered during the traversal. The encountered terms become the outline of the document. A textbook has been automatically generated in the aforementioned way and been published.

A theory of outlines is being developed which relates to thesauri (14). While an outline is intended to reflect some model of the world that makes an impact on readers, a thesaurus and an outline serve the same purpose of organizing concepts into a hierarchy. An outline has an ordering among siblings which is not normally important in thesauri. Sometimes the character of the hierarchical relations in an outline manifest inheritance properties and thus reflect what should be true of thesaurus relations. At other times, the outline is like the high level of a thesaurus in that it identifies the facets of a topic rather than the hierarchical relations.

The ongoing experiments with the MUCH system occur on a network of Unix workstations with a prototype document reuse tool that has a relational database backend and an X-windows interface. The system supports both creation and accessing of hypertext and text. Users can register complex discussions and annotations within the system. Search facilities as well as browse facilities are available. Rough notes may be entered and do not need to be attached to the thesaurus. As all entries in the database are tagged with the name of the person who entered them and the date of entry, certain types of retrieval can be done independently of the thesaurus. Text-to-hypertext and hypertext-to-text tools have been incorporated in the prototype. The direct-manipulation interface allows users to select a link object and be shown a paragraph or conversely. The outline is dynamically generated, and users can select a topic in the outline and see the paragraphs associated with the corresponding link object. The role of the thesaurus is being explored in this new environment.

5. Conclusion

The long-term goal of the MUCH project is to create a thesaurus-based document system which facilitates all aspects of thesaurus and metathesaurus maintenance. The MUCH system also supports direct communication among people. While tools for automatically building thesauri from text are part of the MUCH armamentarium, the resulting thesauri are not adequate for many purposes. Painstaking human effort is required to develop a good thesaurus. To maintain a thesaurus is a difficult job that succeeds only when the thesaurus is regularly used and is changed in response to feedback from users.

A metathesaurus contains other thesauri – leaving them intact – and offers a rich, connected view across all the thesauri. The methodology for connecting thesauri must take advantage of the richness of information in the thesauri and not require substantial human labor. In this way users can be informed of the systematic strategies underlying the thesauri and modifications to thesauri can be readily recompiled into the system.

The directions to take in developing the thesaurus and metathesaurus approach to information are multiple. Directorate XIII-B3 of the European Economic Community has collected a list of about 1,000 thesauri in prominent use in the world. Connecting the thesauri that have been cataloged by DG XIII-B3 and allowing easy access by users could be an enormous benefit to those who want to share terminology. The Commission has outlined a five-step procedure to create this Metathesaurus which the Commission calls the 'Multifunction, Multilingual Thesaurus Database'. Major publishers are developing hypertext systems that include citation databases but also the full text of the documents which are cited (24). After the user has browsed a thesaurus and through it found relevant citations, he or she can elect to browse the relevant documents. The role of the thesaurus in browsing facilities for a single document is only partially understood.

Acknowledgements: The author worked at the National Library of Medicine several years ago but is indebted to colleagues there for an update which allowed this paper to be written. Doug Johnston and Peri Schuyler designed the Model 204 MeSH maintenance system. Peri Schuyler is in charge of the MeSH Section. Joe Hutchins was the systems analyst and principal programmer for the MeSH maintenance system. Doug and others in the MeSH section demonstrated Meta-1 to me. Betsy Humphreys is the project officer for the UMLS. The article by Betsy Humphreys and Donald Lindberg (10) provides an excellent description of the status of Meta-I and was the source for much of the information in this paper. However, no information in this paper can be construed as an official statement from the National Library of Medicine, and the views expressed here are solely the responsibility of this author.

References

(1) Albrechtsen, H.: Software Concepts: Knowledge Organization and the Human Interface. In: Tools for Knowledge Organization and the Human Interface . Frankfurt, Germany: INDEKS Verlag 1990, p. 48-63.
(2) Bain, M., Bland, R., Burnard, L., Duke, J., Edwards, C., Lindsey, D., Rossiter, N., Willet, P.,: Free Text Retrieval Systems: a Review and Evaluation. London: Taylor Graham 1989.
(3) Barr, C.E., Komorowski, H.J., Pattison-Gordon, E., Greenes, R.A.: Conceptual Modeling for the Unified Medical Language System. In: Proceedings of the 12th Annual Symposium on Computer Applications in Medical Care, 1988, p. 148-151.
(4) Cheng, E.: Let our Fingers do the Talking! Electronic Yellow Pages. British Telecom Journal, 7, Winter (1986/1987) No. 4, p.64-66.
(5) National Library and Information Associations Council: Guidelines for Thesaurus Structure, Construction, and Use, New York: American National Standards Institute, 1980.
(6). Craven, T.: Automatic Structure Modification in the Graphic Display of Thesauri. In: Tools for Knowledge Organization and the Human Interface. Frankfurt, Germany: INDEKS Verlag, 1990, p. 146-153.
(7) Dahlberg, I.: Conceptual Compatibility of Ordering Systems. International Classification, 10 (1983) No. 1, p. 5-8.
(8) Evans, D.A.: Pragmatically-structured, Lexical-semantic Knowledge Bases for Unified Medical Language Systems. In: Proceedings of the 12th Annual Symposium on Computer Applications in Medical Care, 1988, p. 169-173.
(9) Finkel, A., Gordon, B., Baker, M., Fanta, C.: Current Medical Information and Terminology. Chicago: American Medical Association, 1981.
(10) Humphreys, B.L., Lindberg, D.: Building the Unified Medical Language System. In: Proceedings of the 12th Annual Symposium on Computer Applications in Medical Care, 1989, p. 475-480.
(11) Klingbiel, W.: Phrase Structure Rewrite Systems in Information Retrieval. Information Processing and Management, 21, No. 2, (1985), p. 113-126.
(12) Lancaster, F.W.: Vocabulary Control for Information Retrieval. Washington, DC: Information Resources Press, 1972.
(13) Lee, J., Malone, T.W.: How Can Groups Communicate When They Use Different Languages? Translating between Partially Shared Type Hierarchies. In: Proceedings Conference on Office Information Systems, March 23-25, 1988, Palo Alto, California. New York: Association Computing Machinery, 1988, p. 23-29.
(14) Mili, H., Rada, R.: Medical Expertext as Regularity in Semantic Nets. Artificial Intelligence in Medicine (to appear 1990).
(15) Mili, H., Rada, R: Inheritance Generalized to Fuzzy Regularity. IEEE Transactions on Systems, Man, and Cybernetics, (to appear 1990).

(16) Minsky, M.: A Framework for Representing Knowledge. In: The Psychology of Computer Vision. Ed. Patrick Winston, New York: McGraw-Hill, 1975, p.211-277.

(17) Niehoff, R., Mack, G.: The Vocabulary Switching System, International Classification, 12 (1985) No. 1, p. 2-6.

(18) Rada, R., Evans, L.: Automated Problem Encoding System for Ambulatory Care," Computers and Biomedical Research, 12, (1979), p. 131-139.

(19) Rada, R., Lu, F., Eng, J., Brylawski, B.: Augmentation and Evaluation of a Medical Classification Structure Through Morphosemantic Analysis. In: Proceedings MEDINFO '86, Amsterdam, NL North-Holland: 1986, p. 1096-1100.

(20) Rada, R., Martin, B.: Augmenting Thesauri for Information Systems. ACM Transactions on Office Information Systems, 5, (1987), No. 4, p. 378-392.

(21) Rada, R., Blum, B., Calhoun, E., Mili, H., Orthner, H., Singer, S.: A Vocabulary for Medical Informatics. Computers and Biomedical Research 20 (1987), p. 244-263.

(22) Rada, R., Mili, H.: A Knowledge-Intensive Learning System for Document Retrieval. In: Knowledge Reorganization and Machine Learning, Ed. Katharina Morik, New York: Springer-Verlag, 1989, p. 65-87.

(23) Rada, R., Zanstra, P., Potharst, J., Barlow, J., de Vries Robbe, P., Bijstra, D.: Expertext for Medical Care and Literature Retrieval. In: Artificial Intelligence in Medicine (in press, 1990.

(24) Rada, R.: Hypertext: from Text to Expertext, London: McGraw-Hill, 1991.

(25) LINK Resources: Electronic Directories. Research Report LINK#0I84, LINK Resources Corporation, 79 Fifth Avenue, New York, NY 10003, September 1987.

(26) Ruge, G., Schwarz, C,: Linguistically Based Term Associations: A New Semantic Component for a Hypertext System. In: Tools for Knowledge Organization and the Human Interface, Frankfurt, Germany: INDEKS Verlag, 1990, p. 88-95.

(27) Seppala P., Salvendy, G.: Impact of Depth of Menu Hierarchy on Performance Effectiveness in a Supervisory Task: Computerized Flexible Manufacturing System. Human Factors, 27 (1985) No. 6, p. 713-722.

(28) Svenonius, E.: Compatibility of Retrieval Languages: Introduction to a Forum. In: International Classification, 10 (1983) No. 1, p. 2-4.

(29) Tuttle, M.S., Blois, M.S., Erlbaum, M.S., Nelson, S.J., Sherertz, D.D.: Toward a Biomedical Thesaurus: Building the foundation of the UMLS. In: Proceedings of the 12th Annual Symposium on Computer Applications in Medical Care, 1988, p. 193.

(30) Wingert, F.: An Indexing System for SNOMED. Methods of Information in Medicine, 25 (1986) No. 1, p. 22–30.