Access to Biomedical Information: The Unified Medical Language System

STEVEN J. SQUIRES

ABSTRACT

THE NATIONAL LIBRARY OF MEDICINE (NLM) is engaged in a long-term project to develop a Unified Medical Language System (UMLS) that will retrieve and integrate information from a variety of information resources. Two UMLS components use fundamental aspects of controlled vocabulary structure and management and their relationship to information retrieval that have general interest for librarianship. The UMLS project is described along with its initial deployment in retrieval environments.

INTRODUCTION

Bibliographic control of information has traditionally focused on locating and describing published documents, and indexing these in useful ways. In every subject domain, the problem of erecting a complete record of existing information is more or less acute, depending on the available support for and interest in comprehensive collections and provision of access. In the biomedical domain, due to its societal importance and to generous government support, the problem of finding and describing the published literature is not great in spite of the size of that literature. As chronicled by Adams (1981) and, more recently, as listed by Tilley (1990), massive government and private efforts are in place for building and maintaining bibliographic and reference databases and online systems that describe and index the biomedical literature.

More and more, however, important information has developed in forms other than the published record. In biomedicine, these include clinical databases and patient records. In these databases, the mechanisms for record creation, maintenance, access, and exchange are not as structured as for bibliographic data. The focus of bibliographic control has had to include describing and structuring records and retrieval tools that permit effective use of information in a large number of diverse information sources.

In spite of the ability of machines to search on any element of stored data, controlled vocabularies are still widely used to index information and to produce effective retrieval. Many different terminologies exist, even within the same subject domain, that have been created to organize and retrieve data for specific purposes. The Unified Medical Language System (UMLS) is conceived as a means of navigating among a disparate array of databases organized using different terminologies. Except perhaps for work in automated indexing, to which the UMLS is not unrelated, this effort is possibly the most important development in biomedical bibliographic control in recent years. This article will describe UMLS components, their potential uses, and some current efforts to incorporate them into retrieval environments. Efforts to evaluate UMLS are noted along with areas for future development.

PURPOSE OF THE UNIFIED MEDICAL LANGUAGE SYSTEM

In the mid-1980s, as the growth and development of electronic means of storing information progressed, and as computational and telecommunications resources for using that information proliferated, the National Library of Medicine recognized a need to assist the biomedical world in using the new resources and capabilities now more or less easily at hand. Its *Long Range Plan* of 1986 presents a comprehensive program of research, resource development, and educational endeavors to provide that assistance. A central part of that plan is the Unified Medical Language System.

Humphreys and Lindberg (1989) and Lindberg and Humphreys (1990) make the case for a UMLS. Their argument starts with the observations that useful biomedical information can be found among an increasingly large number of machine-readable databases, that these databases are different in important ways, and that these differences are among the barriers to effective use. Databases differ by content and by how that content is represented and described. They also differ by means of access. As users are confronted by the ever larger array of different databases, it is increasingly difficult to identify which databases have information relevant to a particular query. Users, too, have different ways of expressing the many concepts

represented in databases and, as a result, formulate queries about those concepts differently. There is a lack of a universally recognized and accepted standard vocabulary for expressing biomedical phenomena and for recording health care events and transactions. Once information is found in a database, the need arises to organize it and possibly evaluate it for its intended use. The UMLS is meant to compensate for these problems, not by imposing uniformity on the diverse world of terminology and databases, but by minimizing the differences about which a user of information sources has to be aware (Lindberg & Humphreys, 1990, p. 121).

These problems are, of course, not new in the information world. Perhaps the most important aspect of the UMLS approach is its "unified" nature, its attempt to provide a single utility through which access to the variety of biomedical databases can be gained, and by which information from them can be easily retrieved and integrated.

DEVELOPMENT OF THE UMLS

The UMLS project was initiated in 1986 by two years of investigation involving research at NLM and research contracts awarded to academic institutions. The initial research resulted in decisions to create three new knowledge sources: a Metathesaurus, a Semantic Network, and an Information Sources Map. The Metathesaurus combines and integrates existing biomedical nomenclatures and relates them to each other. The Semantic Network is a scheme of general categories to organize the terms of the Metathesaurus. The Information Sources Map is a directory of information about biomedical databases that will support source selection and automatic connection and retrieval from them. Each of these will be described in some detail later.

The next three years saw the creation and testing of the three knowledge sources. In 1990, the first versions of the Metathesaurus and the Semantic Network were issued. Revised versions of the first two knowledge sources appeared in 1991 along with the first version of the Information Sources Map. New versions of each component are anticipated annually. The components are issued in multiple formats and in both unit record form and relational form (Cimino et al., 1992, p. 1502). They can be used in MS-DOS, Macintosh HFS, and UNIX environments. The Macintosh version includes MetaCard, a hypercard application for browsing the Metathesaurus (Sherertz et al., 1989), and NET, a graphical browser for the Semantic Network, to facilitate use.[1]

Interest will naturally focus on the new knowledge sources and their eventual use in a fully developed UMLS, but the development strategy itself and the extent of its success should also be evaluated

as an example of cooperative endeavor shared by the public and private sectors. To encourage experimentation and feedback, the UMLS components are being made available free of charge to anyone agreeing contractually to provide feedback and suggestions for improvement (Humphreys & Lindberg, 1992, p. 1496). Based on this feedback, the components will undergo iterative development but will be available in successive if incomplete stages for use. As of January 1993, more than 300 institutions and individuals had asked for copies. Included were universities, hospitals, government research centers or health care agencies, and commercial companies. About 20 percent of the recipients are outside the United States. The usefulness of the UMLS components will be explored by copy recipients in a variety of patient care, medical education, library service, and research environments. Specific applications will involve indexing and coding of data, knowledge representation, natural language processing, user interface development, and information retrieval from multiple databases.

Reports about the project from NLM and from experimental users have appeared since its inception and have included conceptual discussions underlying the creation and content of the knowledge sources, descriptions of the knowledge sources as they appeared, and reports of experimental uses to which the knowledge sources have been put. These reports have largely been made at medical informatics conferences (Annual Symposium on Computer Applications in Medical Care, World Congress on Medical Informatics) and in medical computer journals and not in the general library literature. However, UMLS development makes use of, and has general application for, fundamental concepts about thesaurus construction and the organization of information by means of controlled vocabulary and classification that have long been a part of standard library practices. The UMLS developers have had to confront and resolve problems involving the distinction between word and phrase, the notion of concept, the definition of synonymy, the effectiveness of pre-coordination versus post-coordination, organization through hierarchy, the "relatedness" of concepts, and the usefulness of all of these both for retrieving information and for creating knowledge sources. Lessons learned in this effort could influence nomenclatures and their use everywhere.

It is important to recognize that the knowledge sources being created do not, by themselves, constitute a UMLS system. They are only tools to be exploited by systems developers who, in response to local needs or for enterprising reasons, will create functional components. For example, Lindberg and Humphreys (1989) include among possible functional components a query interpreter capable

of using natural language understanding systems to translate the natural language of user queries or from clinical records into standard expressions. For the next steps in an information quest, a search formulator and transmitter could turn a query into search statements appropriate for a chosen database and would then communicate the search to the database. Following that, information retrieved from databases could, through an output processor, be merged, ranked, and displayed according to parameters defined by the user. Any such components would interact with the UMLS knowledge sources in appropriate ways.

THE UMLS METATHESAURUS

The first UMLS knowledge source to be created was the Metathesaurus (Meta). As with the UMLS as a whole, it is equally important to realize what the metathesaurus is not. Though its name might imply otherwise, it is not created to be a monolithic universally accepted vocabulary to replace all existing biomedical nomenclatures. It is rather a synthesis of existing vocabularies, achieved by linking, merging, and integrating them. Using existing vocabularies gives to Meta an empirical grounding. The Metathesaurus is comprised of biomedical terminology "as it is used" (Tuttle et al., 1988). Thus, Meta endeavors to represent only the meanings of terms that are implicit in the sources from which it was constructed. This means preserving the contexts established for those meanings by the source vocabularies in their structures, including the use of definitions, hierarchies, and other term relationships.

Integrating the thesauri serves two further purposes (Bicknell et al., 1988). It maps them to one another, thereby creating pointers from every concept in the separate thesauri to the most appropriate equivalent concept in the others. This addresses the UMLS objective of translating a user query into a search strategy for a given database that is indexed by a given thesaurus and making this process transparent to the user. Integration also merges the thesauri, thereby creating a more comprehensive knowledge base with a deep level of synonymy. This addresses the UMLS goal of providing an adequate knowledge base for interpreting natural language user queries and linking those queries with appropriate databases.

The biomedical terminologies chosen for integration into the first two versions of the Metathesaurus fell into two sets (Lindberg & Humphreys, 1990, p. 123). The first set included *Medical Subject Headings* (MeSH), *Diagnostic and Statistical Manual of Mental Disorders* (3d rev. ed., American Psychiatric Association, 1987), and the 400 most frequently used terms representing clinical problems and manifestations in clinical records at three COSTAR (Computer

Stored Ambulatory Record) sites. All nonequivalent terms from these
sources became the base set of terms in the Metathesaurus. Terms
from the second set of sources that could be related to the base set
were then included. Thus, not all terms from these vocabularies
became Meta entries. This second set of sources were the *Systematized
Nomenclature of Medicine* (SNOMED) (2d ed., College of American
Pathologists, 1986), the *International Classification of Diseases, 9th
Edition, Clinical Modification* (2d rev. ed., Washington, DC, 1990),
and *Physicians' Current Procedural Terminology* (4th ed., American
Medical Association, 1989). Finally, the first two versions of Meta
included selected terms from *Library of Congress Subject Headings*
mapped to MeSH by NLM staff. Successive iterations of the
Metathesaurus will include fuller integration of all source
vocabularies.

BUILDING THE METATHESAURUS

The means by which the selected thesauri were integrated
incorporated semi-automated lexical matching combined with
knowledge of the relationships among terms explicit in the structures
of the source vocabularies. Tuttle et al. (1988, 1989), Sperzel and Tuttle
(1989), and Sherertz et al. (1989b) demonstrate the utility of automated
lexical matching for finding equivalencies among a diverse set of
vocabularies. Machine versions of the source vocabularies were
obtained and the terms from those vocabularies were expressed in
a uniform manner to facilitate lexical matching among them. The
terms from the first set of source vocabularies were compared and
a single preferred term, or canonical term, was selected for any
identical terms, lexically variant terms, or lexically variant synonyms
of terms. Lexical variants can be terms that are different only because
of case, number, word order, spelling, or punctuation. When terms
from different source vocabularies were found to be identical, or only
lexically variant forms of one another, the preferred term for the
Metathesaurus entry was established by an order of precedence. If
a term from among a set of identical or lexically variant terms was
a MeSH term, that term became the Meta entry. The vocabularies
following MeSH in order of precedence were DWM-IIIR, SNOMED,
ICD, CPT, LCSH, and COSTAR. Once a canonical term was de-
termined, other terms from the set of equivalent terms could be labeled
as lexical variants, synonyms, or lexical variants of synonyms.

Though the term relationships and information about terms that
result from the processes described earlier and the human editing
that followed are stored in several relational database files, a database
management system could be devised to present all the information
about a single term as a comprehensive entry or record for that term.

This conceptual record structure of Meta would then consist of entries for concepts with fields or slots that contain terms related to the concepts or that describe or name attributes of concepts. Tuttle et al. (1989) enumerated the essential slots, as follows:

  Concept Name
  Meta Unique Code(s)
  Syntactic Category (part of speech)
  Lexical Tag (if term is an abbreviation, acronym, etc.)
  Semantic Type (assigned from Semantic Network)
  Source Vocabulary or Vocabularies
  Source Hierarchical Contexts
  Source Definition(s)
  Lexical Variants
  Synonyms
  Related Terms
  Broader Terms
  Narrower Terms

Other attributes of terms include use data, described later, data necessary for thesaurus maintenance, and, if the Meta term is a MeSH term, up to twenty-five data elements derived from the annotations in the MeSH vocabulary.

After the first version of Meta was compiled, the result was subjected to human editing, described by Sperzel et al. (1990). Semantic types and lexical categories were assigned at this step. Editors also evaluated the automated assignments of synonyms, related terms, broader or narrower terms, and lexical variants if these appeared obviously incorrect. The results of human editing had to have their own audit trail, so that new versions of the Metathesaurus computed from updated versions of the original source vocabularies would have the desired result (Sherertz et al., 1990). Tuttle et al. (1992) warn local users of Meta about the consequences of adding local terms to it, since these enhancements would have to be maintained over new releases. He calls for a standard updating method generally adopted that would facilitate both local maintenance and Meta improvement.

Three versions of Meta have been released to date (Meta-1, Meta-1.1, Meta-1.2). The number of concepts grew from approximately 63,000 in Meta-1 to more than 67,000 in Meta-1.1, to approximately 130,000 in Meta-1.2 that was issued in October 1992. Whereas the first two versions contained three kinds of entries—for concepts, related terms, and synonymous terms—the third version has only a single kind of entry, that for concept. Responding to feedback from the first two releases, Meta developers structured the information in Meta to simplify its extraction and manipulation and to make it

easier to conceive of Meta as an abstraction (Tuttle et al., 1993, p. 301). The data that were once distributed in more than fifty separate tables are, in the third version, distributed among only twelve tables. The twelve tables fall into four categories: concepts (one table), relations between concepts (two tables), attributes of concepts (eight tables), and a word-based index Metathesaurus string. In the future, the simplified format may be expressed using the ASN.1 (Abstract Syntax Notation 1) standard, part of the Open System Interconnection Standard. It is hoped that adopting this standard will encourage its use, particularly among the source vocabulary developers, thereby facilitating future collaboration and Meta enhancement (p. 303).

USE DATA IN THE METATHESAURUS

One other attribute of Meta entries remains to be discussed. For those terms in Meta derived from MeSH, "use" data have been gathered and recorded in Meta. This is perhaps the most unique kind of term information in the Metathesaurus. These data may be of three types: occurrence data, data on subheadings used with a term, and co-occurrence data (Humphreys & Schuyler, in press). These kinds of data were originally only computed from MEDLINE but now are included for PDQ (Physician's Data Query), OMIM (Online Mendelian Inheritance in Man), DxPlain, and QMR (Quick Medical Reference). Occurrence data consist of the number of citations to articles in which the concept was a main point, thereby representing that a specific concept is present in an information source and to what degree. Subheading data list which MeSH subheadings have been applied to the concept in indexed citations and the frequency with which each subheading was applied. Such information provides insight into the important separate aspects of a concept. Co-occurrence data record the number of citations in which two terms co-occur as primary concepts. Such data tell us that two terms have been used together in a database and with what frequency. In the Metathesaurus, these co-occurring terms to a concept can be arranged by the semantic type of the co-occurring terms.

Co-occurring data are another example of the empirical nature of Metathesaurus data. That two concepts have been observed in some context to occur together suggests that a relationship exists between them (Nelson et al., 1992, p. 212). That this information has some potential usefulness in information retrieval is suggested by the fact that only a very small percentage of the possible co-occurrences of MeSH terms have actually occurred. A potential use of the different sets of use data will be in determining in what database a query is likely to be successful. Subheading information can be used in interactions with a user to focus or to expand a query (Humphreys

& Schuyler, in press). As effective retrieval use is made of this kind of data, future versions of the Metathesaurus will tabulate use data for more databases, and local implementers may want to add it for local databases or hospital patient records.

Some experimenters have already found interesting applications for the use data. Merz et al. (1993) have developed Question & Answers (Q & A), a system that matches user query terms to Meta to find appropriate retrieval terms and then uses the occurrence data of the Meta terms to estimate the number of articles that would be retrieved by the terms. Depending on this estimate, the program may attempt to improve the search strategy. Q & A is thus a procedure that interacts with Meta for performing retrieval estimation and query refinement before a search.

Miller et al. (1993, p. 88) make use of co-occurrence data to link terms found in patient charts to bibliographic retrieval in MEDLINE. Words from a patient chart are matched to Meta terms. For each identified Meta term so matched, its co-occurring terms are then matched to the other Meta terms from the chart. The list of terms resulting from this two-step matching process represents terms that both appear in the chart and are related via MEDLINE indexing. Searches using these terms are guaranteed to produce retrieval.

THE UMLS SEMANTIC NETWORK

Each concept in the Metathesaurus derives one of its attributes from the second of the UMLS knowledge sources, the Semantic Network. The purpose of this component is to provide a consistent categorization of all concepts in the Metathesaurus and to supply a set of useful relationships among them (McCray, 1989, p. 504). It defines the types of categories to which all concepts in Meta can be assigned and the permissible relationships that can exist between the types.

The work of Miller et al. (1988a, 1988b) established the utility of semantic relationships in medical bibliographic retrieval. They found that vocabularies capture some semantic relationships, in pre-coordinated terms or by applying term subheadings, but that many more are possible. Individual terms may be present in documents or in their indexing that may yield the desired bibliographic retrieval, but often a specific relationship between terms most precisely expresses the topic. Some ability to make use of these relationships in bibliographic retrieval is desirable. Good information retrieval is limited by the lack of such an ability and it is one not handled by Boolean capability (Appel et al., 1988, p. 152). Knowing how two terms are related to each other in a document allows specification of a topic in ways unexpressed by the mere co-existence of terms

or even by terms in close proximity. Also, knowing which relationship between two terms that is operable in each document of a set of documents in which the two terms occur can usefully partition that set (Miller et al., 1988). In the face of an acute need for improving retrieval relevance, pursuit of term relationships as a retrieval device is promising. It is to this goal that the Semantic Network is addressed.

The primary relationship between terms in Meta is a hierarchical one, sometimes called the "is-a" link. Therefore, the types of the semantic network are arranged in a hierarchy. The value of such a structure lies in the inheritance property (McCray, 1989, p. 504). By this property, a category is understood to inherit the "is-a" relationship to each category higher than itself in the hierarchy. A computational advantage of this feature of hierarchical organization is one of efficiency; information about terms or categories found at higher levels need not be repeated at lower levels.

To take full advantage of possible term relationships in information retrieval requires that other nonhierarchical relationships also be identified. Here is where the UMLS differs from other nomenclatures, where only the "is-a" relationship is implied or where associated terms are merely identified as being related. The relationships that are important to define depend on the subject domain of the vocabulary. Therefore, not all possible linguistic relationships between semantic types are in the Semantic Net (McCray & Hole, 1990). Some of the relationships important for the biomedical domain are illustrated by the following general formulations: A causes B, A is a process of B, A is a property of B, A uses B, A treats B, A is exhibited by B, A evaluates B. These relationships will be organized in the Semantic Net into four broad categories: physical relationships (e.g., part of, consists of, contains), temporal relationships (precedes, co-occurs with), functional relationships (causes, produces, affects), and conceptual relationships (measures, assesses) (McCray, 1989, p. 505). In contrast to the hierarchical "is-a" relationship, these other relationships are not necessarily inherited and they may not even hold between any two instances of semantic type terms.

As with terms in the Metathesaurus, types in the semantic net constitute conceptual entries or records, having fields that define them. The slots in each record include the name and system identifier for the type, a positional number from the hierarchy of types, a definition, and the hierarchical links to a type's parents or children. If the semantic type is itself a relationship, the slots comprise the name of the relation, the name of the inverse of the relation, a definition, and the semantic types that can be linked by this relation (Lindberg & Humphreys, 1990, p. 124).

Following are examples of records for a semantic type and a semantic type relationship (McCray & Hole, 1990):

Type: Animal

Identifier: T008

Position: A.6

Definition: An organism with eukaryotic cells, lacking stiff cell walls, ...

Links: is-a Organism; inverse is-a Invertebrate; inverse is-a Vertebrate

Relation: exhibits

Identifier: T136

Inverse Relation: exhibited-by

Definition: Shows or demonstrates

Links: is-a conceptually-related-to

Type Links: Organism-Behavior (p. 130)

The first version of the Semantic Network consisted of 133 types. The 1992 version contains 134 types and 47 relationships. A semantic network of this small size consists of categories that are necessarily broad in scope, and it may consequently have limited effectiveness. The types are assigned, after all, to over 130,000 Meta terms. For example, looking at the set of terms having the same semantic type may not be very useful. The broad scope may have been necessary because the biomedical domain itself is wide, including not only the life sciences and their medical application, but also the social, economic, and demographic aspects of health care delivery. Another characteristic of the network is that the depth of its hierarchy varies— i.e., some categories are not subdivided to the degree possible. The need for hierarchical depth varies, again, depending on the domain. Knowing a term's semantic type along with the level that type occupies in a conceptual hierarchy gives some clues about a term's importance and meaning relative to its domain. It remains to be seen whether hierarchical scope and depth can be usefully exploited for information retrieval.

The possible uses for the information embodied in the Semantic Network are easy to imagine but difficult to realize in an automated environment and await software development for full exploitation. The semantic types should, nevertheless, make it possible for computer systems to organize biomedical concepts effectively and to reason about the possible and probable relationships among different types of concepts (Lindberg & Humphreys, 1989). Such systems may employ linguistic parsing techniques for automated indexing of biomedical literature or automated analysis of clinical data. Use of the Semantic Network may assist in query formulation, interactive query refinement, or simply graphical browsing of the Metathesaurus (McCray & Hole, 1990).

Future editions of the Semantic Network may see additional relationships among existing semantic types and the addition of new

semantic types. Adding new semantic types naturally involves possible reassignment of semantic types to Meta concepts, increasing the complexity of the updating program.

Bishop and Ewing (1992) have suggested that the UMLS developers missed an important link between the Metathesaurus and the Semantic Network by not relating Meta unique term identifiers to semantic types in some way. Meta does not use a hierarchical coding scheme, as do most of the vocabularies being merged. Coding of concepts is helpful for establishing consistent recognition of concepts as opposed to the names for concepts (names may change over time and among environments), for recognizing concepts in different languages, and for efficiently maintaining compatibility between systems. Meta uses random coding, that is, a coding that carries no information about how one concept is related to another. Were the coding itself to reflect in some way the hierarchical relationships among terms as represented by the semantic types, users could more easily extract classes of concepts. Bishop and Ewing (1992) further suggest that, even though most existing coding schemes differ stylistically, the hierarchies they are based on are quite similar and could form the basis of an ideal arrangement of medical knowledge for the future. The UMLS developers might have extended the idea of thesaurus integration, with its empirical founding, to the Semantic Network that purports to organize the concepts in them.

THE PROPERTY OF SEMANTIC LOCALITY

Because it functions as a thesaurus, the Metathesaurus is fundamentally a device for organizing meanings. It should do this in a way that permits a user seeking a term for a meaning to find that term by navigating its relationships to other terms. A thesaurus makes this possible by giving to each of its terms a semantic locality. Semantic locality has other uses as well, including establishing what is generally relevant to a given concept, or what may be relevant in a given situation (Nelson et al., 1992, p. 213).

The provision of semantic locality in the UMLS is particularly generous and goes beyond other thesauri. It is provided by semantic types; by term information that includes synonyms, related terms, and lexical variants; by co-occurrence data; and by contextual data, or, a term's parents, siblings, and children derived from the source vocabularies (Nelson et al., 1992, p. 210). Even though the contextual data may differ or even conflict, the differences may reflect the different intentions and viewpoints of the source vocabularies, and may have value as such. All of these elements help to establish meanings, locate more general and more specific terms, and find potentially useful relationships between terms.

Nelson et al. (1993) attempted to evaluate the semantic locality of UMLS by observing its completeness and redundancy. Redundancy could be said to be high if two Meta terms were found as related terms, as co-occurring terms, and as hierarchically related terms (parent-child or siblings). That the degree of overlap was found to be small suggests that each of the dimensions of semantic locality is unique and valuable (p. 653). To get an idea of whether expected important relationships among terms were indeed expressed by Meta's semantic locality, the authors tested whether concepts linked to an entry by their presence in the definition of that entry were related by other dimensions of semantic locality. Here, only slightly better than half of the direct links between concepts in definitions and the entries they defined were found to exist. Also, about one-fifth of the important concepts in definitions had no corresponding Meta entry. This suggests that enhancements to Meta have to include not only new terms, but also, and just as importantly, new relationships between terms other than those provided by their source thesauri.

INFORMATION SOURCES MAP

The third UMLS knowledge source is called the Information Sources Map. According to Masys and Humphreys (1992), it will address the problem of determining which electronic information sources may be relevant to particular questions and will assist a UMLS user in accessing and using the information found in them. It is comprised of records describing electronic information sources, supplying information on each source's scope, type of information (citation, full text, reference text), language, size, probable utility, access conditions, and updating schedule. In order to make an automatic connection and conduct a successful search, it will contain the data element definitions of each source and the scripts necessary for traversing the communications paths to them. Also, data elements in different information sources which may contain the same value are noted. This information should be useful in retrieval programs designed to conduct automated searches for the same information among multiple information sources.

The first version of the Information Sources Map contains fifty information source records, including those for all the NLM databases, DxPlain from Massachusetts General Hospital, QMR (Quick Medical Reference) from CAMDAT Corporation, and OMIM (Online Mendelian Inheritance in Man) from Johns Hopkins University. Local users of this UMLS component can add records for their own local databases.

The information source records will be indexed using MeSH terms and subheadings, semantic types from the UMLS Semantic

Network, and relations between pairs of semantic types (Semantic Type Relations). Information about relationships among the different information sources will be noted, such as that a database is a superset of, a subset of, or contributes to another.

Masys (1993) has attempted to evaluate the indexing of the records in the Information Sources Map in terms of recall and precision. Fifty clinical medicine queries were translated into MeSH terminology and linked through the Metathesaurus to their semantic types and semantic type relations. The resulting list of semantic types and relations were matched to the indexing terms in the Information Sources Map. Optimal results would have matched queries with all the appropriate databases for searching them. It is not surprising that very high recall or very high precision was achieved only at the expense of the other. The study noted that elements other than the indexing terms that describe the databases of the ISM may prove useful additional filters for matching databases to specific queries. Additional elements include those that denote the intended user audience or the type of content in a database.

Miller et al. (1992), after building a prototype ISM, concluded that additional coding of information sources for "axes of use" would help match resources with queries. Such axes of use included whether the resource was commonly used, possibly used, or unlikely to be used for patient care, clinical research, basic research, or health services research, and whether its coverage in those areas was at a slight, intermediate, or comprehensive level. He went on to say that other possibly useful descriptive attributes to include in the records for sources would be the depth of the material likely to be found in a source (review level, reference level, consensus report level), and the type of content in the source (bibliographic, textual, patient records, directory). In fact, the 1992 version of the Information Sources Map includes axes of use and type of content information.

Following are the textual values of some of the fields in the ISM record for the Developmental and Reproductive Toxicology Database (not the complete record):

Name:	Developmental and Reproductive Toxicology Database
Producer:	National Library of Medicine
Alternate Names:	DART, DAR
Type of Content:	Bibliographic Database
	Indexed Citations
MeSH Indexing:	Abnormalities—chem. induced, epidemiol., etiol.
	Abortion—etiol.

	Alcoholic Intoxication
	Fetal Development—drug effects, radiation effects
	Maternal-Fetal Exchange
	Prenatal Exposure Delayed Effects
Semantic Type Indexing:	Teratogens
	Congenital Abnormality
	Injury or Poisoning
	Hazardous or Poisonous Substance
Semantic Relationship Indexing:	Injury or Poisoning disrupts Embryonic Structure
	Hazardous or Poisonous Substance causes Congenital Abnormality
	Biologically Active Substance affects Biologic Function
Types of Publications Covered:	Journal Articles
	Technical Reports
Users:	Health Care Professionals, Biomedical Researchers
Axes of Use:	Basic Research/commonly
	Health Services Research/unlikely
	Patient Care/commonly
	Environmental Monitoring/possibly

EVALUATION OF THE METATHESAURUS

Many users of the initial versions of Meta have focused on its completeness, asking of it whether important biomedical concepts and relationships are present or asking how well it represents a particular subdomain of biomedicine. These studies typically matched terms from an existing local or official nomenclature for a particular subject domain to Meta to determine the degree of overlap.

Cimino (1992) studied the coverage of clinical laboratory terminology and found Meta to be adequate in terms of concepts represented but inconsistent and insufficient in terms of semantic types for laboratory procedures. Similar studies found the Meta terminology inadequate for the domains of nursing (Zielstorff et al., 1993), clinical radiology (Friedman, 1993), and hypertension (Campbell et al., 1993).

Several studies have focused on the utility of Meta for representing clinical information. A great mass of data are recorded about patients and clinical activity in health centers everywhere. There is great interest in merging and exchanging this information that could become important for outcomes research and cost control research.

The problem is that no standard means of electronically com-
municating this data has gained widespread acceptance and no
existing terminology is recognized as completely adequate to describe
it. Experimenters are assessing Meta for this task.

   Huff and Warner (1990) attempted to match terms used at the
LSD Hospital of the University of Utah for representing clinical
data to words and phrases in Meta. Words matched reasonably well,
but phrase matching was low. Unsuccessful matching at the phrase
level was found to be due primarily to the presence in clinical data
of modifiers and qualifiers (examples are high, low, increased,
decreased, red, painful, increasingly, left, right). Chute et al. (1990)
and Friedman (1993) have concurred in this discovery. These and
other studies insist that modifiers and qualifiers attached to concepts
represent important and distinct information in clinical arenas. They
can significantly alter the meaning of a concept but are not distinct
concepts by themselves. Clearly a means must be found for expressing
modifying attributes within UMLS. Suggestions for addressing this
need have included adding Meta concepts for such modifying concepts
as quality, severity, pattern of occurrence, duration, frequency, trend,
onset, site, occurring with, movement (all possible modifying concepts
for symptoms). Chute et al. (1990, p. 164) suggest representing such
modifiers as relationships within the semantic network. Friedman
(1993) suggests that new semantic types be added, perhaps under the
present Qualitative Concept type, to categorize the needed modifiers.
These would need to include the idea of degree of certainty, degree
of severity, degree of change, and current status for describing patient
observations. Perhaps Meta should provide a way to construct terms
out of existing Meta terms in some systematic way that would
incorporate modifiers and duration.

FUTURE DEVELOPMENTS FOR THE METATHESAURUS

   That these studies have proven influential is borne out by the
ongoing developments and future plans for the Metathesaurus
(Humphreys & Lindberg, 1992). The 1992 (3d) version of the
Metathesaurus provides more complete integration of vocabularies
and classifications already represented, as well as the addition of other
controlled vocabularies.[2] These additions provided for expanded
coverage of clinical terminology, hazardous chemicals, diagnoses and
procedures, and terms from the domains of radiology, epidemiology,
and nursing. Medical device terminology, added from ECRI's
Universal Medical Device Nomenclature System, are important for
health services research and technology assessment. Additional data
from ChemID, the chemical identification file that provides
information about the chemicals cited in the many factual and

bibliographic databases produced by NLM's Toxicology Information Program, will be added to already existing Meta chemical terms, as well as new terms from that file. These data will help direct users to most appropriate databases for questions about the care, handling, and effects of toxic substances.

To increase Meta's ability to facilitate information retrieval in other languages, translations of Meta terms will be added, beginning with selected translations of MeSH developed by international MEDLARS centers. The 1992 edition includes the French translation of MeSH main headings (Thesaurus Biomédical Français / Anglais, 1992), prepared by the Institut National de la Santé et Recherche Médicale (INSERM). This of course introduces into Meta the complexities of multilingual vocabularies and their automatic manipulation, particularly regarding character sets. Walker et al. (1992b) have called for the internationalization of health care terminology by suggesting that UMLS incorporate terminologies that are standard outside the United States, making the Meta a multilingual, multiterminology resource. Many of these international vocabularies already contain mappings to *International Classification of Disease, SNOMED,* and other vocabularies—e.g., the *Read Clinical Codes,* a British terminology, and the German BIAK (Befunddokumentation und Arztbriefschreibung im Krankanhaus). It is argued that such an international terminology resource would facilitate automated language translation and exchange of clinical data and perhaps encourage consistency in health care delivery and development of machine products.

PROJECTS USING THE UMLS

As stated earlier, each of the UMLS components will undergo continual evaluation, development, and enhancement. Their initial versions, however, are already serving as knowledge sources in an array of projects designed to link users and electronic information or to index medical knowledge. As part of its Natural Language Systems Program, the National Library of Medicine is using UMLS in SPECIALIST, a system for parsing and accessing biomedical text (National Institutes of Health, 1992). SPECIALIST uses linguistic and biomedical knowledge for parsing queries and free text in titles and abstracts. NLM will compare retrieval based on this technique compared with standard retrieval using index terms only. To understand the biomedical language it parses, the system requires knowledge of important biomedical concepts, relationships among them, and rules to process the concepts and relations (McCray, 1992, p. 194). The UMLS Metathesaurus is the source of concepts and has been found to improve the quality of the parser and to provide needed

additional search terminology lacking in queries. Knowledge of the UMLS semantic types has aided the parser in identifying what may sensibly co-occur with a concept and helps reduce the number of questionable or invalid parses that are present when only grammatical information is available to SPECIALIST. The system includes a menu-based browser for Meta that allows viewing of term information and display of global searches, such as for all concepts with particular characteristics or all acronyms.

Another NLM expert system under development that will use UMLS components is COACH, a system for assisting GRATEFUL MED users to improve their retrieval from MEDLINE (Kingsland et al., 1992). Though the system will eventually address the problems of too much retrieval and of inappropriate retrieval, the intent of the first version of COACH is to focus on the problem of null retrieval—the search that yields no hits. Extensive analysis of GRATEFUL MED searches has, not surprisingly, confirmed widespread librarian observations of end-user search behavior. GRATEFUL MED users often get no retrieval because of "ANDing into nullspace," using terminology seldom employed in indexing, using terms too specific or not specific enough, using MeSH specialty headings inappropriately, and by using stop words. To address some of these problems, COACH provides a PC-based browser for the Metathesaurus. Using it, one can search Meta using Boolean capabilities to find MeSH headings, hierarchical contexts, child and sibling terms, semantic types, and definitions, all designed to help a user choose additional or replacement terms. Meta, being a rich source of related terms and lexical variants, can be used to augment or replace a user's terms or to map to new terms in accordance with goals of either more or better focused retrieval. COACH is now in alpha test phase with beta release expected soon. To realize its additional goals, the developers intend for COACH, in the future, to incorporate use of semantic type information and co-occurrence data.

High on the list of interests in biomedicine is vocabulary control of patient records, to facilitate their exchange, to more easily retrieve information from them, and to link them to the medical literature. Research has shown that patient encounters often generate questions that could be addressed to machine-readable sources. UMLS could be used for interpreting terms present in patient records, converting these to the vocabularies of information sources, and selecting and connecting automatically to them. It might help to summarize or collocate data from patient records. It might serve as the mapping mechanism between user queries or vocabulary from other sources to patient record databases. Lindberg and Humphreys (1992) propose steps to achieve better structured and maintained automated patient

records that would facilitate their use with UMLS. These proposals include adopting a standard format for recording patient data, using only full terms rather than abbreviations or shorthand expressions, and imposing some vocabulary control over the most standard elements of a record, with minimal use of locally developed vocabularies or extensions of existing ones. Though driven primarily by needs for cost control and outcomes research, the push for standardization of patient data may be helped by the promise of unified mechanisms for information use and retrieval as represented by the UMLS.

Several projects have been inspired by the UMLS paradigm of linking user queries directly to automated searches of databases. The program, Psychtopix, described by Powsner and Miller (1989), uses the machine-readable text of a psychiatry consult as the basis for an automated search of MEDLINE. Words in the consult are matched to a set of predetermined clinical "topics" which then invoke "canned" MEDLINE searches. This method, going from terms to topics rather than from terms to searches, is also used by Interactive Query Workstation (Cimino & Barnett, 1990) and Medline Button (Cimino et al., 1993). These programs depend on Meta for appropriate query interpretation and formulation. In the latter program, *International Classification of Disease, 9th ed., Clinical Modification* terms used to record patient information are mapped to MeSH terms through Meta when MEDLINE searches related to patient care are desired.

In a similar way to the COACH browser, Nelson et al. (1990) used MetaCard to permit a searcher to identify concepts in Meta, post them to a clipboard, and then incorporate them in a search of MEDLINE.

Powsner and Miller (1992) also use Meta to look up words selected by a user from the text of clinical records. After automatically matching the user-selected terms, the user is presented with a set of MeSH terms relevant to his or her input. The user can then select terms from the set and choose Boolean connectors to combine them to form a MEDLINE search.

The structure of Meta inspired Fu et al. (1990) to create a similarly structured patient database where entries describe not medical concepts but medical events. This database of events can then be used to index and accumulate patient information from a variety of sources and may serve as a means of mapping between different clinical databases. Meta is used as the source of terms to fill the attribute slots in medical event entries.

CONCLUSION

The UMLS Metathesaurus and Semantic Network constitute an empirically based taxonomy of biomedicine capable of linking and

mapping to diverse information sources. They provide dimensions of semantic locality that have potential for new ways of information retrieval, and perhaps for new ways of knowledge presentation. Term use data and structured nonhierarchical term relationships, particularly, show promise for managing the ever-growing problem of retrieval relevance. The UMLS components, still under development, are only in their initial stages of implementation, but some useful applications have already been devised for them, as described earlier. It is certainly only in these applications that their real efficacy can be assessed. If that assessment proves to be positive, it will be interesting to explore whether the UMLS approach to terminology management can be usefully applied in other subject domains.

NOTES

[1] The browsers are representative of software tools being developed for classifications and terminologies. Their creators promote their use as essential displays of the features, scope, and usefulness of the complex arrays of associated information present in thesauri that are less easily grasped in printed form (Walker et al., 1992a).
[2] The new 1992 source vocabularies include: *Index for Radiological Diagnoses: Including Diagnostic Ultrasound*, rev., 3d ed., 1986; *AI/RHEUM*, 1992; *COSTART* (Coding Symbols for Thesaurus of Adverse Reaction Terms); *CRISP Thesaurus*, 1992; *Classification of Nursing Diagnoses: Proceedings of the Ninth Conference*, 1991; *Nursing Interventions Classification* (NIC), 1992; and epidemiology terms submitted by McMaster University.

REFERENCES

Adams, S. (1981). *Medical bibliography in an age of discontinuity*. Chicago, IL: Medical Library Association, Inc.
AI/RHEUM. (1992). Bethesda, MD: National Library of Medicine.
Appel, R. D.; Komorowski, H. J.; Barr, C. E.; & Greenes, R. A. (1988). Intelligent focusing in knowledge indexing and retrieval—The relatedness tool. In R. A. Greenes (Ed.), *Proceedings: The twelfth annual symposium on computer applications in medical care* (November 6-9, 1988, Washington, DC) (pp. 152-157). Los Alamitos, CA: IEEE Computer Society Press.
Bicknell, E. J.; Sneiderman, C. A.; & Rada, R. F. (1988). Computer-assisted merging and mapping of medical knowledge bases. In R. A. Greenes (Ed.), *Proceedings: The twelfth annual symposium on computer applications in medical care* (November 6-9, 1988, Washington, DC) (pp. 158-164). Los Alamitos, CA: IEEE Computer Society Press.
Bishop, C. W., & Ewing, P. (1992). Representing medical knowledge: Reconciling the present or creating the future? *M.D. Computing, 9*(4), 218-225.
Campbell, J. R.; Kallenberg, G. A.; & Sherrick, R. C. (1993). The clinical utility of META: An analysis for hypertension. In M. E. Frisse (Ed.), *Supporting collaboration: Proceedings: Sixteenth annual symposium on computer applications in medical care* (A conference of the American Medical Informatics Association, November 8-11, 1992, Baltimore Convention Center, Baltimore, MD) (pp. 397-401). New York: McGraw-Hill, Inc.
Carroll-Johnson, R. M. (Ed.); North American Nursing Diagnosis Association (author). (1991). *Classification of nursing diagnoses: Proceedings of the ninth conference* (Orlando, FL, 1990). Philadelphia, PA: J. B. Lippincott.
Chute, C. G.; Yang, Y.; Tuttle, M. S.; Sherertz, D. D.; Olson, N. E.; & Erlbaum, M. S. (1990). A preliminary evaluation of the UMLS Metathesaurus for patient record

classification. In R. A. Miller (Ed.), *SCAMC proceedings: Fourteenth annual symposium on computer applications in medical care: Standards in medical informatics* (A conference of the American Medical Informatics Association, November 4-7, 1990, Sheraton Washington Hotel, Washington, DC) (pp. 161-165). Los Alamitos, CA: IEEE Computer Society Press.

Cimino, C., & Barnett, G. O. (1990). Standardizing access to computer-based medical resources. In R. A. Miller (Ed.), *SCAMC proceedings: Fourteenth annual symposium on computer applications in medical care: Standards in medical informatics* (A conference of the American Medical Informatics Association, November 4-7, 1990, Sheraton Washington Hotel, Washington, DC) (pp. 33-37). Los Alamitos, CA.: IEEE Computer Society Press.

Cimino, J. J. (1992). Representation of clinical laboratory terminology in the Unified Medical Language System. In P. D. Clayton (Ed.), *Assessing the value of medical informatics: Proceedings: Fifteenth annual symposium on computer applications in medical care* (November 17-20, 1991, Sheraton Washington Hotel, Washington, DC) (pp. 199-203). New York: McGraw-Hill, Inc., Health Professions Division.

Cimino, J. J.; Johnson, S. B.; Aguirre, A.; Roderer, N.; & Clayton, P. D. (1993). The Medline button. In M. E. Frisse (Ed.), *Supporting collaboration: Proceedings: Sixteenth annual symposium on computer applications in medical care* (A conference of the American Medical Informatics Association, November 8-11, 1992, Baltimore Convention Center, Baltimore, MD) (pp. 81-85). New York: McGraw-Hill, Inc.

Cimino, J. J.; Johnson, S. B.; Hripcsak, G.; Sideli, R. V.; Fink, D. J.; Friedman, C.; & Clayton, P. D. (1992). One year's experience with the Unified Medical Language System (UMLS) in academia and patient care. In K. C. Lun, P. Degoulet, T. E. Piemme, & O. Rienhoff (Eds.), *MEDINFO 92* (Proceedings of the Seventh World Congress on Medical Informatics, Geneva Palexpo, Switzerland, 6-10 September 1992) (Part 2, pp. 1501-1505). Amsterdam, The Netherlands: North-Holland, Elsevier Science Publishers B. V.

COSTART: Coding symbols for thesaurus of adverse reaction terms. (1989). 3d ed. Springfield, VA: U.S. Dept. of Commerce, NTIS.

CRISP Thesaurus. (1992). Bethesda, MD: National Institutes of Health.

Diagnostic and statistical manual of mental disorders: DSM-III-R. (1987). 3d ed., rev. Washington, DC: American Psychiatric Association.

Friedman, C. (1993). The UMLS coverage of clinical radiology. In M. E. Frisse (Ed.), *Supporting collaboration: Proceedings: Sixteenth annual symposium on computer applications in medical care* (A conference of the American Medical Informatics Association, November 8-11, 1992, Baltimore Convention Center, Baltimore, MD) (pp. 309-313). New York: McGraw-Hill, Inc.

Fu, L. S.; Bouhaddou, O.; Huff, S. M.; Sorenson, D. K.; & Warner, H. R. (1990). Toward a public domain UMLS patient database. In R. A. Miller (Ed.), *SCAMC proceedings: Fourteenth annual symposium on computer applications in medical care: Standards in medical informatics* (A conference of the American Medical Informatics Association, November 4-7, 1990, Sheraton Washington Hotel, Washington, DC) (pp. 170-174). Los Alamitos, CA: IEEE Computer Society Press.

Huff, S. M., & Warner, H. R. (1990). A comparison of Meta-1 and HELP terms: Implications for clinical data. In R. A. Miller (Ed.), *SCAMC proceedings: Fourteenth annual symposium on computer applications in medical care: Standards in medical informatics* (A conference of the American Medical Informatics Association, November 4-7, 1990, Sheraton Washington Hotel, Washington, DC) (pp. 166-169). Los Alamitos, CA: IEEE Computer Society Press.

Humphreys, B. L., & Lindberg, D. A. B. (1989). Building the Unified Medical Language System. In L. C. Kingsland III (Ed.), *Proceedings: The thirteenth annual symposium on computer applications in medical care* (November 5-8, 1989, Washington, DC) (pp. 475-480). Los Alamitos, CA: IEEE Computer Society Press.

Humphreys, B. L., & Lindberg, D. A. B. (1992). The Unified Medical Language System project: A distributed experiment in improving access to biomedical information. In K. C. Lun, P. Degoulet, T. E. Piemme, & O. Rienhoff (Eds.), *MEDINFO 92*

(Proceedings of the Seventh World Congress on Medical Informatics, Geneva Palexpo, Switzerland, 6-10 September 1992) (Part 2, pp. 1496-1500). Amsterdam, The Netherlands: North-Holland, Elsevier Science Publishers B. V.

Humphreys, B. L., & Schuyler, P. L. (in press). The Unified Medical Language System: Moving beyond the vocabulary of bibliographic retrieval. In N. C. Broering (Ed.), *High performance medical libraries: Advances in information management.* Westport, CT: Meckler Publishing.

Index for radiological diagnoses: Including diagnostic ultrasound. (1986). Rev. 3d ed. Reston, VA: American College of Radiology.

Kingsland, L. C., III; Syed, E. J.; & Lindberg, D. A. B. (1992). Coach: An expert searcher program to assist Grateful Med users searching MEDLINE. In K. C. Lun, P. Degoulet, T. E. Piemme, & O. Rienhoff (Eds.), *MEDINFO 92* (Proceedings of the Seventh World Congress on Medical Informatics, Geneva Palexpo, Switzerland, 6-10 September 1992) (Part 1, pp. 382-386). Amsterdam, The Netherlands: North-Holland, Elsevier Science Publishers B. V.

Lindberg, D. A. B., & Humphreys, B. L. (1989). Computer systems that understand medical meaning. In J. R. Scherrer, R. A. Coté, & S. H. Mandil (Eds.), *Computerized natural medical language processing for knowledge representation: Proceedings of the IFIP-IMIA WG6 international working conference on computerized natural medical language processing for knowledge representation* (Geneva, Switzerland, 12-15 September, 1988) (pp. 5-17). Amsterdam, The Netherlands: North-Holland, Elsevier Science Publishers B. V.

Lindberg, D. A. B., & Humphreys, B. L. (1990). The UMLS knowledge sources: Tools for building better user interfaces. In R. A. Miller (Ed.), *SCAMC proceedings: Fourteenth annual symposium on computer applications in medical care: Standards in medical informatics* (A conference of the American Medical Informatics Association, November 4-7, 1990, Sheraton Washington Hotel, Washington, DC) (pp. 121-125). Los Alamitos, CA: IEEE Computer Society Press.

Lindberg, D. A. B., & Humphreys, B. L. (1992). The Unified Medical Language System (UMLS) and computer-based patient records. In M. J. Ball & M. F. Collen (Eds.), *Aspects of the computer-based patient record* (pp. 165-175). New York: Springer-Verlag.

Masys, D. R. (1993). An evaluation of the source selection elements of the prototype UMLS information sources map. In M. E. Frisse (Ed.), *Supporting collaboration: Proceedings: Sixteenth annual symposium on computer applications in medical care* (A conference of the American Medical Informatics Association, November 8-11, 1992, Baltimore Convention Center, Baltimore, MD) (pp. 295-298). New York: McGraw-Hill, Inc.

Masys, D. R., & Humphreys, B. L. (1992). Structure and function of the UMLS information sources map. In K. C. Lun, P. Degoulet, T. E. Piemme, & O. Rienhoff (Eds.), *MEDINFO 92* (Proceedings of the Seventh World Congress on Medical Informatics, Geneva Palexpo, Switzerland, 6-10 September 1992) (Part 2, pp. 1518-1521). Amsterdam, The Netherlands: North-Holland, Elsevier Science Publishers B. V.

McCloskey, J. C., & Bulechek, G. M. (Eds.). (1992). *Nursing interventions classification* (NIC). St. Louis, MO: Mosby Year Book.

McCray, A. T. (1989). The UMLS semantic network. In L. C. Kingsland III (Ed.), *Proceedings: The thirteenth annual symposium on computer applications in medical care* (November 5-8, 1989, Washington, DC) (pp. 503-507). Los Alamitos, CA: IEEE Computer Society Press.

McCray, A. T. (1992). Extending a natural language parser with UMLS knowledge. In P. D. Clayton (Ed.), *Assessing the value of medical informatics: Proceedings: Fifteenth annual symposium on computer applications in medical care* (November 17-20, 1991, Sheraton Washington Hotel, Washington, DC) (pp. 194-198). New York: McGraw-Hill, Inc., Health Professions Division.

McCray, A. T., & Hole, W. T. (1990). The scope and structure of the first version of the UMLS semantic network. In R. A. Miller (Ed.), *SCAMC proceedings: Fourteenth annual symposium on computer applications in medical care:*

Standards in medical informatics (A conference of the American Medical Informatics Association, November 4-7, 1990, Sheraton Washington Hotel, Washington, DC) (pp. 126-130). Los Alamitos, CA: IEEE Computer Society Press.

Merz, R. B.; Cimino, C.; Barnett, G. O.; Blewett, D. R.; Gnassi, J. A.; Grundmeier, R.; & Hassan, L. (1993). Q & A: A query formulation assistant. In M. E. Frisse (Ed.), *Supporting collaboration: Proceedings: Sixteenth annual symposium on computer applications in medical care* (A conference of the American Medical Informatics Association, November 8-11, 1992, Baltimore Convention Center, Baltimore, MD) (pp. 498-502). New York: McGraw-Hill, Inc.

Miller, P. L.; Barwick, K. W.; Morrow, J. S.; Powsner, S. M.; & Riely, C. A. (1988a). Semantic relationships and medical bibliographic retrieval: A preliminary assessment. *Computers and Biomedical Research, 21*(1), 64-77.

Miller, P. L.; Smith, P.; Morrow, J. S.; Riely, C. A.; & Powsner, S. M. (1988b). Semantic relationships and MeSH. In R. A. Greenes (Ed.), *Proceedings: The twelfth annual symposium on computer applications in medical care* (November 6-9, 1988, Washington, DC) (pp. 174-179). Los Alamitos, CA: IEEE Computer Society Press.

Miller, P. L.; Wright, L. W.; Frawley, S. J.; Clyman, J. I.; & Powsner, S. M. (1992). Selecting relevant information resources in a network-based environment: The UMLS information sources map. In K. C. Lun, P. Degoulet, T. E. Piemme, & O. Rienhoff. (Eds.), *MEDINFO 92* (Proceedings of the Seventh World Congress on Medical Informatics, Geneva Palexpo, Switzerland, 6-10 September 1992) (Part 2, pp. 1512-1517). Amsterdam, The Netherlands: North-Holland, Elsevier Science Publishers B. V.

Miller, R. A.; Gieszczykiewicz, F. M.; Vries, J. K.; & Cooper, G. F. (1993). CHARTLINE: Providing bibliographic references relevant to patient charts using the UMLS Metathesaurus knowledge sources. In M. E. Frisse (Ed.), *Supporting collaboration: Proceedings: Sixteenth annual symposium on computer applications in medical care* (A conference of the American Medical Informatics Association, November 8-11, 1992, Baltimore Convention Center, Baltimore, MD) (pp. 86-90). New York: McGraw-Hill, Inc.

National Institutes of Health. (1992). *National Library of Medicine: Programs & services fiscal year 1991*. Washington, DC: U.S. Department of Health and Human Services, Public Health Service.

National Library of Medicine. (1986-1987). *Long Range Plan* (In seven parts). Washington, DC: U.S. Department. of Health and Human Services.

Nelson, S. J.; Fuller, L. F.; Erlbaum, M. S.; Tuttle, M. S.; Sherertz, D. D.; & Olson, N. E. (1993). The semantic structure of the UMLS Metathesaurus. In M. E. Frisse (Ed.), *Supporting collaboration: Proceedings: Sixteenth annual symposium on computer applications in medical care* (A conference of the American Medical Informatics Association, November 8-11, 1992, Baltimore Convention Center, Baltimore, MD) (pp. 649-653). New York: McGraw-Hill, Inc.

Nelson, S. J.; Sherertz, D. D.; Tuttle, M. S.; & Erlbaum, M. S. (1990). Using MetaCard: A hypercard browser for biomedical knowledge sources. In R. A. Miller (Ed.), *SCAMC proceedings: Fourteenth annual symposium on computer applications in medical care: Standards in medical informatics* (A conference of the American Medical Informatics Association, November 4-7, 1990, Sheraton Washington Hotel, Washington, DC) (pp. 151-154). Los Alamitos, CA: IEEE Computer Society Press.

Nelson, S. J.; Tuttle, M. S.; Cole, W. G.; Sherertz, D. D.; Sperzel, W. D.; Erlbaum, M. S.; Fuller, L. L.; & Olson, N. E. (1992). From meaning to term: Semantic locality in the UMLS Metathesaurus. In P. D. Clayton (Ed.), *Assessing the value of medical informatics: Proceedings: Fifteenth annual symposium on computer applications in medical care* (November 17-20, 1991, Sheraton Washington Hotel, Washington, DC) (pp. 209-213). New York: McGraw-Hill, Inc., Health Professions Division.

Physicians' current procedural terminology. (1989). 4th ed. Chicago, IL: American Medical Association.

Powsner, S. M., & Miller, P. L. (1989). Linking bibliographic retrieval to clinical reports: PsychTopix. In L. C. Kingsland III (Ed.), *Proceedings: The thirteenth*

annual symposium on computer applications in medical care (November 5-8, 1989, Washington, DC) (pp. 431-435). Los Alamitos, CA: IEEE Computer Society Press.

Powsner, S. M., & Miller, P. L. (1992). From patient reports to bibliographic retrieval: A Meta-1 front-end. In P. D. Clayton (Ed.), *Assessing the value of medical informatics: Proceedings: Fifteenth annual symposium on computer applications in medical care* (November 17-20, 1991, Sheraton Washington Hotel, Washington, DC) (pp. 526-530). New York: McGraw-Hill, Inc.

Sherertz, D. D.; Olson, N. E.; Tuttle, M. S.; & Erlbaum, M. S. (1990). Source inversion and matching in the UMLS Metathesaurus. In R. A. Miller (Ed.), *SCAMC proceedings: Fourteenth annual symposium on computer applications in medical care: Standards in medical informatics* (A conference of the American Medical Informatics Association, November 4-7, 1990, Sheraton Washington Hotel, Washington, DC) (pp. 141-145). Los Alamitos, CA: IEEE Computer Society Press.

Sherertz, D. D.; Tuttle, M. S.; Blois, M. S.; & Erlbaum, M. S. (1988). Intervocabulary mapping within the UMLS: The role of lexical matching. In R. A. Greenes (Ed.), *Proceedings: The twelfth annual symposium on computer applications in medical care* (November 6-9, 1988, Washington, DC) (pp. 201-206). Los Alamitos, CA: IEEE Computer Society Press.

Sherertz, D. D.; Tuttle, M. S.; Cole, W. G.; Erlbaum, M. S.; Olson, N. E.; & Nelson, S. J. (1989a). A HyperCard ® implementation of Meta-1: The first version of the UMLS Metathesaurus. In L. C. Kingsland III (Ed.), *Proceedings: The thirteenth annual symposium on computer applications in medical care* (November 5-8, 1989, Washington, DC) (pp. 1017-1018). Los Alamitos, CA: IEEE Computer Society Press.

Sherertz, D. D.; Tuttle, M. S.; Olson, N. E.; Erlbaum, M. S.; & Nelson, S. J. (1989b). Lexical mapping in the UMLS Metathesaurus. In L. C. Kingsland III (Ed.), *Proceedings: The thirteenth annual symposium on computer applications in medical care* (November 5-8, 1989, Washington, DC) (pp. 494-499). Los Alamitos, CA: IEEE Computer Society Press.

Sperzel, D. W.; Erlbaum, M. S.; Fuller, L. L.; Sherertz, D. D.; Olson, N. E.; Schuyler, P. L.; Hole, W. T.; Savage, A.; Passarelli, P.; & Tuttle, M. S. (1990). Editing the UMLS Metathesaurus: Review and enhancement of a computed knowledge source. In R. A. Miller (Ed.), *SCAMC proceedings: Fourteenth annual symposium on computer applications in medical care: Standards in medical informatics* (A conference of the American Medical Informatics Association, November 4-7, 1990, Sheraton Washington Hotel, Washington, DC) (pp. 136-140). Los Alamitos, CA: IEEE Computer Society Press.

Sperzel, W. D., & Tuttle, M. S. (1989). Updating the UMLS Metathesaurus: a model. In L. C. Kingsland III (Ed.), *Proceedings: The thirteenth annual symposium on computer applications in medical care* (November 5-8, 1989, Washington, DC) (pp. 488-493). Los Alamitos, CA: IEEE Computer Society Press.

Systematized nomenclature of medicine. (1986). 2d ed. Skokie, IL: College of American Pathologists.

Thesaurus biomédical français/anglais. (1992). Paris, France: Institut National de la Santé et Recherche Médicale (INSERM).

Tilley, C. B. (1990). Medical databases and health information systems. In M. E. Williams (Ed.), *Annual Review of Information Science and Technology* (Vol. 25, pp. 313-382). Amsterdam, The Netherlands: Elsevier Science Publishers.

Tuttle, M. S.; Blois, M. S.; Erlbaum, M. S.; Nelson, S. J.; & Sherertz, D. D. (1988). Toward a bio-medical thesaurus: Building the foundation of the UMLS. In R. A. Greenes (Ed.), *Proceedings: The twelfth annual symposium on computer applications in medical care* (November 6-9, 1988, Washington, DC) (pp. 191-195). Los Alamitos, CA: IEEE Computer Society Press.

Tuttle, M. S.; Sherertz, D. D.; Erlbaum, M. S.; Olson, N. E.; & Nelson, S. J. (1989). Implementing Meta-1: The first version of the UMLS Metathesaurus. In L. C. Kingsland III (Ed.), *Proceedings: The thirteenth annual symposium on computer*

applications in medical care (November 5-8, 1989, Washington, DC) (pp. 483-487). Los Alamitos, CA: IEEE Computer Society Press.

Tuttle, M. S.; Sherertz, D. D.; Erlbaum, M. S.; Sperzel, W. D.; Fuller, L. F.; Olson, N. E.; Nelson, S. J.; Cimino, J. J.; & Chute, C. G. (1992). Adding your terms and relationships to the UMLS Metathesaurus™. In P. D. Clayton (Ed.), *Assessing the value of medical informatics: Proceedings: Fifteenth annual symposium on computer applications in medical care* (November 17-20, 1991, Sheraton Washington Hotel, Washington, DC) (pp. 219-223). New York: McGraw-Hill, Inc., Health Professions Division.

Tuttle, M. S.; Sperzel, W. D.; Olson, N. E.; Erlbaum, M. S.; Suarez-Munist, O.; Sherertz, D. D.; Nelson, S. J.; & Fuller, L. F. (1993). The homogenization of the Metathesaurus ® schema and distribution format. In M. E. Frisse (Ed.), *Supporting collaboration: Proceedings: Sixteenth annual symposium on computer applications in medical care* (A conference of the American Medical Informatics Association, November 8-11, 1992, Baltimore Convention Center, Baltimore, MD) (pp. 299-303). New York: McGraw-Hill, Inc.

United States. Health Care Financing Administration. (1991). *International classification of disease(s), ninth revision, clinical modification.* Washington, DC: USGPO.

Walker, D. C.; Cimino, J. J.; Gabrieli, E. R.; Giere, W.; Lamberts, H.; Linnarsson, R.; Moore, W. G.; Read, J. D.; Rothwell, D. J.; Satomura, Y.; Shepherd, S. G. S.; Sherertz, D. D.; Sun Rong; Walker, J. C.; & Wohlfahrt, M. (1992a). Electronic presentations of health care terminologies: Some terminology browsers. In K. C. Lun, P. Degoulet, T. E. Piemme, & O. Rienhoff (Eds.), *MEDINFO 92* (Proceedings of the Seventh World Congress on Medical Informatics, Geneva Palexpo, Switzerland, 6-10 September 1992) (Part 2, p. 1615). Amsterdam, The Netherlands: North-Holland, Elsevier Science Publishers B. V.

Walker, D. C.; Walters, R. F.; Cimino, J. J.; Dujols, P.; Li Ensheng; Giere, W.; Kiuchi, T.; Lamberts, H.; Moore, W. G.; Roger, F. H.; Satomura, Y.; & Stitt, F. W. (1992b). Internationalization of health care terminology. In K. C. Lun, P. Degoulet, T. E. Piemme, & O. Rienhoff (Eds.), *MEDINFO 92* (Proceedings of the Seventh World Congress on Medical Informatics, Geneva Palexpo, Switzerland, 6-10 September 1992) (Part 2, pp. 1444-1451). Amsterdam, The Netherlands: North-Holland, Elsevier Science Publishers B. V.

Zielstorff, R. D.; Cimino, C.; Barnett, G. O.; Hassan, L.; & Blewett, D. R. (1993). Representation of nursing terminology in the UMLS Metathesaurus: A pilot study. In M. E. Frisse (Ed.), *Supporting collaboration: Proceedings: Sixteenth annual symposium on computer applications in medical care* (A conference of the American Medical Informatics Association, November 8-11, 1992, Baltimore Convention Center, Baltimore, MD) (pp. 392-396). New York: McGraw-Hill, Inc.