The Unified Medical Language System

D. A. B. Lindberg,
B. L. Humphreys,
A. T. McCray

National Library of Medicine,
Bethesda MD,
USA

Abstract: In 1986, the National Library of Medicine began a long-term research and development project to build the Unified Medical Language System® (UMLS®). The purpose of the UMLS is to improve the ability of computer programs to "understand" the biomedical meaning in user inquiries and to use this understanding to retrieve and integrate relevant machine-readable information for users. Underlying the UMLS effort is the assumption that timely access to accurate and up-to-date information will improve decision making and ultimately the quality of patient care and research. The development of the UMLS is a distributed national experiment with a strong element of international collaboration. The general strategy is to develop UMLS components through a series of successive approximations of the capabilities ultimately desired. Three experimental Knowledge Sources, the Metathesaurus®, the Semantic Network, and the Information Sources Map have been developed and are distributed annually to interested researchers, many of whom have tested and evaluated them in a range of applications. The UMLS project and current developments in high-speed, high-capacity international networks are converging in ways that have great potential for enhancing access to biomedical information.

Key-Words: Unified Medical Language System, UMLS, Semantic Network, National Library of Medicine, NLM, Telecommunication, Information Management, IAIMS

Introduction

In 1986, the National Library of Medicine (NLM) began a long-term research and development project to build the Unified Medical Language System (UMLS). The purpose of the UMLS is to improve the ability of computer programs to "understand" the biomedical meaning in user inquiries and to use this understanding to retrieve and integrate relevant machine-readable information for users [1]. More specifically, the UMLS project is an effort to overcome two significant barriers to effective retrieval of machine-readable biomedical information. The first is the variety of ways the same concepts are expressed in different machine-readable sources and by different people. The second is the distribution of useful information among many disparate databases and systems. Advances in technology, such as more powerful workstations, storage devices, and telecommunications capabilities, and improvements in organizational infrastructure, as typified by the National Network of Libraries of Medicine™ (NN/LM™) [2] and Integrated Advanced Information Management Systems (IAIMS) [3, 4], are necessary but not sufficient to connect health care practitioners and researchers to pertinent machine-readable information. The critical and inherently most difficult requirement is the conceptual connection between the user's question and the available machine-readable information [5].

Underlying the UMLS effort is the assumption that timely access to accurate and up-to-date information will improve decision-making and ultimately the quality of patient care and research. A growing body of evidence supports this assumption [6, 7]. The UMLS project assumes that the amount of useful biomedical information will continue both to increase and to be dispersed among many databases and systems. The rapidly increasing number of biomedical sources accessible via the Internet illustrates this phenomenon. The UMLS strategy recognizes that many of the differences in the terminology used in databases and by users reflect important distinctions in purpose and perspective [8]. Although current efforts to standardize the record structure, transmission formats, and terminology of specific types of biomedical information [9, 10] may reduce the complexity of the UMLS task, they will not eliminate it.

In short, the UMLS project assumes a continuing need to navigate a large array of diverse machine-readable sources to obtain information relevant to a particular user's practice or research question. Disparities in the way concepts are expressed will continue to exist in different sources, for example, in patient record systems

and in bibliographic databases. The UMLS approach is to help applications to exploit this diversity for the benefit of health professionals and biomedical researchers. While the results of the UMLS effort may be useful in related endeavors, the project is not an attempt to develop a single standard biomedical vocabulary or classification, or to build a knowledge base covering all of biomedicine, nor is it an attempt to define the structure and content of computer-based patient records.

NLM's mission is to support biomedical research and to improve health care delivery by providing ready access to published biomedical information. Index Medicus™, MeSH™, MEDLARS™, the National Network of Libraries of Medicine, TOXNET[1]™, and DOCLINE™ [11] all contribute to this mission. NLM's support for training in medical librarianship and medical informatics, for medical informatics research, and for innovative systems development also contributes to an effective infrastructure that supports enhanced information services. NLM focuses its resources on the creation and maintenance of services that are not readily subject to private development and distribution. The long-term maintenance of the knowledge structures required by the UMLS presents this kind of challenge and is therefore an appropriate role for NLM.

Development Strategy and History

The development of the UMLS is a distributed national experiment with a strong element of international collaboration [12]. To address the complex problem of relating user inquiries to the content of biomedical information sources, NLM has assembled a multi-disciplinary in-house research group[2] and contracted with a number of primarily university-based medical informatics research groups throughout the United States[3]. The UMLS research team incorporates: experience in the development of different types of information sources, e.g., patient record systems, expert systems, bibliographic databases; expertise in a range of disciplines, e.g., medical informatics, computer science, linguistics, library and information science; and access to members of the user groups the UMLS intends to serve. From its inception the UMLS project has also sought input from a wide range of intended users of UMLS products, including many outside the United States [13]. The general strategy is to develop the UMLS components through a series of successive approximations of the capabilities ultimately desired. Rapid development and broad distribution of early UMLS products will allow subsequent expansions in scope and complexity to be based on feedback from real applications in a variety of biomedical environments. The success of the UMLS effort is dependent on collaborators who are willing to apply its experimental products.

In the first phase of the project (1986–1988), the UMLS research team investigated user needs, developed tools for the research effort, identified required UMLS capabilities, examined alternative methods for delivering these capabilities, and defined in general terms the necessary components of the system [14]. Two types of components were deemed essential: new machine-readable knowledge sources and sophisticated interface programs. The interface programs would use the highly structured information about biomedical terminology and databases contained in the knowledge sources to interpret user inquiries, to identify and locate relevant sources of information, and to execute successful searches on the user's behalf.

Definition of the Three UMLS Knowledge Sources

From the outset, it was assumed that a "Metathesaurus" which linked terminology and concepts from a range of vocabularies and classifications would be needed. The UMLS team had no preconceptions about the specific form of this knowledge source, however, or about the methods for building it [14]. Serious consideration was given to creating a new canonical classification of biomedical concepts to which existing vocabularies could be mapped. The UMLS-funded project to create "generic frames" for patient findings [15, 16] explored this approach. A semantic network was proposed as a structure for such a new taxonomy [17]. As research proceeded, it became clear that development of a new detailed taxonomy of the requisite scope (all of biomedicine) was a monumental undertaking with no guarantee of an end product more useful than existing vocabulary systems. Such an undertaking was not in keeping with the strategy of rapid distribution and feedback on early versions of UMLS products. Other UMLS-funded work indicated that a new canonical taxonomy was not essential to achieving the UMLS goal of aiding the retrieval of information from disparate machine-readable sources. Direct linking of alternative names for concepts taken from exist-

1 As part of the MEDLARS system, TOXNET is a computerized collection of files on toxicology, hazardous chemicals and related areas.

2 In addition to the authors, Wiliam Hole, M. D., Lawrence Kingsland III, Ph. D., Daniel Masys, M. D., R. P. C. Rodgers, M. D., Harold Schoolman, M. D., and Peri Schuyler lead UMLS research activities at NLM.

3 The current UMLS contractors are Brigham and Women's Hospital (PI: Robert Greenes, M. D., Ph. D.), Columbia University (PI: James Cimino, M. D.), Lexical Technology, Inc. (PI: Mark Tuttle), Massachusetts General Hospital (PI: G. Octo Barnett, M. D.), University of Pittsburgh (PI: Randolph Miller, M. D.) with subcontractor University of Utah (PI: Homer Warner, M. D., Ph. D.), and Yale School of Medicine (PI: Perry Miller, M. D., Ph. D.). The MPC Corporation (PIs: Randolph Miller, M. D., University of Pittsburgh; David Evans, Ph. D., Carnegie-Mellon University; subcontractor PI: Homer Warner, M. D., Ph. D., University of Utah) and the University of California, San Francisco (PI: Marsden S. Blois, M. D., Ph. D.) were UMLS contractors from 1986–1988.

ing machine-readable vocabularies emerged as a potentially viable way to build the Metathesaurus [18]. This approach exploited both automated lexical matching techniques [19] and the structured "knowledge" embedded in existing biomedical vocabularies, classifications, and databases such as MEDLINE® [20]. The methodology selected to build the Metathesaurus is an example of the reuse of knowledge developed for other purposes discussed by Musen [21].

Once the outline of the Metathesaurus was defined, the UMLS project team decided that a separate, associated UMLS Semantic Network was needed, not of the individual concepts in the Metathesaurus, but of the semantic types or broad categories of concepts within it. The assignment of semantic types to concepts in the Metathesaurus provides a consistent high level of categorization of these concepts and also links them to the biomedical "common sense" represented by the relationships among semantic types in the Network [22–24].

The initial definition of the characteristics of the third UMLS Knowledge Source, the Information Sources Map, occasioned less debate. The UMLS research team readily agreed that it should contain both human readable and machine-processable information about the scope and content of publicly available machine-readable biomedical information sources. This information is needed to support automated or semi-automated scource selection. The other key component of Information Sources Map records will be procedural information needed to effect successful automated searches of the selected sources [25]. Details about the structure and content of the three UMLS Knowledge Sources appear in subsequent sections of the paper.

Building, Distributing, and Applying the UMLS Knowledge Sources

The highest priority for the next UMLS development phase (1989 to 1991) was the production of initial versions of the Knowledge Sources. NLM issued the first experimental editions of the Metathesaurus and the Semantic Network on CD-ROM in the fall of 1990 [26]. A year later, the first version of the Information Sources Map was released, along with updated versions of the Metathesaurus and the Semantic Network. New experimental editions of all three UMLS Knowledge Sources now appear on CD-ROM annually. NLM is considering making a UMLS knowledge source server based on the client-server paradigm available on the Internet. The knowledge server would allow users, developers, and programs to navigate and retrieve information stored in the data files.

In addition to encouraging testing and feedback on useful improvements, broad dissemination of the early versions of the Knowledge Sources promotes the development of prototypes of the interface programs required to deliver the UMLS functionality to end users. The combination of centralized development of the core of the Knowledge Sources and decentralized development of the applications programs that make use of them was considered likely to foster progress toward the complex goals of the UMLS project. Experience to date supports this strategy, although some aspects of maintenance of the Knowledge Sources will become more decentralized over time and some utility programs are being developed centrally for release with the Knowledge Sources.

The UMLS project priorities for the 1992–1994 period are to develop an array of useful applications that rely on the UMLS Knowledge Sources, to expand and refine the Knowledge Sources based on feedback from early applications, and to establish robust production systems and procedures for maintaining and distributing the Knowledge Sources [12]. NLM itself and its UMLS research contractors are among those developing specific applications programs that make use of the UMLS Knowledge Sources. NLM has also awarded grants for UMLS-related projects and provided small contracts for testing the use of the Knowledge Sources with existing applications at several institutions [5]. Some significant applications have emerged which should be ready for general distribution within the next two years. In the meantime, each successive edition of the UMLS Knowledge Sources incorporates substantial enhancements in content or format or both [27]. French translations of MeSH terms were added in a recent addition, and additional translations of Metathesaurus terminologies will be incorporated in the future. Steady progress is being made toward sustainable production mechanisms for maintaining, enhancing, and distributing these complex products [28–32].

UMLS Knowledge Sources

The UMLS Knowledge Sources contain information useful for developing intelligent interfaces to biomedical information systems. The knowledge stored in the Metathesaurus and Semantic Network should help interfaces to map user queries to information in a wide range of biomedical information systems. The knowledge stored in the Information Source Map should assist in the identification of the most appropriate information source or sources for the query posed. All three Knowledge Sources have been designed to allow for local addenda that can work in concert with the regularly released NLM files [33].

Metathesaurus

Biomedical vocabularies have been developed for a variety of disciplines and for a range of information sources, including bibliographic databases, factual databases, clinical record systems, and expert systems. The Metathesaurus may be seen as a thesaurus that transcends these individual thesauri, or controlled vocabularies, by virtue of the lexical and semantic links that it provides [34, 35]. The Metathesaurus contains information about biomedical terms from a continuously increasing set of controlled vocabularies and classifications. Additional information is added in the process of constructing the Metathesaurus, but the original mean-

ing of a term in its source vocabulary is always preserved. The 1993 version contains terms from 15 biomedical vocabularies [36]. In some cases, all terms from a vocabulary are included, while in other cases, only selected terms are included. As the Metathesaurus continues to evolve, more and more vocabularies will be represented in their entirety. Although extensive, the Metathesaurus is not meant to be a complete source of biomedical concepts. Its scope is determined by the scope of the vocabularies contained within it.

The Metathesaurus is organized by concept, or meaning. In this sense it is a true thesaurus in the tradition of Peter Mark Roget, or "basically a tool for transforming ideas into words" [37]. Entries in the Metathesaurus connect alternate names for the same concept, such as synonyms, lexical variants, and translations. Strings that are lexical variants of each other are first grouped together as a single term with one string designated as the preferred form of that term. Terms that mean the same thing are then linked together as alternate names of the same concept, with one term designated as the preferred name of the concept. The designation of preferred forms and preferred names is done by an algorithm based on an order of precedence among the source vocabularies. In some cases, identically spelled strings mean very different things in different vocabularies. For example, "dressing" in MeSH is an entry term to "bandages", and in the Nursing Interventions Classification it has the meaning of "to dress". These terms would, thus, be treated as different concepts in the Metathesaurus.

The identification of relationships among different concepts offers great potential for improved information systems. The Metathesaurus incorporates all inter-concept relationships present in its source vocabularies, adds relationships between concepts from different vocabularies, and provides empirically derived co-occurrence data for some information sources [38].

A concept may be related hierarchically to another concept within the same source. It may be a parent, child or sibling of one or more concepts in the source vocabulary. These relationships are represented in the Metathesaurus, thereby enabling a user to choose the most appropriate terms when formulating a search strategy. Examination of the relevant hierarchy might, for example, make it clear that the initial search term chosen was either too broad or too narrow. For certain sections of the MeSH hierarchy some of the implicit links between child and parent concepts have been labelled with a valid relationship from the Semantic Network. All anatomical, disease, and psychiatry and psychology terms have been labelled, as well as sections of biological phenomena, including physiology.

Broader, narrower, and other close relations between concepts are labelled during Metathesaurus construction. Not all close relationships among concepts in different Methathesaurus vocabularies have been identified, however. This is an iterative process that begins with lexical programs and is refined by a variety of techniques including human review and revision. The number of links among concepts will gradually increase. The greater the number of connections that can be identified among concepts across vocabularies, the more likely it is that information will be found in a variety of relevant information sources.

When a concept present in one vocabulary does not appear as a single concept in another vocabulary, but can be represented or closely approximated by a combination of concepts in the second vocabulary, the Metathesaurus may store this combination as an "associated expression". The associated expression can be used to construct an appropriate search statement in the database that is coded with this other vocabulary. For example, the DSM-III-R concept "Amphetamine or similarly acting sympathomimetic delirium" is not found as a single concept in MeSH. In order to find this concept in MEDLINE, the MeSH heading "Delirium" qualified by the subheading "chemically induced" might be used. Similarly, the MeSH heading "Aortography" is not directly found in the Library of Congress Subject Headings (LCSH). A Boolean combination of the LCSH headings "Aorta" and "Radiography" might be used when searching for this concept in online catalogs indexed by the Library of Congress.

A somewhat different and potentially quite powerful set of links is provided between concepts that co-occur in a particular database. These co-occurrence data are pre-computed before each release, making it possible for an application to make use of information much of which cannot be computed in real time or without large computational resources. The co-occurrence of findings and the diseases with which they are associated in AI/RHEUM has been included in the most recent release. For example, "partial hearing loss" and "hemoptysis" both co-occur with the disorder "Wegener's granulomatosis". The majority of the co-occurrence data in the Metathesaurus is derived from the MEDLINE file. The frequency of co-occurring MeSH terms together with the frequency of the subheadings applied has been calculated for the main points in more than eight years of MEDLINE citation records. Main points in citation records are those MeSH index terms that are marked with an asterisk. Because Metathesaurus concepts have been assigned to semantic types, it is possible to present a view of the co-occurrence data at a higher level of generality, to display, for example, the relative frequency with which a particular disorder co-occurs with drugs, organisms or geographic areas. The co-occurrence data can be used by search interfaces to provide the user with a view of what aspects of a topic have or have not been written about in the literature. This, in turn, can help the user search for articles of interest from the categories that are known to occur.

For some information sources, so-called locator information has been computed. Concepts are marked if they appear in selected sources, in particular, MEDLINE, Online Mendelian Inheritance in Man (OMIM), Physician Data Query (PDQ) System, DXplain, Quick Medical Reference (QMR), and AI/RHEUM. In some cases frequency information is listed together with an indication of what the

frequency count is measuring. For MEDLINE it would be the number of times the concept (either a MeSH heading or subheading) has appeared as a main concept in articles indexed in a particular segment of the database. These data are incomplete in the current Metathesaurus except for MEDLINE and AI/RHEUM. System developers might consider adding location information for their own information sources. This would ensure an even stronger link between a local source and the extensive information available in the Metathesaurus and the other two Knowledge Sources.

In addition to the inter-concept data included in the Metathesaurus, many attributes of individual concepts are also included. Some key attributes were created expressly for the Metathesaurus, others are taken from its source vocabularies. Definitions and other kinds of notes or annotations give a more extensive indication of the meaning of a concept. In the most recent release of the Metathesaurus, definitions from the 27th edition of the Dorland's Illustrated Medical Dictionary [39] were added to the Metathesaurus. The Dorland dictionary has some 112,000 entries of which only a small number has been added to the Metathesaurus to date. In subsequent versions many more will be added. Concepts derived from MeSH and AI/RHEUM often include a definition, as well. Metathesaurus entries may include multiple definitions from different sources. In this case, each definition is labelled with its source. In addition, for some controlled vocabularies, a scope note or annotation is included that, while not precisely a definition, does give useful information about the intended use or scope of a concept in that vocabulary.

Special lexical entities such as acronyms, abbreviations, trade names, and drug identification numbers are explicitly labelled. This information, together with syntactic category and inflectional variant information, is useful for natural language processing. Inflectional variants are included only for those vocabularies that explicitly store them. The MeSH vocabulary, in addition to having human-assigned variants, includes algorithmically generated variants for all of its terms. Although a number of algorithms exists for recognizing and/or generating lexical variants, each has been developed for a specific purpose and has certain virtues as well as limitations [40, 41]. The advantage of handling lexical variation algorithmically is that fewer data need to be stored; the disadvantage is that since some of these phenomena depend on the particular lexical item, a program will sometimes give the wrong result. Beginning in 1994, NLM will distribute a set of lexical programs with the UMLS Knowledge Sources that will include both rules and known exceptions to these rules.

All strings are represented in the word index that accompanies the Metathesaurus [42]. The index can be used to identify all concepts, terms and strings that contain a particular word. By searching the word index it is possible, for example, to identify all Metathesaurus concepts that include the word "heart". Some 88 concepts are found in the current version, including, "American Heart Association", "heart aneurysm", "heart auscultation", and "heart catheterization". Note that concepts such as "cardiac volume", "myocardial contraction" and "coronary artery bypass" are found as well, since they each have synonymous terms that contain the word "heart".

Semantic Network

The UMLS Semantic Network provides a consistent view of the concepts represented in the UMLS Metathesaurus. Semantic networks attempt to impart common sense knowledge to computers, allowing them to "reason" and draw conclusions about entities by virtue of the categories to which they have been assigned. The UMLS Semantic Network is a structure for categorizing objects in the biomedical domain. The scope is thus broader than any single vocabulary represented in the Metathesaurus, yet the granularity is uneven. So, while semantic types have been included for organisms, anatomical structures, biologic function, chemicals, behaviors and other activities and concepts and ideas, the depth of these categories varies. Actual use of the Network in a range of applications should help determine which categories will be further refined in subsequent releases of this Knowledge Source.

The early versions of the Network were developed based on analysis of the vocabularies included in the Metathesaurus and based on experiments using the UMLS test collection of queries and MEDLINE citation records [43]. Analysis of existing structured vocabularies yielded a set of high-level categories that resulted in the initial set of semantic types, and work with the UMLS test collection resulted in the initial set of relationships. Concurrently with the work on the Semantic Network, experiments were conducted in making explicit the relationships between MeSH child and parent terms in certain sections of the vocabulary. This led to further candidate relationships for inclusion in the Network. Participation by all UMLS research collaborators resulted in the version of the Network which was released in the fall of 1990 [17, 44]. The current version includes 132 semantic types and 47 relationships between them.

Each concept in the Metathesaurus is assigned to one or more of the semantic types in the Network based on the meaning or meanings that the concept has in its source vocabularies. Assigning semantic types to Metathesaurus concepts involves algorithmic procedures as well as extensive review by subject matter experts. Wherever possible, default semantic types are assigned to concepts by a program. This is possible because most of the constituent vocabularies in the Metathesaurus are already structured, providing useful semantic information. These default assignments are subsequently reviewed by experts who determine if the correct assignment has been made and whether any types need to be added. For some concepts it is not possible to assign default semantic types reliably either because the concept comes from an unstructured or loosely structured source vocabulary, or because its position in a

structured vocabulary does not map easily to a semantic type. In this case, the semantic types are assigned by subject matter experts. In either case, whether the initial assignment has been done algorithmically, or whether it has been done by a subject matter expert, there is further review to ensure accuracy and consistency.

The primary relation in the Semantic Network is the 'isa' link. This links semantic types of greater and lower specificity, establishes the hierarchy of types within the Network, and is used for deciding on the most specific semantic types available for assignment to a Metathesaurus concept. The isa link allows nodes in a hierarchy to inherit information from higher level nodes. The inheritance property allows efficient storage of information, since information that holds true for a higher level node need not be repeated for all lower level nodes. It allows certain generalizations to be captured that otherwise would appear as isolated facts. For example, by grouping all biologic functions together and by grouping all organisms together, it is possible to make one (common sense) statement like "biologic functions are processes of organisms". Procedurally, then, each of the descendants of biologic function and organism inherits this information that was stated only once.

By traversing the isa links it is possible to compute an interpretation for any given node in the Network. For example, a leaf node in the Network is "Medical Device". This is a "Manufactured Object", which is a "Physical Object", which, in turn, is an "Entity". Similarly, a "Disease or Syndrome" is a "Pathologic Function", which is a "Biologic Function", which, in turn, is a "Natural Phenomenon or Process". Simply by traversing the Network, it is possible to see that a medical device is an object and that a disease is a process. By inheritance, any properties that are associated with objects are automatically shared by medical devices, and any properties associated with processes are automatically shared by diseases. Note that these inferences can be made in the absence of any other definitional information and can be done with ease by program.

A non-hierarchical relationship may be thought of as a property that relates concepts or classes of concepts in a network. The non-hierarchical relationships in the Network fall into four categories: physical, functional, temporal, and conceptual relationships. The relations are stated between high-level nodes in the Network whenever possible and are generally inherited by all the children of those nodes. The links indicate what relationships are possible (or permitted). For example, a drug may treat a disease, it may prevent a disease, or it may complicate a disease. A drug may even cause a disease, though the reverse (a disease causing a drug) is not permitted.

In addition to giving an indication of the meaning of individual Metathesaurus concepts, the Semantic Network importantly provides an overall semantic structure for Metathesaurus concepts. Since Metathesaurus concepts are derived from a number of sometimes quite disparate thesauri which have their own structure, the Network serves as a unifying force. It groups together all concepts that share a particular semantic type and allows generalizations to be made about that set of objects. Thus, all diagnostic procedures would be grouped together regardless of whether they appear in the CPT vocabulary, the ICD-9-CM, or MeSH. This means that a generalization that states, for example, that diagnostic procedures measure biological function is applicable to this entire set of concepts.

Information Sources Map

The Information Sources Map (ISM) is a knowledge source which has been developed to describe computerized biomedical information sources. The goal is to provide users with a path to the most appropriate databases based on the particular query posed. ISM records contain highly structured information, drawn in some cases from the other UMLS Knowledge Sources, as well as information intended primarily for humans to read. The current version contains data on some 64 information sources, many of which have been developed and are maintained at NLM, together with others that have been developed by other institutions. The information sources are varied and include not only major bibliographic databases for biomedical research, clinical practice, and bioethics, but also diagnostic expert systems such as AI/RHEUM, DXplain, Iliad, Quick Medical Reference, and factual databases concerned with drugs, toxicology, environmental health, genetics, and protein and nucleic acid sequences. Future editions of the ISM will describe many more information sources.

Four elements in the ISM are used to index the conceptual scope of the information sources: relevant MeSH terms, MeSH subheadings which denote the contexts in which the main MeSH headings are applicable, semantic types from the UMLS Semantic Network, and semantic links, which link two semantic types with a relation from the Semantic Network. The application of indexing terms to information sources is similar to the indexing of the biomedical literature, except that in the case of the literature, the most specific applicable term is chosen, while in the case of the ISM, the most generally applicable term is chosen [25]. An example will illustrate the indexing done for ISM records. The Environmental Mutagen Database Backfile is a database created by the Oak Ridge National Laboratory. It contains citations to publications from 1950–1991 on agents that have been tested for genotoxic activity. Some sample MeSH terms and subheadings that have been assigned to this database are "DNA Damage/drug effects", "Mutation", and "Genes, Lethal/radiation effects". Sample semantic types assigned are "Acquired Abnormality", "Genetic Function", and "Hazardous or Poisonous Substance". Sample Network relations assigned are "Hazardous or Poisonous Substance affects Biologic Function", and "Biologically Active Substance causes Congenital Abnormality".

In addition to characterizing subject scope, ISM records include narrative descriptions of the databases; an

indication of who the intended audience for a database might be; the type of information that is contained, e.g., bibliographic database, knowledge base, full-text database, clinical protocols; the probable uses of the database, e.g., for clinical practice or health services research; the organization that provides the database; the names and addresses of contact individuals; the name of the host system; and sample records from the database itself.

Development of the procedural component of the ISM is in its early stages. Certain emerging standards, such as the ANSI Z39.50 information retrieval protocol, are being considered in the development of the procedural component.

Applying the UMLS Knowledge Sources

Information Retrieval

To improve access to machine-readable biomedical information, the UMLS Knowledge Sources must be exploited by intelligent user interface programs. The UMLS model of biomedical information retrieval includes the Knowledge Sources, many target machine-readable information sources, smart interface programs, and an involved user. For a successful outcome, the user must be willing to interact with the smart programs to clarify ambiguous inquiries, to select among alternatives presented by the system, and to evaluate the relevance of information retrieved. This view of the role of the user is similar to current thinking about how health professionals should interact with expert systems [45].

Although user interaction is essential to the UMLS model, the amount, type, and timing of that interaction will be variable, based on a number of factors, including user interface design, the complexity of the information need, the extent to which the Knowledge Sources cover the topic of interest, and the characteristics and preferences of the user. Current retrieval applications involving the UMLS Knowledge Sources illustrate a range of approaches to assisting the user. Further development and evaluation will be needed to identify the approaches that are most effective for specific tasks, environments, and users.

Effective searching of databases indexed by human-assigned subject headings or codes often depends on the translation of the meaning of a user query into the controlled terminology used in the target database. The synonyms, lexical variants, semantic types, inter-concept relationships, and concept usage information in the UMLS Metathesaurus and the relationships among semantic types represented in the UMLS Semantic Network are potentially useful in this process. A range of applications has explored how information from these two Knowledge Sources can best be organized, displayed, and employed on behalf of the user.

MicroMeSH [46, 47] and MetaCard [48, 49] offer innovative solutions to the problem of displaying complex relationships among strings, terms, and concepts in MeSH and the Metathesaurus, respectively. Researchers at Brigham & Women's Hospital have used a network approach to displaying and navigating the UMLS Semantic Network and the inter-concept relationships from the Metathesaurus [e.g., 50]. Other systems exploit UMLS knowledge without attempting to display its structure. SAPHIRE [51] uses the Metathesaurus' representation of synonymy to identify MeSH terms in free text as part of a concept-based automatic indexing and retrieval strategy which endeavors to apply probabilistic retrieval methods to concepts rather than words. SPECIALIST [52, 53] makes use of a range of information in both Knowledge Sources in research on natural language processing of biomedical text. Coach, an expert system designed to assist Grateful Med users to conduct more effective searches, consults UMLS information on the user's behalf, but also allows the user to explore the Metathesaurus directly through a browser application [54, 55].

Several UMLS investigators are testing the hypotheses that many user inquiries, particularly in clinical environments, are in fact specific instances of a limited number of query types or "generic queries" and that an interface program can offer more intelligent and useful assistance to the user if the query type can be determined. Psychtopix [56] identifies a set of important search topics in the field of psychiatry and then maps information from psychiatric consultation reports to this set, using a knowledge base of DSM-III-R concepts and manually developed MeSH search strategies. The Interactive Query Workstation (IQW) [57] uses the semantic types of the terms in the user's query to determine a subset of relevant query types. Its associated Q & A query formulation assistant uses Metathesaurus co-occurrence data to help users to construct successful searches [58]. Researchers at Columbia are exploring several approaches to identifying the appropriate query type. These include natural language processing techniques that consult the Metathesaurus and the Semantic Network and the development of subsets of query types likely to be of interest in particular clinical contexts [59].

Research done to date confirms that the UMLS Metathesaurus and Semantic Network can be used effectively to aid retrieval of information relevant to specific patient problems. Powsner et al. [60] have built a front-end that uses the Metathesaurus to find and list MeSH terms that are related to words or phrases the user has marked in a machine-readable patient record. The program passes a MEDLINE search strategy to the Grateful Med Search Engine after the user selects terms from the list. The MEDLINE Button [61] attempts to use ICD9-CM to MeSH mapping information in the Metathesaurus to construct MEDLINE searches relevant to particular diagnoses. Using full-text patient records from the MARS system [62], CHARTLINE [63] identifies terms in segments of a patient's chart that are also present in the Metathesaurus, displays co-occurring MeSH terms relevant to this particular patient, and conducts a MEDLINE

search on MARS using terms selected from the display by the user.

While several researchers have concentrated on the problem of connecting clinical information to relevant MEDLINE citations, others have applied the UMLS Knowledge Sources to the task of selecting the databases most likely to contain information relevant to particular queries. The Physician's Information Assistant [64, 65] uses the MetaCard browser interface and the concept location and co-occurrence information in the Metathesaurus to select one or more relevant information sources and then searches these for the user. IQW [57] directs queries to several types of databases, including full-text sources, drug databases, and the DXplain diagnostic system. Miller et al. [66] and Masys [67] have conducted preliminary tests of source selection algorithms that rely on matching information about the user's search terms obtained from the Metathesaurus and the Semantic Network with information about the subject scope of various databases in the prototype Information Sources Map. This work indicates that the Information Sources Map can be used to identify databases likely to be relevant to particular inquiries. A version of the Information Sources Map is currently available through Yale's NetMenu campus-wide network interface [68], which also provides access to a Metathesaurus browser. Users of NetMenu's Information Sources Map receive assistance in selecting relevant databases and then are connected automatically to those they select.

Although much research, development, and evaluation work remains, two general conclusions can be drawn from investigations done to date. First, the UMLS Knowledge Sources are useful for their intended purpose, which is to facilitate retrieval of relevant information from a variety of machine-readable sources. Second, the most effective retrieval interfaces are likely to employ a range of different approaches to assisting users, depending on user preference, the type of inquiry, and the amount of machine-readable "context", available to assist in automated interpretation of the user's meaning. The utility of offering a variety of user aids is amply illustrated in a number of currently available user interfaces, including Grateful Med.

Indexing and Data Creation Applications

Designed to support information retrieval, the content and structure of the UMLS Metathesaurus and Semantic Network are also potentially useful to those involved in data creation, indexing, or encoding. By linking concept names from different terminologies, the Metathesaurus creates a richer set of synonyms and concept relationships than is present in any single source. This larger entry vocabulary can help people or programs to locate the appropriate preferred term or code in the specific vocabulary or classification being applied, assuming that it is represented in the Metathesaurus. Hersh [51] and Wagner and Cooper [69] have applied the Metathesaurus in automated indexing of biomedical text and image descriptions, respectively. Chute et al. have applied UMLS Knowledge Sources to latent semantic indexing of diagnoses [70]. The MedIndex [71] expert indexing assistant program makes use of knowledge in the Metathesaurus and the Semantic Network.

Because the Metathesaurus provides uniform access to terminology from an array of vocabulary sources, it facilitates the review and analysis of existing controlled vocabularies and the construction of lists of concepts and terms suitable for specific data creation and indexing tasks. Eisner [72] has constructed a core vocabulary for use in describing the content of dental curricula by extracting some terminology from the Metathesaurus and augmenting it with terms from other sources. Borrowing from one or more existing controlled vocabularies is more efficient than creating a new vocabulary de novo and offers better potential for linking the information being captured to related information in other machine-readable sources.

As stated previously, the scope of the UMLS Metathesaurus is determined by the combined scope of the vocabularies and classification it encompasses. Its coverage of specific clinical concepts that are needed in patient records mirrors that of its source vocabularies and increases as more clinical vocabularies are incorporated. The UMLS project links vocabularies at the individual concept level and groups all concepts by broad semantic types. It does not attempt to reconcile the differences in hierarchical perspective and specificity among its source vocabularies. For this reason, the Metathesaurus and the Semantic Network do not constitute a single consistent classification suitable for indexing or coding detailed clinical content [73]. The Metathesaurus adds value to existing clinical vocabularies by linking them to a rich array of synonyms, variants, related concepts, hierarchical perspectives, and other useful semantic and syntactic information. The added features provided by the Metathesaurus can therefore facilitate manual and automated encoding of patient data in one or more of its constituent vocabularies. The Metathesaurus can also support automated connections between patient records and other types of information that can improve health care decisions, such as practice guidelines, expert systems, and the current literature [74]. The combination of the Metathesaurus and the Semantic Network may also be useful in evaluating the coverage of existing clinical vocabularies and in building better ones [75].

Several studies have examined the extent to which different editions of the UMLS Metathesaurus cover clinical concepts [76–82]. The addition of concepts from a variety of clinical vocabularies and classifications is a high priority for the UMLS project and is proceeding rapidly influenced by feedback from these studies. The 1993 edition includes all preferred terms and codes for ICD9-CM diagnoses, and work is beginning on incorporation of concepts and terminology from SNOMED III.

The UMLS and Emerging International High-Speed Networks

Low-cost powerful workstations, high-capacity world-wide networks, and the move to flexible client-server architectures are all contributing to the rapid growth in the amount and types of machine-readable information that are technically accessible to anyone with an Internet connection. The volume of information available has been described as overwhelming and bewildering. At the same time the number of Internet users is also growing rapidly, and the use of electronic mail, bulletin boards, FTP, etc. is transforming the way many people work. A number of important tools, such as Knowbots, WAIS, Gopher, and World Wide Web [83], and their many combinations and derivatives, are emerging to assist users in navigating the Internet and in identifying and locating potentially useful information sources. The use of these tools has increased the visibility of the problems the UMLS project is attempting to solve. For most users, technical access to hundreds or thousands of databases is of no value unless it is accompanied by some means of determining which ones contain information useful in the current circumstances and by the ability to frame search inquiries in terms the relevant databases can understand. If these conditions are not met, access to hundreds of information sources is actually less useful than access to a few.

As technical barriers to information access are removed, better semantic connections become even more important. Fortunately, the UMLS project and current network developments are converging in ways that are potentially beneficial to the worldwide biomedical community. The UMLS Knowledge Sources and the intelligent interface programs that make use of them can supply, for the biomedical domain, the missing conceptual link that is needed to isolate relevant information in the masses of machine-readable data available on the Internet. In turn, the Internet provides a powerful tool for distributed access to and maintenance of the UMLS Knowledge Sources, including the very large Metathesaurus files. Use of high-performance computers accessible on the network can speed some of the time-consuming vocabulary and database analysis tasks associated with Metathesaurus construction. Internet access can also solve some of the problems encountered by those attempting to apply the Knowledge Sources in modest hardware and software environments.

The increasing availability of biomedical databases on the Internet simplifies the task of the UMLS Information Sources Map. The ongoing development of tools and protocols for connecting to and searching Internet-accessible information sources is likely to provide solutions to many of the basic connectivity issues that must be addressed to achieve a fully functional UMLS Information Sources Map. The combination of these networking advances and the semantic connections provided by the UMLS should speed progress toward the goal of seamless retrieval and integration of machine-readable biomedical information, including images and computer-based patient records.

The UMLS project has much to gain and much to offer in the new era of high-performance computing and communications. The UMLS development team is only beginning to identify and test the ways in which these technical developments can help to make the UMLS goals a reality for today's health professionals and biomedical researchers.

REFERENCES

[1] Lindberg DAB, Humphreys BL. Computer systems that understand medical meaning. In: *Computerized Natural Medical Language Processing for Knowledge Representation.* Scherrer JR, Cote RA, Mandil SH (eds). Amsterdam: Elsevier Science Publishers, 1989; 5–17.

[2] Bunting A. The Nation's Health Information Network: History of the Regional Medical Library Program, 1965–1985. Bull Med Libr Assoc 1987; 75 (3 Suppl): 1–62.

[3] Integrated Academic Information Management Systems (IAIMS) model development. Bull Med Libr Assoc 1988; 76: 221–67.

[4] Lorenzi NM (ed). Symposium: A decade of IAIMS. Bull Med Libr Assoc 1992; 80: 241–93.

[5] Humphreys BL, Lindberg DAB. The UMLS project: Making the conceptual connection between users and the information they need. Bull Med Libr Assoc 1992; 81 (2): 170–7.

[6] Lindberg DAB, Siegel ER, Rapp BA, Wallingford KT, Wilson SR. Use of MEDLINE by physicians for clinical problem solving. JAMA 1993; 269: 3124–9.

[7] Marshall JG. The impact of the hospital library on clinical decision making: The Rochester Study. Bull Med Libr Assoc 1992; 80: 169–78.

[8] Rossi Mori A, Berrauer J, Pakarinen V, et al. Models for representation of terminologies and coding systems in medicine. In: De Moor GJE, McDonald CJ, Noothoven van Goor J (eds). *Progress in Standardization in Health Care Informatics.* Amsterdam: IOS Press, 1993; 92–104.

[9] McDonald CJ. ANSI's Health Informatics Planning Panel (HISPP) – The Purpose and Progress. In: *Progress in Standardization in Health Care Informatics.* De Moor GJE, McDonald CJ, Noothoven van Goor J (eds). Amsterdam: IOS Press, 1993; 14–9.

[10] ASN.1 Specifications. Bethesda MD: National Center for Biotechnology Information, Nov 1991.

[11] Dutcher GA. DOCLINE: A national automated interlibrary loan request routing and referral system. Inf Technol Libr 1989; 8: 359–70.

[12] Humphreys BL, Lindberg DAB. The Unified Medical Language System Project: a distributed experiment in improving access to biomedical information. In: *MEDINFO 92: Proceedings of the 7th World Congress on Medical Informatics.* Amsterdam: North-Holland Publ Comp, 1992; 1496–1500.

[13] Humphreys BL, Lindberg DAB, Hole WT. Assessing and enhancing the value of the UMLS Knowledge Sources: In: *Proceedings of the 15th Annual Symposium on Computer Applications in Medical Care.* Clayton PD (ed). New York: McGraw Hill, 1991; 78–82.

[14] Humphreys BL, Lindberg DAB. Building the Unified Medical Language System. In: Kingsland LC III (ed). *Proceedings of the 13th Annual Symposium on Computer Applications in Medical Care.* New York: IEEE Computer Society Press, 1989; 475–80.

[15] Masarie FE Jr, Miller RA, Bouhaddou O, Giuse NB, Warner HR. An interlingua for electronic interchange of medical information: using frames to map between clinical vocabularies. Comput Biomed Res 1991; 24: 379–400.

[16] Evans DA. Pragmatically-structured, lexical-semantic knowledge bases for Unified Medical Language Systems. In: Greenes RA (ed). *Proceedings of the 12th Annual Symposium on Computer Applications in Medical Care.* New York: IEEE Computer Society Press, 1988; 169–73.

[17] Barr CE, Komorowski HJ, Pattison-Gordon E, Greenes RA. Conceptual modeling for the Unified Medical Language System. In: Greenes RA (ed). *Proceedings of the*

12th Annual Symposium on Computer Applications in Medical Care. IEEE Computer Society Press, 1988; 148–51.

[18] Tuttle MS, Blois MS, Erlbaum MS, Nelson SJ, Sherertz DD. Toward a biomedical thesaurus: building the foundation of the UMLS. In: Greenes RA (ed). *Proceedings of the 12th Annual Symposium on Computer Applications in Medical Care.* New York: IEEE Computer Society Press, 1988; 191–5.

[19] Sherertz DD, Tuttle MS, Blois MS, Erlbaum MS. Intervocabulary mapping within the UMLS: the role of lexical matching. In: Greenes RA (ed). *Proceedings of the 12th Annual Symposium on Computer Applications in Medical Care.* New York: IEEE Computer Society Press, 1988; 201–6.

[20] Cimino JJ, Mallon LJ, Barnett GO. Automated extraction of medical knowledge from MEDLINE citations. In: Greenes RA (ed). *Proceedings of the 12th Annual Symposium on Computer Applications in Medical Care.* New York: IEEE Computer Society Press, 1988; 180–4.

[21] Musen MA. Dimensions of knowledge sharing and reuse. Comp Biomed Res 1992; 25: 435–67.

[22] McCray AT. The UMLS semantic network. In: Kingsland LC III (ed). *Proceedings of the 13th Annual Symposium on Computer Applications in Medical Care.* New York: IEEE Computer Society Press, 1989; 503–7.

[23] McCray AT, Hole WT. The scope and structure of the first version of the UMLS semantic network. In: Miller RA (ed). *Proceedings of the 14th Annual Symposium on Computer Applications in Medical Care.* New York: IEEE Computer Society Press, 1990; 126–30.

[24] McCray AT. Representing Biomedical Knowledge in the UMLS Semantic Network. In: Broering NC (ed). *High Performance Medical Libraries.* Westport CT: Meckler Publ, 1993 (in press).

[25] Masys DR, Humphreys BL. Structure and function of the UMLS Information Sources Map. In: MEDINFO 92: Proceedings of the 7th World Congress on Medical Informatics. Lun KC, Degoulet P, Piemme TE, Rienhoff O (eds). Amsterdam: North-Holland Publ Comp, 1992; 1518–21.

[26] Lindberg DAB, Humhreys BL. The UMLS knowledge sources: tools for building better user interfaces. In: Proceedings of the 14th Annual Symposium on Computer Applications in Medical Care. Miller RA (ed). New York: IEEE Computer Society Press, 1990; 121–5.

[27] Tuttle MS, Sperzel WD, Olson NE et al. The homogenization of the Metathesaurus schema and distribution format. In: Proceedings of the 16th Annual Symposium on Computer Applications in Medical Care. Frisse ME (ed). New York: McGraw Hill, 1992; 299–303.

[28] Tuttle M, Sherertz D, Erlbaum M, Olson N, Nelson S. Implementing Meta-1: the first version of the UMLS Metathesaurus. In: Proceedings of the 13th Annual Symposium on Computer Applications in Medical Care. Kingsland LC III (ed). New York: IEEE Computer Society Press, 1989; 483–7.

[29] Sherertz DD, Olson NE, Tuttle MS, Erlbaum MS. Source inversion and matching in the UMLS Metathesaurus. In: Proceedings of the 14th Annual Symposium on Computer Applications in Medical Care. Miller RA (ed). New York: IEEE Computer Society Press, 1990; 141–5.

[30] Sherertz DD, Olson NE, Tuttle MS, Sperzel WD, Erlbaum MS, Fuller LF. The META-1* Engine: a database methodology used in building the UMLS Metathesaurus. In: MEDINFO 92: Proceedings of the 7th World Congress on Medical Informatics. Lun KC, Degoulet P, Piemme TE, Rienhoff O (eds). Amsterdam: North-Holland Publ Comp, 1992; 144–9.

[31] Sperzel D, Erlbaum M, Fuller L et al. Editing the UMLS Metathesaurus: review and enhancement of a computed knowledge source. In: Proceedings of the 14th Annual Symposium on Computer Applications in Medical Care. Miller RA (ed). New York: IEEE Computer Society Press, 1990; 136–40.

[32] Sperzel WD, Tuttle MS, Olson NE et al. The Meta-1.2 engine: a refined strategy for linking biomedical vocabularies. In: Proceedings of the 16th Annual Symposium on Computer Applications in Medical Care. Frisse ME (ed). New York: McGraw Hill, 1992; 204–6.

[33] Tuttle MS, Sherertz DD, Nelson SJ et al. Adding your terms and relationships to the UMLS Metathesaurus. In: Proceedings of the 15th Annual Symposium on Computer Applications in Medical Care. Clayton PD (ed). New York: McGraw Hill, 1991; 219–23.

[34] Humphreys BL, Schuyler PL. The Unified Medical Language System: Moving Beyond the Vocabulary of Bibliographic Retrieval. In: Broering NC (ed). *High Performance Medical Libraries.* Westport CT: Meckler Publ, 1993 (in press).

[35] Schuyler PL, Hole WT, Tuttle MS, Sherertz DD. The UMLS Metathesaurus: Representing different views of biomedical concepts. Bull Med Libr Assoc 1993; 81: 217–22.

[36] Metathesaurus vocabularies are listed in the UMLS documentation: UMLS Knowledge Sources. 3rd Experimental Edition – August 1992. Documentation. Bethesda Md: National Library of Medicine.

[37] Roget's International Thesaurus. Fourth Edition. New York: Harper & Row Publ, 1984.

[38] Neslon SJ, Tuttle MS, Cole WG et al. From meaning to term: semantic locality in the UMLS Metathesaurus. In: Proceedings of the 15th Annual Symposium on Computer Applications in Medical Care. Clayton PD (ed). New York: McGraw Hill, 1991; 209–13.

[39] Dorland's Illustrated Medical Dictionary. 27th edition. Philadelphia: W. B. Saunders Company, 1988.

[40] Byrd RJ, Klavans JL, Aronoff M, Anshen F. Computer methods for morphological analysis. In: Proceedings of the 24th Annual Meeting of the ACL. Morristown NJ: Association for Computational Linguistics 1986; 120–7.

[41] See the extensive references listed in: Morphology and Computation. Sproat R (ed). Cambridge Ma: MIT Press, 1992.

[42] The UMLS documentation includes the C code that generates the word index: UMLS Knowledge Sources. 3rd Experimental Edition – August 1992. Documentation. Bethesda Md: National Library of Medicine.

[43] Schuyler PL, McCray AT, Schoolman HM. A test collection for experimentation in bibliographic retrieval. In: MEDINFO 89: Proceedings of the Sixth Conference on Medical Informatics. Barber B, Cao D, Qin D, Wagner G (eds). Amsterdam: North-Holland Publ Comp, 1989; 910–12.

[44] Miller PL, Barwick KW, Morrow JS, Powsner SM, Riely CA. Semantic relationships and medical bibliographic retrieval: a preliminary assessment. Comput Biomed Res 1987; 21: 64–77.

[45] Miller RA, Masarie FE Jr. The demise of the "Greek Oracle" model for medical diagnostic systems. Meth Inf Med 1990; 29: 1–2.

[46] Lowe HJ, Barnett GO. MicroMeSH: a microcomputer system for searching and exploring the National Library of Medicine's Medical Subject Headings (MeSH) vocabulary. In: Proceedings of the 11th Annual Symposium on Computer Applications in Medical Care. Stead WE (ed). New York: IEEE Computer Society Press, 1987; 717–20.

[47] Lowe H, Barnett GO, Scott J, Mallon L, Ryan-Blewett D. Remote access MicroMeSH: evaluation of a microcomputer system for searching the MEDLINE database. In: Proceedings of the 13th Annual Symposium on Computer Applications in Medical Care. Kingsland LC III (ed). New York: IEEE Computer Society Press, 1989; 445–7.

[48] Sherertz DD, Tuttle M, Cole W, Erlbaum M, Olson N, Nelson S. A HyperCard implementation of Meta-1: the first version of the UMLS Metathesaurus. In: Proceedings of the 13th Annual Symposium on Computer Applications in Medical Care. Kingsland LC III (ed). New York: IEEE Computer Society Press, 1989; 1017–18.

[49] Sherertz DD, Tuttle MS, Erlbaum MS, Sperzel WD, Fuller LF, Nelson SJ. m-CD: a HyperCard interface to Meta-1 on a CD-ROM. In: Proceedings of the 14th Annual Symposium on Computer Applications in Medical Care. Miller RA (ed). New York: IEEE Computer Society Press, 1990; 1052–3.

[50] Komorowski HJ, Greenes RA, Barr C, Pattison-Gordon E. Browsing and authoring tools for a Unified Medical Language System. In: Proceedings of RIAO 88 Conference on User-oriented Content-based Text and Image Handling. Fluhr C, Walker

D (eds). Cambridge Ma: MIT Press, 1988; 624–41.

[51] Hersh WR. Evaluation of Meta-1 for a concept-based approach to the automated indexing and retrieval of bibliographic and full-text databases. Med Decis Making 1991; 11 (4 Suppl): 120–4.

[52] McCray AT. Extending a natural language parser with UMLS knowledge. In: Proceedings of the 15th Annual Symposium on Computer Applications in Medical Care. Clayton PD (ed). Applications in Medical Care. New York: McGraw Hill 1991; 194–8.

[53] McCray AT, Aronson AR, Browne AC et al. UMLS knowledge for biomedical language processing. Bull Med Libr Assoc 1993; 81: 184–94.

[54] Kingsland LC III, Syed EJ, Lindberg DAB. COACH: an expert searcher program to assist Grateful Med users searching MEDLINE. In: MEDINFO 92: Proceedings of the 7th World Congress on Medical Informatics. Lun KC, Degoulet P, Piemme TE, Rienhoff O (eds). Amsterdam: North-Holland Publ Comp 1992; 382–6.

[55] Kingsland LC, Harbourt AM, Syed EJ, Schuyler PL. Applying the UMLS Knowledge Sources in an expert searcher environment. Bull Med Libr Assoc 1993; 81: 178–83.

[56] Powsner SM, Miller PL. Linking bibliographic retrieval to clinical reports: PsychTopix. In: Proceedings of the 13th Annual Symposium on Computer Applications in Medical Care. Kingsland LC III (ed). New York: IEEE Computer Society Press, 1989; 431–5.

[57] Cimino C, Barnett GO, Hassan L, Blewett DR, Piggins JL. Interactive Query Workstation: Standardizing access to computer-based medical resources. Comput Meth Progr Biomed 1991; 35: 293–9.

[58] Merz RA, Cimino C, Barnett GO et al. Q & A: a query formulation assistant. In: Proceedings of the 16th Annual Symposium on Computer Applications in Medical Care. Frisse ME (ed). New York: McGraw Hill, 1992; 498–504.

[59] Cimino JJ, Aguirre A, Johnson SB, Peng P. Generic Queries for Meeting Clinical Information Needs. Bull Med Libr Assoc 1993; 81: 195–206.

[60] Powsner SM, Miller PL. From patient reports to bibliographic retrieval: a Meta-1 front-end. In: Proceedings of the 15th Annual Symposium on Computer Applications in Medical Care. Clayton PD (ed). New York: McGraw Hill, 1991; 526–30.

[61] Cimino JJ, Johnson SB, Aguirre A, Roderer N, Clayton PD. The Medline button. In: Proceedings of the 16th Annual Symposium on Computer Applications in Medical Care. Frisse ME (ed). New York: McGraw Hill, 1992; 81–5.

[62] Yount RJ, Vries JK, Council CD. The Medical Archival System: An information retrieval system based on distributed parallel processing. Inform Processing Management 1991; 27: 379–91.

[63] Miller RA, Gieszczykiewicz FM, Vries JI, Cooper GF. CHARTLINE: providing bibliographic references relevant to patient charts using the UMLS Methathesaurus Knowledge Sources. In: Proceedings of the 16th Annual Symposium on Computer Applications in Medical Care. Frisse ME (ed). New York: McGraw Hill, 1993; 91–8.

[64] Nelson SJ, Sherertz DD, Tuttle MS, Erlbaum MS, Smith LG. Browsing MIM. PDQ, CDD, and MEDLINE using MetaCard. In: Proceedings of the 14th Annual Symposium on Computer Applications in Medical Care. Miller RA (ed). New York: IEEE Computer Society Press, 1990; 1054–5.

[65] Nelson SJ, Sherertz DD, Tuttle MS. Issues in the development of an information retrieval system: The Physician's Information Assistant. In: MEDINFO 92: Proceedings of the 7th World Congress on Medical Informatics. Lun KC, Degoulet P, Piemme TE, Rienhoff O (eds). Amsterdam: North-Holland Publ Comp, 1992; 371–5.

[66] Miller PL, Wright LW, Frawley SJ, Clyman JI, Powsner SM. Selecting relevant information resources in a network-based environment: the UMLS Information Sources Map. In: MEDINFO 92: Proceedings of the 7th World Congress on Medical Informatics. Lun KC, Degoulet P, Piemme TE, Rienhoff O (eds). Amsterdam: North-Holland Publ Comp, 1992; 1512–7.

[67] Masys DR. The evaluation of the source selection elements of the prototype UMLS Information Sources Map. In: Proceedings of the 16th Annual Symposium on Computer Applications in Medical Care. Frisse ME (ed). New York: McGraw Hill, 1992; 295–8.

[68] Clyman JI, Powsner SM, Paton JA, Miller PL. Using a network menu and the UMLS Information Sources Map to facilitate access to online reference materials. Bull Med Libr Assoc 1993; 81: 207–16.

[69] Wagner MM, Cooper GF. Evaluation of a Meta-1-based automatic indexing method for medical documents. Comput Biomed Res 1992; 25: 336–49.

[70] Chute CG, Yang Y, Evans DA. Latent Semantic Indexing of medical diagnoses using UMLS semantic structures. In: Proceedings of the 15th Annual Symposium on Computer Applications in Medical Care. Clayton PD (ed). New York: McGraw Hill, 1992; 185–9.

[71] Humphrey SM. Indexing biomedical documents: From thesaural to knowledge-based retrieval systems. Artif Intell Med 1992; 4: 29–56.

[72] Eisner J. The developing Electronic Curriculum Consortium. J Dent Educ, 1990; 54: 598–9.

[73] Bishop CW, Ewing P. Presenting medical knowledge: Reconciling the present or creating the future? MD Comput 1992; 9: 218–25.

[74] Lindberg DAB, Humphreys BL. The Unified Medical Language System (UMLS) and computer-based patient records. In: Aspects of the Computer-based Patient Record. Ball MJ, Collen MF (eds). New York: Springer-Verlag, 1992; 165–75.

[75] Cimino JJ, Hripcsak G, Johnson SB, Friedman C, Fink DJ, Clayton PD. UMLS as knowledge base – a rule-based expert system approach to controlled medical vocabulary management. In: Proceedings of the 14th Annual Symposium on Computer Applications in Medical Care. Miller RA (ed). New York: IEEE Computer Society Press, 1990; 175–9.

[76] Campbell JR, Kallenberg GA, Sherrick RC. The clinical utility of META: an analysis for hypertension. In: Proceedings of the 16th Annual Symposium on Computer Applications in Medical Care. Frisse ME (ed). New York: McGraw Hill, 1992; 397–404.

[77] Chute CG, Yang Y, Tuttle MS, Sherertz DD, Olson NE, Erlbaum MS. A preliminary evaluation of the UMLS Metathesaurus for patient record classification. In: Proceedings of the 14th Annual Symposium on Computer Applications in Medical Care. Miller RA (ed). New York: IEEE Computer Society Press, 1990; 161–5.

[78] Cimino JJ. Representation of clinical laboratory terminology in the Unified Medical Language System. In: Proceedings of the 15th Annual Symposium on Computer Applications in Medical Care. Clayton PD (ed). New York: McGraw Hill, 1991; 199–203.

[79] Friedman C. The UMLS coverage of clinical radiology. In: Proceedings of the 16th Annual Symposium on Computer Applications in Medical Care. Frisse ME (ed). New York: McGraw Hill, 1992; 307–16.

[80] Huff SM, Warner HR. A comparison of Meta-1 and HELP terms: implications for clinical data. In: Proceedings of the 14th Annual Symposium on Computer Applications in Medical Care. Miller RA (ed). New York: IEEE Computer Society Press, 1990; 166–9.

[81] Zielstorff RD, Cimino C, Barnett GO, Hassan L, Blewett DR. Representation of nursing terminology in the UMLS Metathesaurus: a pilot study. In: Proceedings of the 16th Annual Symposium on Computer Applications in Medical Care. Frisse ME (ed). New York: McGraw Hill, 1992; 392–6.

[82] Sato L, McClure RC, Rouse RL, Schatz CA, Greenes RA. Enhancing the Metathesaurus with clinically relevant concepts: anatomic representations. In: Proceedings of the 16th Annual Symposium on Computer Applications in Medical Care. Frisse ME (ed). New York: McGraw Hill, 1992; 388–91.

[83] Schwartz MF, Emtage A, Kahle B, Neuman BC. A comparison of Internet resource discovery approaches. Comput Systems 1992; 5: 461–93.

Address of the authors:
Donald A. B. Lindberg, M. D.,
National Library of Medicine,
Bethesda MD, 20894,
USA