Editorial Comments JAMIA

Making the Conceptual Connections:

The UMLS after a Decade of Research and Development

The UMLS Section in this issue of JAMIA is dedicated to the memory of Marsden Scott Blois, Jr., a pioneer in medical concept representation. Dr. Blois, an internationally recognized physician-investigator and medical informatician, made important contributions to melanoma research and to biomedical informatics during his distinguished and varied career. His keen insight in the early days of the UMLS project contributed significantly to the future direction and success of the project.

This issue of JAMIA helps mark the tenth anniversary of the Unified Medical Language System (UMLS) project. From the beginning, the project has focused on overcoming the barriers users face when attempting to interact with computerized health information systems. The UMLS developers and their collaborators envisioned and then created a set of knowledge sources designed to support the development of sophisticated and accessible information systems. The UMLS knowledge sources have grown to now encompass the Metathesaurus (the largest of the knowledge sources), the Semantic Network, the Information Sources Map, and the SPECIALIST lexicon with its accompanying lexical programs. The Metathesaurus integrates more than 30 biomedical thesauri. The most recent release contains 331,756 concepts named by 739,439 different terms, including translations of some of the terminology into several other languages. Metathesaurus concepts are assigned semantic types from the Semantic Network, which, through its 135 semantic types and 51 relationships, provides a unifying semantic structure for the Metathesaurus terminology. The Information Sources Map is designed to serve as a resource for identifying databases and other information resources that are relevant to users' particular queries. The SPECIALIST lexicon is designed for use in natural language processing applications. The accompanying lexical programs work together with the lexicon (currently containing some 90,000 lexical records) in recognizing lexical variation in biomedical terminologies and texts. The UMLS data are distributed on CD-ROM, but, increasingly, researchers are gaining access by using the UMLS Knowledge Source Server, which provides flexible Internet-based access to all the knowledge sources.[1]

The papers included here necessarily represent only a small portion of the active research that has been conducted on the UMLS over the past ten years. For more detail, readers should examine the recently compiled bibliography of some 280 selected citations on the UMLS project.[2] The current issue reflects a spectrum of publications that characterize UMLS research and development efforts. The contributions include a historical perspective on the UMLS project by its leaders[3]; a viewpoint paper by a prominent group not directly involved in core UMLS efforts on the role the UMLS now plays in addressing basic issues related to terminology development and use[4]; a paper on anatomic knowledge representation that proposes certain extensions to the UMLS Semantic Network and Metathesaurus[5]; a paper that describes semantically based methods to improve UMLS content quality and maintenance techniques[6]; and three reports on experiments applying the UMLS to specific clinical objectives.[7–9] Two papers[7,9] represent international work on using and extending UMLS knowledge in health information systems.

A decade of UMLS research has seen the development and testing of a rich set of continuously evolving knowledge sources. These knowledge sources are distributed regularly to the research community, a decision that was motivated by the belief that systems and

tools improve through actual testing and use. Since early 1997, when the eighth annual release of the knowledge sources was announced, over 500 individuals and organizations worldwide have requested and received access to this eighth edition. Since the first release of the knowledge sources in 1990, investigators have experimented with the UMLS data, and particularly its Metathesaurus, in a variety of application areas.[2,3] The UMLS—which incorporates multiple terminologies, each designed for its own purposes and users—has, through its rich set of interrelationships, resulted in an evolving system that is greater than the sum of its component parts.[4] As the coverage of the UMLS grows and tools to improve its update and maintenance flourish,[6] it should find increased applicability in supporting research and education in the basic clinical sciences (e.g., Rosse et al.[5]), in bibliographic and information science research and products (e.g., the Internet Grateful Med Fact Sheet[10]), and in the clinical arena (e.g., Joubert et al.,[7] Cooper and Miller,[8] and Bodenreider et al.[9])—ALEXA T. MCCRAY, PHD, AND RANDOLPH A. MILLER, MD

References ■

1. McCray AT, Razi AM, Bangalore AK, Browne AC, Stavri PZ. The UMLS knowledge source server: a versatile Internet-based research tool. Proc AMIA Annu Fall Symp. 1996; 164–8. URL: ⟨http://umlsks.nlm.nih.gov/⟩.
2. Selden C, Humphreys BL. Unified Medical Language System (UMLS): January 1986 through December 1996. Current Bibliographies in Medicine. Bethesda, MD: National Library of Medicine, 1997.
3. Humphreys BL, Lindberg DAB, Schoolman HM, Barnett GO. The Unified Medical Language System: An informatics research collaboration. J Am Med Inform Assoc. 1998;5: 1–11.
4. Campbell KE, Oliver DE, Shortliffe EH. The UMLS: A collaborative approach toward solving terminologic problems. J Am Med Inform Assoc. 1998;5:12–16.
5. Rosse C, Mejino JL, Modayur BR, Jakobovits R, Hinshaw KP, Brinkley JF. Motivation and organizational principles for the Digital Anatomist Symbolic Knowledge Base: An approach toward standards in anatomical knowledge representation. J Am Med Inform Assoc. 1998;5:17–40.
6. Cimino JJ. Auditing the Unified Medical Language System with semantic methods. J Am Med Inform Assoc. 1998;5: 41–51.
7. Joubert M, Fieschi M, Robert J-J, Volot F, Fieschi D. UMLS-based conceptual queries to biomedical information databases: an overview of the project ARIANE. J Am Med Inform Assoc. 1998;5:52–61.
8. Cooper GF, Miller RA. An experiment comparing lexical and statistical methods for extracting MeSH terms from clinical free text. J Am Med Inform Assoc. 1998;5:62–75.
9. Bodenreider O, Burgun A, Botti G, Fieschi M, Le Beux P, Kohler F. Evaluation of the use of the Unified Medical Language System as a medical knowledge source: a case study. J Am Med Inform Assoc. 1998;5:76–87.
10. Internet Grateful Med Fact Sheet. Bethesda, MD: National Library of Medicine, 1996. URL: ⟨http://igm.nlm.nih.gov/⟩.

Affiliations of the authors: National Library of Medicine, Bethesda, MD (ATM); Vanderbilt University Medical Center, Nashville, TN (RAM).

Correspondence and reprint requests: Alexa T. McCray, PhD, National Library of Medicine, Building 38A, Room BIN-28H, 8600 Rockville Pike, Bethesda, MD 20894.
e-mail: ⟨mccray@nlm.nih.gov⟩.

Received for publication: 9/12/97; accepted for publication: 9/16/97.

■ **JAMIA.** 1998;5:129–130.

■

It's the Information That's Important, Not the Technology

This issue of JAMIA contains a paper by Dr. Nordyke and Dr. Kulikowski that describes the interplay between a single specialty practice and a structured medical record/computer database over a period of 35 years.[1] The story paints a clear picture of the many ways in which access to longitudinal information about a population of like patients can be used to improve practice and to satisfy intellectual curiosity over the course of the day.

The story presents a contrast to much of the work on computer-based patient records. The sole focus is on the organization and utilization of information. The computer is used only for those parts of the process where it does something that cannot be done reasonably any other way (retrieval and analysis of a population of records) or something that can be done without added work (generation of patient reports). Structured data-capture forms and printed reports are used as the provider interface, avoiding use of the computer where it can be expensive and troublesome. Along the way, the practice experimented with a variety of decision support strategies, but found that they got in the way of practice and were less useful than straightforward access to the data about their patients.

At first glance, the reader may question whether this story is relevant today. Technology is rapidly removing the barriers to direct provider–computer interaction, and we are learning how to incorporate decision support into the flow of practice. Work on these innovations is important, and it will increase the art of the possible over time. Few of the people involved in this work, however, have been as effective in using data to inform and change their practices as have the authors of this case study.

The story has an important message for Medical Informatics today. We can make a big difference if we help people organize and access data to understand and improve practice patterns. We can establish credibility by using whatever processes and technology will work now to achieve that goal. In the process we can buy time to solve the challenges of technology and knowledge/data representation/exchange that must to be solved if we are to achieve what this one specialty practice did—across sites, specialties, and a patient's life time.—WILLIAM W. STEAD, MD

Reference ■

1. Nordyke RA, Kulikowski CA. An informatics-based chronic disease practice: case study of a 35-year computer-based longitudinal record system. J Am Med Inform Assoc. 1998;5: 88–103.

Affiliation of the author: Vanderbilt University, Nashville, TN.

Correspondence and reprints: William W. Stead, MD, The Annette and Irwin Eskind Biomedical Library, Vanderbilt University Medical Center, Nashville, TN 37232-8340. e-mail: ⟨bill.stead@mcmail.vanderbilt.edu⟩.

Received for publication: 9/18/97; accepted for publication: 9/18/97.

■