JAMIA
Letters to the Editor

UMLS Concept Indexing for Production Databases: A Feasibility Study

To the Editor:—In the recently published study by Nadkarni et al.,[1] the authors used text-mining software to extract concepts from clinical documents. Matching of these concepts was attempted with the UMLS 99 Metathesaurus. Matches were then categorized as true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) from 8,745 terms in a "training set" and 1,701 terms in a "test set," for a total of 10,446 terms. True positives were reported as 82.6 percent for the training set and 76.3 percent for the test set.

In 1999, we carried out an almost identical study using the identical version of the UMLS, on a larger scale, which resulted in very similar results that were presented at the 1999 AMIA Annual Symposium.[2] In our study, 4,994 of the most frequently referenced terms were chosen from 1,000,000 terms randomly extracted from the general Mayo Clinic Master Sheet Index and the Impression/Report/Plan section of the Mayo Clinic clinical notes system, to form a general medicine set. The Mayo Clinic Department of Dermatology independently developed a lexicon of 9,050 unique terms describing lesions photographed in their practice, which formed a specialty-specific set. We used automated term composition and the UMLS to assess match rates. In addition, we looked at match rates on our total 14,044 terms based on filtering using the UMLS semantic types.

Comparison of the data from the two studies (Table 1) reveals striking similarities.

What we recognized in 1999, which was omitted from the analysis of Nadkarni et al., was that other metrics are important in the clinical interpretation of these data. Representing the data as shown in Table 1 allows for useful combinations. The true-positive rate is the number of true positives divided by the sum of true positives and false negatives (TP/[TP + FN]), yielding sensitivity. Similar calculation of specificity (TN/[TN+FP]), positive predictive value (TP/[TP + FP]), and positive likelihood ratio (sensitivity/[1 − specificity]) can be carried out. When these combinations are done, it is evident that concept matching in the UMLS is actually much better than was implied by the only true-positive incidence quoted by Nadkarni et al. (Table 2).

The differences in these metrics across the data sets are related to the relatively liberal definition of true positives and the relatively strict definition of false positives given by Nadkarni et al. Unlike them, we did take negation into account when determining true positives. Their definition of false positive was limited to acronyms, abbreviations, spelling/gram-

Table 1 ■

Comparison of Data

	Nadkarni et al.[1]		McDonald et al.[2]	
	Training Set	Test Set	General Test Set	Dermatology Test Set
True positives	7,227 (82.6%)	1,298 (76.3%)	4,213 (84.4%)	6,947 (76.8%)
False positives	96 (1.1%)	34 (2.0%)	509 (10.2%)	964 (10.7%)
True negatives	1,306 (14.9%)	257 (15.1%)	245 (4.9%)	1,029 (11.4%)
False negatives	116 (1.3%)	112 (6.6%)	27 (0.5%)	110 (1.2%)
TOTALS	8,745	1,701	4,994	9,050

mar errors, and proper names, whereas we had each "match" judged by a practicing internist to make the determination of true positive or false positive, regardless of term classification.

By applying automated term composition with filters based on the UMLS semantic types, we showed that we could balance sensitivity and specificity to optimize the other metrics (Table 3).

Our study publication predates that of Nadkarni et al. by almost 14 months. It is clear from the data of both studies that concept indexing with the UMLS is actually highly sensitive, with quite a high positive predictive value. This conclusion was omitted by Nadkarni et al. but is worthy of further analysis as the UMLS, and the algorithms that use it, increase their specificity to match the already quite excellent sensitivity.

Surely, for a vocabulary to be useful it must evolve and its content must grow. The UMLS now contains more than 700,000 concepts, but it still does not cover all clinically useful terminology. As Cimino[3] states, "…a formal methodology is needed for expanding content." Chute et al.[4] reinforce this statement with the argument that "in the absence of a single, all-embracing health care terminology, there need to be coordination and organizing support for interrelated terminologies…" and that "developers of clinical classifications must consider ways they can develop their systems to become part of an integrated set of terminology systems."

If terms are added to a vocabulary indiscriminately, however, redundancy and combinatorial explosion may make the vocabulary unwieldy and difficult to search in a timely fashion. "An alternative approach is to enumerate all the atoms of a terminology and allow users to combine them into necessary coded terms, allowing compositional extensibility."[3,5,6] One risk of this approach is its potential for making the use of the vocabulary more complex.

We hypothesized that automated term composition as developed and tested in a randomized controlled trial[7] would allow large-scale coverage of specialty-specific and general local vocabularies. This automated process would facilitate the appropriate inclusion of such terms into a larger vocabulary without creating redundancy.

As we noted in 1999, user-directed composition may allow salvage of many of the false-positive and true-negative matches, thus significantly increasing the incorporation rate.[8] The true-negative terms, which do not yield to user-directed post-coordination of concepts to form a positive exact match, could form a set of terms that could be considered for incorporation into larger vocabularies without the onus of redundancy.

Table 2 ■

Comparison of Metrics

	Nadkarni et al.[1]		McDonald et al.[2]	
	Training Set	Test Set	General Test Set	Dermatology Test Set
Sensitivity (%)	98.4	92.1	99.4	98.4
Specificity (%)	93.2	88.3	32.5	51.6
Positive predictive value (%)	98.7	97.4	89.2	87.8
Positive likelihood ratio (%)	14.4	7.8	1.5	2.0

Table 3 ■

Metrics Derived by Use of Semantic Type Filtering

	General Test Set	Dermatology Test Set
Sensitivity (%)	88.1	75.2
Specificity (%)	73.7	82.2
Positive predictive value (%)	98.4	95.1
Positive likelihood ratio	3.34	4.22

Given the large size of both the specialty-specific and local general terminological corpi used in our study, this method should be generalizable to other local specialty-specific and general terminology sets. These results help solidify the need for compositional mechanisms for terminological representation and show the utility of the considerable synonymy offered by the UMLS. Future research should focus on how to integrate colloquial terminologies such as the UMLS with formal reference terminologies.—FURMAN S. MCDONALD, MD, PETER L. ELKIN, MD

References ■

1. Nadkarni P, Chen R, Brandt C. UMLS concept indexing for production databases: a feasibility study. J Am Med Inform Assoc. 2001;8:80–91.
2. McDonald FS, Chute CG, Ogren PV, Wahner-Roedler D, Elkin PL. A large-scale evaluation of terminology integration characteristics. Proc AMIA Annu Symp. 1999:864–7.

Affiliation of the authors: Mayo Clinic, Rochester, Minnesota.

Correspondence and reprints: Peter L. Elkin, MD, Associate Professor of Medicine and Medical Informatics, Mayo Medical School, Mayo Clinic, Rochester, MN 55905; e-mail: <elkin.peter@mayo.edu>.

Received for publication: 4/20/01; accepted for publication: 4/20/01.

3. Cimino JJ. Desiderata for controlled medical vocabularies in the 21st century. Methods Inf Med. 1998;37:394–403.
4. Chute CG, Cohn SP, Campbell JR. A framework for comprehensive health terminology systems in the United States: development guidelines, criteria for selection, and public policy implications. J Am Med Inform Assoc. 1998;5(6):503–10.
5. Cote RA, Robboy S. Progress in medical information management: Systematized Nomenclature of Medicine (SNOMED). JAMA. 1980;243:756–62.
6. Evans DA, Rothwell DJ, Monarch IA, Lefferts RG, Cote RA. Towards representations for medical concepts. Med Decis Making. 1991;11:S102–8.
7. Elkin PL, Bailey KR, Chute CG. A randomized controlled trial of automated term composition. Proc AMIA Annu Symp. 1998:765–9.
8. Elkin PL, Mohr DN, Tuttle MS, et al. Standardized problem list generation, utilizing the Mayo canonical vocabulary embedded within the Unified Medical Language System. Proc AMIA Annu Fall Symp. 1997:500–4.

Dr. Nadkarni replies:—We thank Drs. McDonald and Elkin for pointing out the relevance of their 1999 paper[1] to ours.[2] (This excellent paper, which we had not read earlier, provided a valuable education for us.) However, we would politely demur to the description of their study (in the second paragraph of their letter) as "almost identical" to ours. The very real differences in our respective study designs and objectives, as we describe below, led us to significantly different conclusions about the utility of the 1999 UMLS for our respective purposes.

The study of McDonald et al. tried to match the contents of two hand-curated vocabularies with the UMLS, using an automated term composition approach. One of these vocabularies was a lexicon developed for dermatology, while the other comprised the 5,000 most common terms extracted from the Mayo clinical notes system. Curation involved correction of obvious spelling errors and removal of duplicates. Their objective was to quantify the value of algorithms for the electronic interrelation of different terminologies, a very valuable objective articulated by Chute et al.[3]

Our study, on the other hand, tried to match phrases in documents (not necessarily equivalent to terms) to UMLS concepts through an entirely electronic process. We took phrases as they existed in the documents, so malformed phrases or misspellings contributed to failure of concept recognition. Our objective was to determine the potential utility of electronic concept indexing in assisting document retrieval. We therefore had to classify and quantify the types of errors that occur when matching is attempted, in finer detail than merely as "false positives" or "false negatives." Thus, we used categories such as algorithm failure (phrase too long), proper names confused with concepts, acronyms, and so on.

The first table in the letter, which quantifies TP, FP, TN and FN, is interesting. However, in our work, we had not attempted to quantify "true negatives," so we had to spend some time trying to determine how McDonald and Elkin obtained the number of 1,306 TNs from our published work. After some guesswork, we finally figured out that this number represented the sum of the following match categories:

Redundant UMLS concepts	490
Homonyms	481
Concepts not in UMLS	158
General form of concept missing from UMLS	127
Concept not useful for indexing	25
Too many non-stop words (algorithm failure)	25
TOTAL	1,306

From the perspective of McDonald and Elkin, which is to facilitate matching of terms between vocabularies using automated term composition, some of these categories (such as "Concepts not in UMLS") could indeed be classified, for their purpose, as true negatives. However, a failure of our algorithm, which handles a maximum of five non-stop words per phrase, could hardly be classified so charitably. The categorization of homonyms as true negatives is also unduly charitable. A 1994 paper by Rindflesch and Aronson[4] shows that it is possible to disambiguate different meanings of a homonym like "immunology," which can refer to a laboratory test panel or to the study of a biological function, but that this requires natural language processing of the document using numerous special-purpose rules that are specific for individual homonyms.

From our perspective of trying to match phrases as they are encountered in text to UMLS concepts, however, all the above categories must be classified as failures, whether they are true negatives or not. This is because they cause the matching algorithm (more precisely, our present algorithm; better ones will, no doubt, be developed by others) to strike out. Thus, if an isolated homonymous phrase is encountered in a document and cannot be disambiguated, it cannot be indexed. The value of a concept index, whose whole purpose is to make it possible to retrieve a document in response to a query, is therefore reduced. (In such a case, the user must fall back on the word index, if it exists, but then numerous false positives are retrieved.) This is our reason for taking a somewhat more pes-

simistic view than McDonald and Elkin, as reflected in our conclusions. We reiterate that our guarded view applies to our objective, not to that of McDonald et al.

The similarities of our work to that of McDonald et al., as correctly pointed out in the letter, is in the matching algorithm, which tries exact matches first and automated term composition later, although McDonald et al. also introduce the innovative step of using UMLS semantic types to assist the matching process. The authors' point that concepts cannot be proliferated endlessly by combination of existing concepts (because of combinatorial explosion) is well made; we fully agree. Eventually, however, the decision to create a compound concept is often a pragmatic one. (Thus, "blood pressure" began as a compound concept, pressure of the blood, but clinicians now consider this concept atomic: high blood pressure, another compound concept, is itself such an important and common condition that even lay persons are supposed to know about it.)—PRAKASH NADKARNI, MD

References ■

1. McDonald FS, Chute CG, Ogren PV, Wahner-Roedler D, Elkin PL. A large-scale evaluation of terminology integration char
2. Nadkarni P, Chen R, Brandt C. UMLS concept indexing for production databases: a feasibility study. J Am Med Inform Assoc. 2001;8:80–91.
3. Chute CG, Cohn SP, Campbell JR. A framework for comprehensive health terminology systems in the United States: development guidelines, criteria for selection, and public policy implications. J Am Med Inform Assoc. 1998;5(6):503–10.
4. Rindflesch TC, Aronson AR. Ambiguity resolution while mapping free text to the UMLS Metathesaurus. Proc Annu Symp Comput Appl Med Care. 1994:240–4.

Affiliation of the author: Yale University School of Medicine, New Haven, Connecticut.

Correspondence and reprints: Prakash M. Nadkarni, MD, Center for Medical Informatics, Yale University School of Medicine, P.O. Box 208009, New Haven, CT 06520-8009; e-mail: <prakash.nadkarni@yale.edu>.

Received for publication: 5/2/01; accepted for publication: 5/2/01.

■ **J Am Med Inform Assoc.** 2001;8:512–515.

Medical Informatics and Health Care in Oman

To the Editor:—I would like to introduce the readers of JAMIA to informatics contributions in the Sultanate of Oman health care system. The Ministry of Health provides health care services in a three-tiered system—primary health care (health centers, extended health centers), secondary health care (regional referral hospitals), and tertiary health care (tertiary referral teaching hospitals and a university hospital).

At present, data collection in the primary care tier is performed primarily by a health officer, who does a statistical compilation of the data generated in the clinics. There is no a standard data collection format; the format varies from clinic to clinic. For example, in a general outpatient department clinic, for example, the records include the various diseases seen and the number of patients examined by the general duty doctors. In a diabetes clinic, nurses complete paper reports recording, among other data, the number of patients attending the clinic, their age groups, their medications, whether cases are old or new, and patients' end-organ involvement. These data are then sent on a floppy disk to the regional statistical officer.

At a secondary health care facility (regional hospital), a medical record system collects information for better planning of health services. The source of data is the patient in a direct admission or, if the patient has been referred to the hospital, the referral letter. The medical records department collects all the data. The data entry is done using custom-designed software, which has a user-friendly data entry screen. In addition to the standard information—e.g., demographic information and admission details—the system collects information about the patient's tribe, the name of the sheikh, and whether the patient is local or expatriate (since expatriate patients must pay unless they are employed by the government).

The components of the hospital record are:

- *Operational information,* including basic details about patients admitted to the hospital; admissions in different departments, such as surgery, medicine, and pediatrics; and the ratio of referrals from various hospitals
- *Tactical information,* including the number of available beds in various departments; the number of patients in various wards, and durations of stay; the number of local and expatriate admissions in various wards; and referrals and the type of referrals

- *Strategic information,* which helps in the planning of future health facilities. For example, football/soccer is very popular, and if many referrals for sports injuries are received from a basic health facility, an orthopedic doctor might be stationed at the facility or doctors there might be trained in sports medicine.

The aggregated details from the medical records help create a morbidity profile, and this statistical result helps in planning future hospitals and deciding the types of facilities (e.g., oncology) to be included in a new hospital. The collected data have relevance to the health care of this country, since they provide a basic patient record database that can be used to develop the system further and a central registry of patients, especially patients with chronic diseases like diabetes and hypertension.

At the tertiary health care facility (referral tertiary teaching hospital and the university hospital), a well-developed integrated health care system consists of data on patient management, medical records, pharmacy, clinical physiology, laboratory management, nurse scheduling, and finance and administration. The information can be accessed by doctors and nurses from desktop terminals in wards as well as clinics.

Scope for Improvement

Even the best systems may be improved. Areas of the existing system in which changes are needed include some patient details; lists of completed procedures, e.g., laboratory investigations, surgical procedures, and radiographic or other special examinations; and allergies and special conditions, such as glucose-6-phosphate dehydrogenase deficiency or sickle cell anemia.

Suggestions for further development in the secondary health care area include:

- The system should be accessible to health workers, so that a medical records person is not always needed to handle the system.
- Minimal training of staff in referral clinics should be provided, so that appointments can be generated in the clinics themselves.

Planning for the Future

The awareness of informatics among health care providers at the primary health care level is nil. At the regional level, health care providers do not access the hospital system, since only the medical records system is in place. Current plans are to implement a more comprehensive laboratory system along with investigative databases, and to integrate these with the total system. Doctors and nurses are being given a preliminary introduction to use of the system.

At the tertiary level, doctors and nurses learn to use the system on the job.

At present, there are no plans for implementation of medical informatics at the primary health care level, and there is no specific curriculum for medical informatics in the university medical college. There is also no one specifically trained in medical informatics in Oman. We do, however, have a two-pronged strategy aimed at 1) an integrated health care management system and 2) medical informatics education. These proposals are applicable to our particular system of health service.

Integrated Health Care Management System

We are planning an integrated health management system, beginning at the primary health care level. An integrated health care system at this level should cater to the needs of the institution, which may be an health center or an extended health center (polyclinic).

- Stage One—Computerized medical records system, laboratory, and radiology records and automated reporting system that can be accessed by a health care provider from anywhere in the health care facility
- Stage Two—Networking of this system with the secondary health care facility (regional referral hospital), so that any patient admitted, treated, and discharged from a regional hospital can be followed up at a primary health care facility, with the primary care physician having online access to the detailed patient record from the regional hospital

The secondary health care facility (regional hospital) should have a full-fledged integrated hospital system (IHS) that includes the patient record and records of investigative (e.g., laboratory and radiology) and surgical procedures that have been performed. Digital cataloguing of x-rays and other images would be desirable.

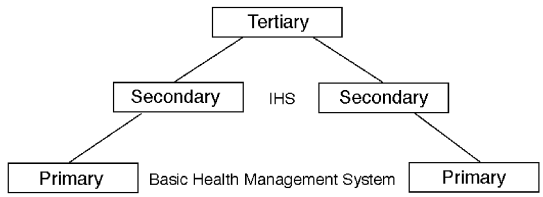

Figure 1 Strategic vision of interconnecting information systems in support of Oman's health care system.

The tertiary health care facility already has a system, which needs to be upgraded and networked with the primary and secondary health care facilities (Figure 1). A master network would be useful, given the geography of this country and its large area. For example, a patient's initial treatment and follow-up for hypertension or diabetes may occur at health care facilities in different regions. With central networking and use of the patient's ID, the health care provider could access the patient's records, including details of medications, previous hospitalizations, and the course of the disease up to now. This information would help the health care provider provide efficient, cost-effective, and precise patient care.

Medical Informatics Education and Training

Training Students. We plan to introduce a medical informatics course as part of the curriculum of the medical school at Sultan Qabus University. The university can create the course or collaborate with institutions already running such a course, such as the Royal College of Surgeons in Edinburgh, Scotland.

Training Health Care Providers. Once a core of trained medical informatics people is in place, a phased program for training other health care providers could begin. Health providers at the primary health care level should be trained first. They can be trained first in basic computer skills and eventually in medical informatics. This core group of health care providers trained in medical informatics can become master trainers of other providers at the various health facilities.

Medical informatics workshops, seminars, discussions, and presentations can be used to introduce people to this field. At these presentations, health care providers can test-drive a system to discover its uses and benefits. Having health care providers as master trainers and trainees means that everyone speaks "the same language," and trainees will not have to deal with too much computer jargon.

Innovations in the education system combined with an integrated health system will allow the people of Oman to experience the best possible health care.—ASHWANI DHAR, MBBS

Affiliation of the author: Ministry of Health Hospital, Oman.

Correspondence and reprints: Dr. Ashwani Dhar, MBBS, P.O. Box 86, Postal Code 315, Oman; e-mail: <ashdhar@omantel.net.om>.

Received for publication: 5/3/01; accepted for publication: 5/21/01.