Enhancements of Anatomical Information in UMLS Knowledge Sources

Cornelius Rosse MD, DSc[1], Mohamed Ben Said, MD, MS[2],
Kraig R. Eno, BSc[1], and James F. Brinkley, MD PhD[1]
[1]Digital Anatomist Program, Department of Biological Structure,
University of Washington, Seattle,WA 98195
[2] Service d'Information Medicale (SIM), Hopital Tenon Paris 75020 - France

ABSTRACT

Although anatomical terminology forms a part of biomedical structured vocabularies, available sources lack the requisite granularity, semantic types and relationships for comprehensively and consistently representing anatomical concepts in machine readable form. Thoracic angiology was selected as a proof of concept experiment for in depth representation of symbolic information in gross anatomy through the enhancement of semantic types, concepts and relationships in UMLS. Provided the representation of concepts is comprehensive, hierarchies generated with four types of simple relationships are capable of displaying anatomical information from the systemic view point with sufficient detail to meet the needs of applications in basic science education and in the practice of surgical subspecialties.

INTRODUCTION

The development of structured medical vocabularies is motivated largely by the advantages that can be realized through computerizing the clinical record, because it is recognized that standardized and automated representations of clinical information can translate into efficiency and cost saving in clinical care [1]. Several vocabularies have, therefore, been constructed from a clinical view point [2,3,4] and their evaluations also tend to focus on assessing the occurrence of the terms they contain or lack when compared with different types of clinical records (discharge summaries, radiological case reports, etc.).

Such a clinical focus, however, imposes limitations on these information sources, in that many of the fundamental concepts that are implied in the use of clinical terminology are not entered into these systems, because the terms denoting them do not appear explicitly in clinical records. As a consequence, the representation of basic science concepts tends to be too shallow for these information sources to meet the needs of knowledge representation in biomedical domains that are not directly concerned with clinical care (for example, see Table 1). These shortcomings are surfacing as a result of the establishment of on-line databases that have a potential for significantly enhancing and facilitating the applications of basic science information in the training of health care providers and in the delivery of health care. The development of knowledge-based end-user programs for training and for delivering patient care calls for the comprehensive and specific representation of concepts in the biomedical sciences that are basic to clinical medicine as well as in clinical medicine itself.

The limitations of UMLS [5] and SNOMED [3] for a basic science application became apparent in our attempts to establish a symbolic knowledge base in gross and neuroanatomy for the Digital Anatomist information framework [6,7]. We were motivated to take advantage of the schema of the UMLS Metathesaurus because it assures linkages, via the UMLS semantic network, to other biomedical information sources. More importantly, the Metathesaurus provides a stable environment for incorporating new concepts and developing new relationships [8]. However, both UMLS and SNOMED lack the appropriate granularity and semantic relationships for a satisfactory and comprehensive representation of gross and neuroanatomy.

This report considers the requirements of a symbolic knowledge base in anatomy, assesses the anatomical content of available knowledge sources, and reports on the representation of thoracic vasculature as a prototype for the elementary modeling of anatomical information.

REQUIREMENTS

The minimum requirements of a symbolic knowledge base for each anatomic region and body system should include 1. semantic types of anatomic concepts, 2. a comprehensive set of concepts with a high degree of specificity, 3. a set of terms that are associated with each concept reflecting usage in different fields, 4. definitions of concepts and their semantic types, and 5. the relationships that link concepts and semantic types into different hierarchies.

The Digital Anatomist information sources are concerned with macroscopic anatomical entities and may be regarded as a prototype for a knowledge base in anatomy [6,7]. The spatial database for a particular anatomical region contains image files for each segmented structure that has dimensions of approximately 1mm or larger. In order to develop a knowledge-based interactive module for applications such as simulated dissection or surgery, each anatomic

0195–4210/95/$5.00 © 1995 AMIA, Inc.

entity must be linked to specific identifiers that distinguish not only different structures, but similar structures that recur on the two sides of the body, as well as those that recur segmentally (e.g. each of the eleven posterior intercostal arteries on both the right and left sides, and also the branches of each of these arteries). The representation of semantic types and different relationships should permit interaction with the spatial data in different contexts (e.g. from a systemic view point: "Display the arteries that supply the esophagus"; or from a regional view point: "Display the structures anterior to the left main bronchus in the pulmonary hilum").

METHOD OF APPROACH

We selected the thorax as a prototype region for developing and evaluating elementary components of a symbolic knowledge base in anatomy, and began by restricting our attention to thoracic vasculature (angiology). First we have proposed candidate anatomical semantic types for thoracic vasculature as children of the semantic types currently represented in UMLS, generating an -IS A- hierarchy (Fig.1). Second, we have generated a comprehensive set of anatomical concepts from an American [9] and a European [10] textbook of anatomy, and a surgical anatomy text [11]. We have assigned each concept to a semantic type and tested the validity of the candidate semantic types by evaluating whether they appropriately and usefully capture generalizable characteristics of the concept. Conforming to the syntax of the UMLS Metathesaurus, we assigned a preferred term to each concept and designated other terms that are in use to denote that concept as synonyms. The set of preferred terms was then mapped to the UMLS Metathesaurus [12] and SNOMED [3], and Nomina Anatomica [13] was checked in order to determine which concepts were or were not represented in these vocabularies.

A unique University of Washington/Digital Anatomist (UWDA) identifier was assigned to each concept and to all the terms associated with it. The UWDA identifier was cross-referenced with the equivalent UMLS identifier and SNOMED code when the concept was represented in these sources.

Hierarchies were generated using the UMLS relationships -IS A- and -PART OF-, and two new, UWDA relationships, -BRANCH OF-, in the case of arteries and -TRIBUTARY OF-, in the case of veins; both relationships are children of the UMLS relationship -PHYSICALLY RELATED TO-.

We entered the information into a relational database by creating two tables, one containing terminology, and the other links between concepts. The TERMS and LINKS tables were stored using Sybase, a commercial database management system running on a NeXT computer. Information in the two tables can be related using the UWDA concept identifier.

A row in the TERMS table stores a term, its role (semantic type, preferred term, or synonym) and its UWDA, UMLS, and SNOMED identifiers. Preferred terms and synonyms for a concept are assigned the same UWDA identifier.

The LINKS table stores parent-child relationships between terms by recording the identifiers of the two nodes, along with a code for the link type. A series of locally-written Unix utilities are used to extract subsets of the data in formats for easy viewing and printing, as shown in Tables 1 and 2.

We are currently working on an interactive editor which will allow domain experts to view and change the information in the database directly, using a graphical user interface.

RESULTS

Anatomical Semantic Types

Anatomical semantic types are groups or classes of anatomical entities that are distinguishable from one another on the basis of a set of attributes (*differentia*) that are shared by all instances in one group. In the field of gross anatomy the most specific UMLS semantic types are "Body Part, Organ or Organ Component", "Body Space or Junction" and "Body Location or Region". Figure 1 shows anatomical semantic types linked by the -IS A- relationship at successive nodal levels for "Blood Vessel" as one of the children of "Body Part, Organ or Organ Component", before reaching leaf concepts. A similar hierarchy was generated for angiology as offspring of "Body Space or Junction" yielding anatomical semantic types at three nodal levels such as "Orifice or Ostium", "Vascular Anastomosis" "Arteriovenous Anastomosis", etc. In making concept assignments, we have found it necessary to incorporate anatomical variants as semantic types distinct from "Congenital Abnormality", a UMLS semantic type. Textual definitions were generated for anatomical semantic types as an initial step toward machine-readable definitions once an appropriate language for concept representation has been identified [14, 15].

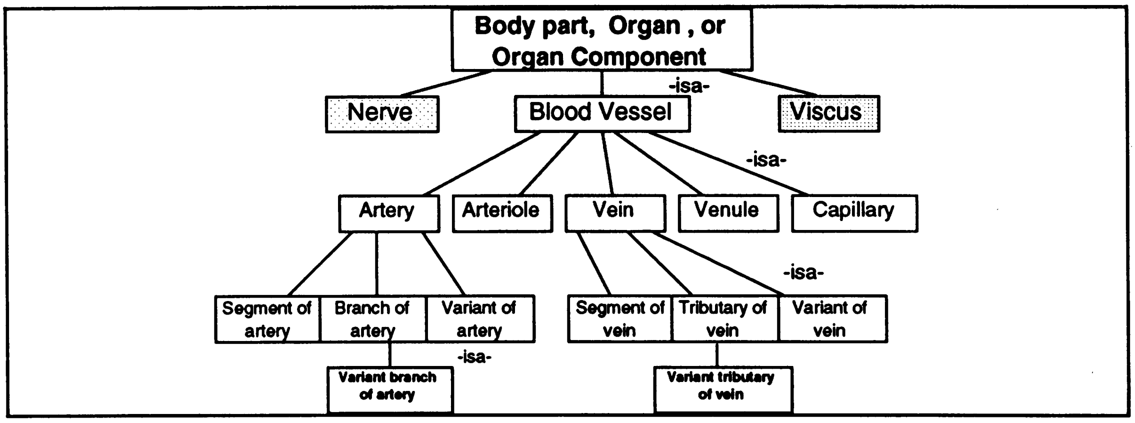

Fig. 1 Anatomical semantic types for angiology represented as offspring of a UMLS semantic type.

Concepts

Our objective was to represent every macroscopically identifiable blood vessel in the thorax as a discrete anatomical concept and assign a unique identifier to it. This specificity takes into account laterality and duplication of segmentally recurring anatomical structures.

Such comprehensiveness and granularity distinguish the UWDA vocabulary from other information sources. For instance, Nomina Anatomica designates *Arteriae intercostales posteriores* and SNOMED includes *Posterior intercostal arteries*, both without further specification. UWDA adopts these concepts and terms and extends them: Each of the eleven posterior intercostal arteries on both the right and left sides of the body are entered as discrete concepts. Moreover, the same is true of each of the branches of each of these arteries (see Table 2).

Terms

Preferred terms for concepts were assigned from an anatomical, rather than a clinical, view point. For instance the UWDA preferred term for a terminal branch of the left coronary artery is "Anterior interventricular branch of the left coronary artery", which is descriptively correct and is the English language equivalent of the Nomina Anatomica term. The UWDA term list also includes "Left anterior descending artery" and the acronym "LAD", which are commonly used clinically, as synonyms, assigning them the same UWDA concept identifier as to the preferred term. When a Nomina Anatomica term exists for a concept, it is always included as a synonym. If a concept is named in SNOMED, the term is included either as a preferred term or a synonym.

Terms designating parts and generations of branches of the aorta within the thorax account for 958 instances in the term list (737 preferred terms and 221 synonyms). Of those terms, 34 have UMLS identifiers, 152 have SNOMED identifiers, (several have both), 60 are present in Nomina Anatomica and 672 appear only in the UWDA term list.

Relationships

The relationships that can be represented among angiology concepts within the thorax using four links capture a large amount of detailed anatomical information from a systemic view point. All semantic type assignments are specified in machine-readable files by -IS A- links to anatomical semantic types, which themselves are arranged hierarchically using the same mechanism. The -IS A- hierarchy encodes the semantic types and links all individual concepts of thoracic angiology to the two broad gross anatomic semantic types currently represented in UMLS. The -PART OF- hierarchy captures detailed topographic information about thoracic vasculature,including such clinically important information as normal anatomical variants (Table 1).

The -BRANCH OF- relationship encodes a hierarchy for all arteries and their macroscopically identifiable branches within the thorax, and displays their derivation from the aorta or the pulmonary trunk (Table 2). The -TRIBUTARY OF- relationship establishes corresponding hierarchies for the veins in the thorax that terminate in the superior vena cava or in the pulmonary veins. Tables 1 and 2 provide examples of the extent and granularity of anatomical information that can be retrieved from the database with the aid of quite simple relationships because there is in depth and comprehensive representation of concepts.

Concept	Semantic Type	UMLS	SNOMED
3734 Aorta	Artery	3483	37633095
3736 Ascending aorta	Segment of artery	3956	37633103
3738 Supravalvular ridge	Segment of artery	--	--
3740 Bulb of aorta	Segment of artery	--	37633105
3745 Aortic sinus	Segment of artery	37197	37633107
3748 Anterior aortic sinus	Segment of artery	--	--
3752 Ostium of right coronary artery	Orifice or Ostium	--	37633145
3754 Ostium of third coronary artery (right)	Variant of Orifice	--	--
3755 Absent ostium of right coronary artery	Variant of Orifice	--	--
3756 Left posterior aortic sinus	Segment of artery	--	37633109
3760 Ostium of left coronary artery	Orifice or Ostium	--	37633130
3762 Ostium of third coronary artery (left)	Variant of Orifice	--	--
3763 Absent ostium of left coronary artery	Variant of Orifice	--	--
3764 Right posterior aortic sinus	Segment of artery	--	37633110

Table 1. An example of the tab delimited listing of the -PART OF- hierarchy for a segment of the aorta. Preferred terms of concepts are preceded by the UWDA identifiers and are followed by the anatomical semantic type, UMLS and SNOMED identifiers.

2772 Second part of right subclavian artery							
	3901 Right costocervical trunk						
		2933 Right superior intercostal artery					
			2124 First right posterior intercostal artery				
				.			
				.			
				.			
				2263 Second right posterior intercostal artery			
					3370 Precentral branches of second right posterior intercostal artery		
					2401 Dorsal branch of second right posterior intercostal artery		
						3862 Spinal branch of dorsal branch of second right posterior interc	
							2729 Postcentral branch of spinal branch of seco
							2296 Prelaminar branch of spinal branch of seco
							3404 Anterior radicular branch of spinal branch
							3635 Posterior radicular branch of spinal branch
							2669 Segmental medullary branch of spinal bran
						3874 Medial branch of dorsal branch of second right posterior inter	
							3909 Postlaminar branch of dorsal branch of sec
							3385 Articular branches of dorsal branch of seco
							3869 Muscular branches of dorsal branch of med
							2637 Cutaneous branches of dorsal branch of me
						2308 Lateral branch of dorsal branch of second right posterior inter	
							2705 Muscular branches of lateral branch of seco
							2215 Cutaneous branches of lateral branch of sec
					2593 Collateral branch of second right posterior intercostal artery		
						3707 Muscular branches of collateral branch of second right posteri	
					2634 Lateral cutaneous branch of second right posterior intercostal artery		
						3006 Anterior branch of lateral cutaneous branch of second right po	
							2662 Mammary branches of anterior branch of se
						2750 Posterior branch of lateral cutaneous branch of second right po	
					2557 First right bronchial artery		
	3007 Right deep cervical arery						

Table 2. An example of a tab delimited listing of the -BRANCH OF- hierarchy showing branches and subbranches of the second part of the right subclavian artery. UWDA concept identifiers precede each preferred term. The text component of most entries does not fit the printed page format of the table, but it is not relevant to illustrating the organization of the hierarchy.

DISCUSSION

Although the systematization of anatomical terminology has a long history, machine-readable representations of anatomical information remain inadequate for building symbolic models that meet the needs of knowledge-based educational and clinical applications. The work we report is the first step toward providing elementary modules of a symbolic knowledge base in anatomy as one of the components of the Digital Anatomist client-server framework for structural information [6,7]. The first practical application of the evolving structured vocabulary is the Digital Anatomist Interactive Atlas [7,16,17]. This user interface, designed for educational applications of the anatomical information source, can be readily modified to serve the needs of clinical specialties and subspecialties such as radiology,cardiology, and general and thoracic surgery. Moreover, since most of the biomedical information used in clinical practice, education and scholarly communication is explicitly or implicitly linked to some anatomical concept, by providing comprehensive and consistent representation of anatomical concepts, our work should facilitate the development of a variety of computerized information systems used in clinical and academic settings (databases, expert systems, financial programs, educational programs, etc). The most effective way to assure reusage of the anatomical knowledge representations we develop is to incorporate them into UMLS.

ACKNOWLEDGEMENTS

This work was supported by NLM contract LM4-3546. We gratefully acknowledge the encouragement and advice we have received from Betsy Humphreys, Alexa McCray and Mark Tuttle and we thank Gordon L. Briggs for making SNOMED International available for our work.

References

[1] United States General Accounting Office. Standards for Automated Medical Records. Report of Congress GAO/IMTEC-93-17, 1993.

[2] World Health Organization.International Classification of Diseases: Ninth Revision. Geneva: World Health Organization, 1977.

[3] Côté RA, Rothwell DJ, Palotay LJ, Beckett RS, Brochu L, eds. The Systematized Nomenclature of Medicine: SNOMED International. Northfield,IL: College of American Pathologists. 1993.

[4] Chisholm J The Read clinical classification. Br. Med. J 1990; 300:1092.

[5] Lindberg DAB, Humphreys BL, McCray AT The Unified Medical Language System Meth Inform Med 1993; 32:281-291.

[6] Brinkley JF, Prothero JS, Prothero JW, Rosse C A framework for the design of knowledge-based systems in structural biology. Proc. 13th Annual Symp. on Computer Applications in Medical Care. Washington, D.C., 61-65, 1989.

[7] Brinkley JF, Eno KR, Sundsten JW, Conley DM, Rosse C. 1995 A distributed framework for distance learning in anatomy: The Digital Anatomist Interactive Atlas. 19th SCAMC, New Orleans, LA.

[8] Tuttle MS, Olson NE, Campbell K, Sheretz DD, Stuart JN, Cole WG Formal properties of the Metathesaurus 18th SCAMC 145-149, 1994.

[9] Hollinshead WH, Rosse C Textbook of Anatomy 4th edition, Harper & Row, Publishers, Philadelphia, PA, 1985.

[10] Williams PL, Warwick R, Dyson M, Bannister LH Gray's Anatomy 37th ed, Churchill Livingstone, New York, NY, 1989.

[11] Hollinshead WH Anatomy for Surgeons, Vol.2., 2nd edition, Harper & Row, New York.

[12] Tuttle MS, Blois MS, Erlbaum MS, Nelson SJ, Sherertz DD Toward a biomedical thesaurus: building the foundation of the UMLS. In: Greenes, R.A. (ed). Proc 12th Ann Symposium on Comp Appl in Med Care. New York: IEEE Computer Society Press 191-195, 1988.

[13] Nomina Anatomica Prepared by Subcommittees of the International Anatomical Nomenclature Committee. 3rd edition Churchill Livingstone,Edinburgh, 1989.

[14] Evans D, Cimino JJ, Hersh WR, Huff SM, Bell DS Toward a medical-concept representation language J Am Med Informatics Assoc 1994; 1:207-217.

[15] Campbell KE, Das AK, Musen MA Logical foundation for representation of clinical data J Am Med Informatics Assoc 1994' 1:218-232.

[16] Bradley SW, Rosse C, Brinkley JF 1995 Web-based Access to an online atlas of anatomy: The Digital Anatomist common gateway interface. 19th SCAMC, New Orleans, LA.

[17] Conley DM, Sundsten JW, Ratiu P, RauschningW, Rosse C 1995 The Digital Anatomist Series: 3-D, segmented, dynamic atlases of body regions. 19th SCAMC, New Orleans, LA.