PUTTING FLESH ON THE BONES: ISSUES THAT ARISE IN CREATING ANATOMICAL KNOWLEDGE BASES WITH RICH RELATIONAL STRUCTURES

Fritz Lehmann and Doug Foxvog,
fritz@cyc.com and foxvog@cyc.com
Cycorp, 3721 Executive Center Dr., Austin, Texas, 78731-1645 USA

Abstract

The Cyc system has a large knowledge base with detailed knowledge about its concepts. Cyc has axioms and rules for common-sense real-world knowledge embedded in a large taxonomy covering "everything". This is used with Cyc's logical inference engine to cover medical subject matter in conceptual depth with machine-usable semantics. Cyc's semantic definitions of relationships are more detailed than those in other systems. For gross anatomy we described the spatial positions, orientations, and functions of, as well as relationships between, bones, muscles, vessels, nerves, organs, etc. Cyc knows which junctions of vessels are distal to which others, which organs are directly served, etc. Several general Knowledge Representation and ontology issues arose while we described human gross anatomy to Cyc including: logical representation of "prototypical" relations, connections, and alignments; composite informative names versus opaque name-tokens; abstract versus concrete series; body regions and anatomical directions; non-exact mappings to external standard terms; directness of functional service by vessels; body systems versus locations in the body; specialized prototypes; and the veterinary organ taxonomy problem.

Introduction

Our system, Cyc, includes an automated logical inference system, a justification system, web interfaces, and a large knowledge base including not only terms but also detailed knowledge about each concept, including axioms and rules for common-sense real-world knowledge (Lenat 1995, Lenat et al. 1990). These are embedded in a large taxonomy covering "everything". Thus our knowledge structure is different from term hierarchies such as the UMLS Metathesaurus (UMLS 1998) or other "vocabulary servers". It is intended to cover the same subject matter and more, but in much greater conceptual depth with machine-usable information (in particular, richly axiomatized relational information and attributes). Our system is described in our World Wide Web page at **http://www.cyc.com**.

We are now working to cover various fields related to medicine and health care, such as anatomy, diseases, treatments, outcomes, etc., and we are establishing cross-link citations from our conceptual ontology to other systems' term hierarchies. For the latter, we have established citations pointing from Cyc's anatomy concepts to corresponding terms and codes in WordNet (Miller et al. 1993, 1998), MeSH (MeSH 1997), SENSUS, and SNOMED (Côté et al. 1993) We plan to add citations to the Read Codes (Read 1997). We plan to do the same for tissues and biochemistry. Later, when our ontology is rich enough to provide all the necessary base-concepts, we would like to describe (in logical semantic definitions) the meanings of the ICD-9/10 (ICD-9/CM 1997, ICD-10 1992), CPT (CPT 1996), and Read Codes (Read 1997) for disorders and procedures.

Our semantic definitions of relationships are more detailed than those we have seen in other systems. In gross anatomy, we describe the spatial positions, orientations, and functions of, as well as relationships between, bones, muscles, vessels, nerves, organs, etc. using the Cyc ontology. Cyc knows, from its more primitive ontology, that a blood vessel is a tube through which a fluid ordinarily flows, that its junctions are junctions of "pipes", and that "distal" means ordinal distance from the heart along blood vessel pathways. Cyc knows which junctions of vessels are distal to which others, which organs are directly served by which vessels, etc. Such facts are represented in the CycL logical language as "ground-level" axioms in a Human Body Structure "Microtheory" (a special context for relevant facts), for example:

```
In Mt : HumanBodyStructureMt
(distalTo
    (The
      (LeftFn GonadalVein))
    (The
      (LeftFn RenalVein))).
```

During about a year of work, some interesting general knowledge representation and "ontology" issues arose in what had seemed in advance to be a fairly routine "knowledge engineering" and entry task: describing human gross anatomy. These are described in the following section.

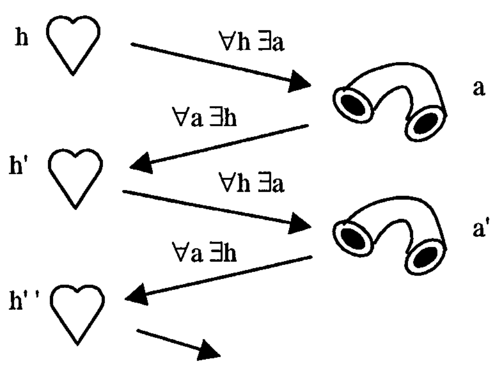
Semantic Issues

Issue 1. Logical Representation of Prototypical Relations

Straightforward logical representation of body part relations is too burdensome. In English, we say "the heart" is connected to "the aorta", but in a logical description of reality it is individual hearts that are connected to individual aortas. If we formalize the general rule in logic, using the notation (predicate argument1 argument2 argument3 ...) and (connective proposition1 proposition2 proposition3 ...), as:

```
ForAll h,v:
(implies                                    [=>]
 (and (instanceOf h Heart)
      (instanceOf v Vertebrate)
      (anatomicalParts v h))          [h is part of v]
 (thereExists a                          [such that]
  (and (instanceOf a Aorta)
       (anatomicalParts v a)
       (connectedTo a h))))
```

some of the meaning of the rule is expressed, but not enough. The above says that **for every individual heart of a vertebrate, something exists which is an aorta in that vertebrate and is connected to that heart**. Unfortunately that still allows the heart to be connected to several aortas, and fails to say that for every aorta there exists exactly one heart connected to it. Many re-formulated logical fixes of this problem lead to their own logical problems:



In the diagram above, a heart has an aorta, and an aorta has a heart, etc., but the quantified formulae fail to make sure that the heart of the aorta of the original heart is the *same heart.*

Moreover, even if it is done completely correctly, the resulting formula is too long and complicated for people to read, or to be used efficiently in machine inference.

Instead, we use a "prototype" individual, the "typical person" individual, and use a functional "The" to denote "The Heart", "The Liver", "The Aorta", of that person, etc. This works only for unique body parts: (The Heart). For paired parts we use (The (LeftFn Lung)) where "LeftFn" is a function returning the leftmost of two bilaterally symmetrical body parts. This will not work for multifarious body parts like alveoli, hairs, etc. Parts of a paired body part that themselves have no symmetric counterpart, such as the middle lobe of the right lung, are unique body parts and are referred to accordingly. (The left lung has no middle lobe.)

We put prototype-related statements in a special BodyStructure "microtheory" or context called "VertebrateBodyStructureMt" as follows:

```
In VertebrateBodyStructureMt:
(connectedTo (The Heart) (The Aorta)).
```

which is a much simpler statement: in the "prototypical vertebrate", "The Heart" and "The Aorta" are connected. (Actually the connection used is more specific than connectedTo --- see below.)

In our work on prototypes, we allow for exceptions. The reason that anatomical prototypes work at all is that people tend to be "cut from the same pattern." Down to a certain level of detail, they have most internal part properties and relations in common. But the Cyc system recognizes that in certain respects an individual may depart seriously from the "normal" anatomical pattern.

Prototypical body parts are mapped to an individual's particular body parts using BodyPartFn for unique body parts and BodyPartCollectionFn for multiferous parts. Cyc only has to be told:

```
(alignedAlongLength
   (The (LeftFn RadialArtery))
   (The (LeftFn Radius-Bone))),
```

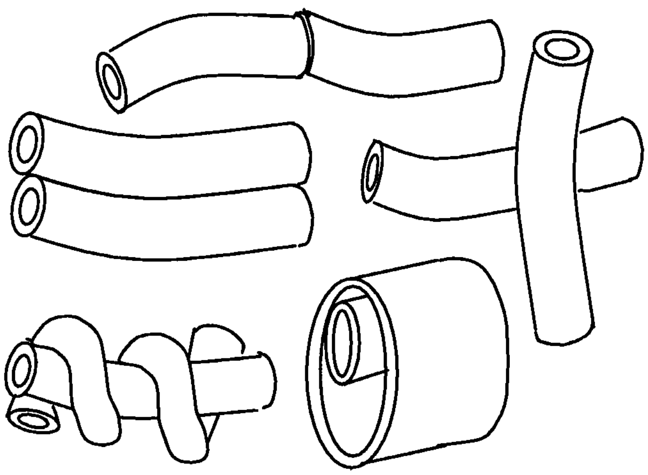
meaning "the left radial artery is aligned along the left radius", to know that Ann's left radial artery is aligned along her radius:

```
(alignedAlongLength
   (BodyPartFn Ann (LeftFn RadialArtery))
   (BodyPartFn Ann (LeftFn Radius-Bone))).
```

Issue 2. Connections and Alignments

In stating the connections and alignments between parts, we have specialized the concept for "physically connected" into dozens of more informative connection predicates. The connection predicates (relations) say *how* the two items are connected. We have supplied logical definitions (for these connection relations) that are used in logical machine inference. The connection relations form a *subsumption hierarchy of relations*, having connectedTo at the top (most general) with more specific connection relations below it. There are several "dimensions" to this specificity. The rigidity of the connection, the degrees of freedom, the forms of the related objects (like cavity, vessel, fiber), the relative spatial alignment (along, inserted into, wrapped around, etc.), and whether fluid can flow between the items, are among the many ways of specializing "connectedTo".

Two pipes may be aligned with each other in several possible spatial relations.



Similarly, a cylinder can be alongside a sheet or membrane in several different ways. In most cases the two items may be either connected or merely juxtaposed. There may be specific connection types such as synovial, grafted, sutured, or even stapled. We have predicates for the various juxtapositions, and for the various corresponding connection types in which the so-juxtaposed body parts are connected.

The following is a list of some of the anatomy-relevant connection types, excluding things like screwedIn needed for other domains (although some prostheses are screwed into bone) and mere juxtaposition types. More complicated connection types that include extra parameters are also excluded from this list. Each of these connection types has one or more rules detailing what the name is meant to express to the reader.

- connectedTo
  - connectedAtContact
    - adjacentPathsAtJunction
    - cavityConnectedAlongPathSide
    - connectedAlongEdge
      - connected-EdgeToEdge
        - connected-EdgeToEdge-Acute
      - connected-EdgeToSurface
        - connected-EdgeAlongCylinder
        - connected-SheetTransectsAlong
      - connected-EntireEdge
    - connectedAlongSurface
      - sheetSurfaceConnected
    - connectedAtEnd
      - connected-AbutsSurface
        - pipeAbuts-DeadEnd
      - endToEndConnected
      - pipeEndsAtCavity
    - connectedAtSpot
  - cavityConnectedAlongPathSide
  - connected-LengthAlongSurface
  - in-Embedded
    - embeddedCylinderInSheet
    - in-Lodged
    - muscleInsertion
    - muscleOrigin
  - connectedAlongInside
  - connectedAlongLength
    - connectedAlongAlignedLengths
  - connectedAround
  - stuckTo
- connectedViaConnector
- connectedTo-Rigidly
- rotationallyConnectedTo
- connectedTo-SemiRigidly
  - connectedViaFlexibleConnector
    - flapHingedTo
- connectedToInside
- directlyServesBodyParts
  - innervates
  - muscleInsertion
  - muscleOrigin

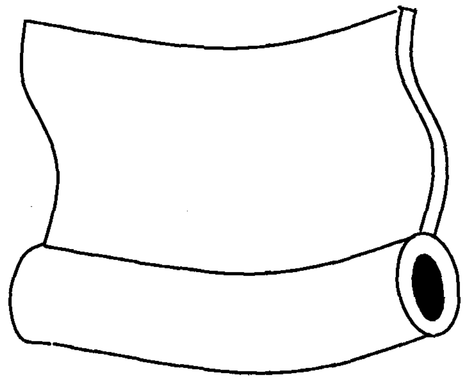
The indenting in the list shown above does not reflect the full hierarchy of connection types. Even after the irrelevant relations were edited out; duplicates in the indented list, due to being under multiple parent relations, were removed. This subsumption relation between relations is used directly for machine inference.

For example, the aorta – heart connection has actually been asserted as

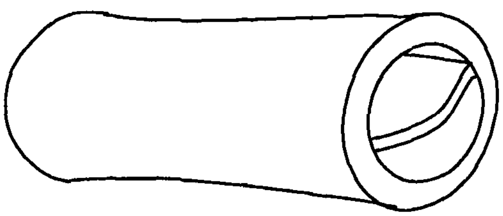
(pipeEndsAtCavity
(The AscendingAorta)
(The (LeftFn VentricleOfHeart))).

This implies the aforementioned (connectedTo (The Heart) (The Aorta)) because (The (LeftFn VentricleOfHeart)) is a part of (The Heart), (The AscendingAorta) is a part of (The Aorta), (The Heart) and (The Aorta) are disjoint, the predicate connectedTo subsumes the predicate pipeEndsAtCavity, and the predicate connectedTo is a symmetric predicate (i.e. (connectedTo X Y) implies (connectedTo Y X)). Cyc knows all this automatically.

To express the connection between the Mesentery sheet and the Small Intestine we write:
(connected-EdgeAlongCylinder
(The Mesentery)
(The SmallIntestine)):



To express the connection between the Tectorial Membrane and the Cochlear Tube in the Inner Ear we write:
(connected-SheetTransectsAlong
(The (LeftFn TectorialMembrane))
(The (LeftFn CochlearTube))):



Issue 3. Composite Informative Names Versus Opaque Names

We could have created all anatomy concepts with long single names, possibly helpful to human beings, but not to computers, like FourthDigitOfLeftHand or, equally good for the machine, Qx5897xypvtdl5a.

Instead, in many cases we followed medical practice in using composite, informative functional designators rather than simple names:

(Nth (The (LeftFn FingerSeries)) 4)

means the fourth digit of the left-hand finger series counting laterally from the thumb.

The composite description with nested functions allows the Cyc program to draw various fairly general inferences automatically. Cyc knows that the above example is part of the left hand, is between the third and fifth digits, that it is connected synovially to the fourth metacarpal bone by the fourth metacarpal joint, etc., without having explicit statements of all such facts

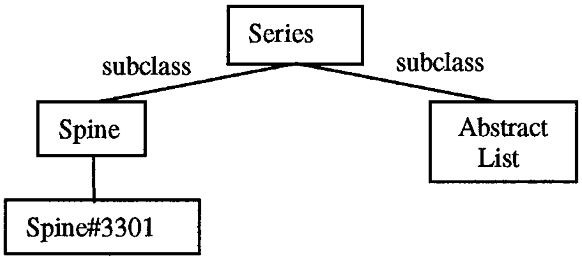
Issue 4. Abstract Versus Concrete Series

Representing parts of the body that are deemed to be in a series, or customarily are counted, forced us to face the deep issue of an abstract structure versus its real-world embodiments. The vertebrae, for example, are in several series and we customarily count from the cephalic end of such a series to designate a particular vertebra.[1] We have, in Cyc, axioms about sequences as abstract entities -- that there may be a first and a last, that no element has two immediate successors or two immediate predecessors, the notion of "betweenness", and so on. We'd like these features to apply to the real-world series, such as the vertebrae *in situ*.

We considered two ways to accomplish this:

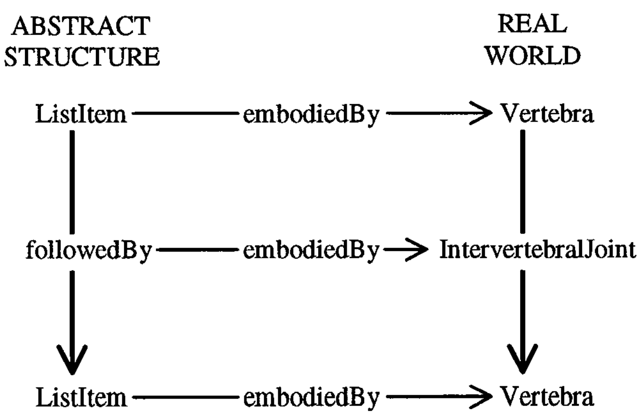
Approach 1. Treat real-world structures as specializations of general abstract structures. Assert that every real-world embodiment of an abstract structure is an instance of the general (including the abstract) collection of such structures. In this view, any triangular pretzel is a triangle. This makes every defined collection of real-world series (such as the collection of all series of vertebrae, or the collection of all rows of teeth, or the collection of all series of dorsal rami) a subset of the collection of all series. A drawback of this approach is that Cyc cannot efficiently conclude that two isomorphic (structurally identical) series are "the same", even in the very abstract domains (like mathematical Graph Theory) in which this is desirable.

[1] The ordering is purely conventional, and often depends on the culture. In English, the fingers are counted from the thumb to the little finger, whereas in Burushaski they are counted from the little finger to the thumb (the 5th finger).



Approach 2. Treat real-world structures and abstract structures as separate but corresponding systems. Assert that for every structured real-world thing, there is a corresponding abstract structure that it "embodies". In this view, any triangular pretzel is not a triangle, but it has a corresponding abstract structure which is a triangle. Abstract objects within the abstract structure have corresponding real objects in the real-world thing, and, similarly, abstract relations among abstract objects have corresponding real relations among the real objects. For example, the "succession" relation between the Nth and N+1st item in an abstract list of vertebrae corresponds to the actual joint between a vertebra and the vertebra below it. This is the more "pure and correct" approach, but it has the drawback that either a.) all the axioms for, say, series have to occur in the Knowledge base twice, once for the abstract and once for the real, or b.) to conclude something (like betweenness) for real objects, Cyc has to map to the abstract domain, deduce the relations, and reverse-map back to the real objects. Both solutions are very burdensome.

Approach 2 has a so-called "commuting square":



If the top and bottom embodiments hold, and the top ListItem is followedBy the bottom one, then there *must* be an intervertebral joint between the vertebrae. Now, on the abstract side, two structurally isomorphic structures are deemed "the same thing" whereas on the real-world side two structurally isomorphic structures remain two separate individuals.

We recognize the formal merits of Approach 2, but for the reasons of not wanting to duplicate axioms, and not wanting to map back and forth across the "embodiedBy" links during reasoning, we have stayed with Approach 1 so far, for pragmatic convenience.

This structural issue will be also be important when we represent genetic structures like nucleic acid sequences and amino acid sequences.

Issue 5. Body Regions and Directions

Various anatomical "directions" including distal, proximal, ventral, dorsal, lateral, and caudal, have been defined. Originally we considered mapping these to our pre-existing "up" and "down" based predicates, but this could not be done.

"Distal" does not really mean "distant from". It means "more distant along a certain type of path". For the cardiovascular system, it is distance along vessels (primarily veins & arteries) from the heart. The aorta goes up from the heart at first, to the left, and then down; nonetheless, all junctions in the aortic arch, and in the rest of the aorta, are distal to the aortic valve.

There are two overlapping body region/direction systems commonly used – one related to body structure (caudal-cephalic, ventral-dorsal, ...) and the other related to orientation with respect to the external world, assuming that the animal is in "standard position" (superior-inferior, posterior-anterior). In human beings, with an erect spine assumed, the mapping between the two systems in the body is different than that for most other vertebrates, which normally have a relatively horizontal spine. However, in the head the mappings are the same. Cyc assertions are written with respect to body structure (so that the assertions are not restricted to human beings), but the system can answer questions posed in either modality.

Certain anatomical directions (palmar, plantar, lingual, ...) apply only with respect to certain body parts (hand, foot, teeth, ...) or have different meanings when used with specific body parts (e.g., dorsal in the foot), so Cyc restricts or modifies their meanings appropriately.

We use body direction terminology for two related sets of concepts: relative directions between body parts and regions of individual body parts. When we consider the relative directions between body parts, we note that common terminology would allow two ways of making every statement; e.g.

(dorsalTo (The ThoracicVertebrae) (The Heart)) and
(ventralTo (The Heart) (The ThoracicVertebrae)).

To obviate the need for duplicate assertions, we maintain only one predicate from each inverse pair. When specifying regions however, both region functions (e.g. DorsalRegionFn and VentralRegionFn) are necessary and provided. Thus both (DorsalRegionFn (The Spleen)) and (VentralRegionFn (The Spleen)) exist in the system.

We created an elaborate system of Paths, Traversals, and Trajectories to represent the relations between various segments and junctions of vessels. There are some predicates (relations) for which an existing path system must specified as part of every assertion or query, whereas other predicates assume the existence of a conventional, unspecified path system. Predicates like "betweenOnPath" work for all path types. "betweenOnPath" is used to define "distalToFrom" (a ternary predicate specifying the reference organ to be measured from), which is used in turn to define "distalTo" (a binary predicate that assumes a standard core organ for a given path type). Thus "distalTo" for the bronchi means something very different from "distalTo" for the arteries.

Issue 6. Non-exact Citations to Terms of External Standards

In mapping Cyc medical concepts to concepts in cited external standards and thesauri like WordNet, MeSH, SNOMED, etc., we found that only sometimes is there a direct one-to-one mapping to cite. For this kind of exact citation we use the ternary predicate:

```
(synonymousExternalConcept
    CYC-CONCEPT
    EXTERNAL-SYSTEM-NAME
    STRING)
```

as in, for example:

```
(synonymousExternalConcept
    Cerebrum
    MeSH-Information1997
    "Telencephalon | A8.186.211.730.885").
```

In our web-based interface, clicking the mouse-button on "MeSH-Information1997" yields a fuller bibliographic citation.

Where the citation to MesH is only approximate, but close, we use overlappingExternalConcept in the same way:

```
(overlappingExternalConcept
    Eyelash
    SNOMED-Information1996
    "T-01530 Eyelashes")
```

The more interesting case is when there is some transformation necessary between the meaning of the string in the cited external system and some concept (or some complex expression made of concepts) in Cyc. For this we use MeaningInSystemFn, a function that takes the name of a cited external system and the cited string in that system, and return as the value whatever (abstract) "meaning" the cited string has in that system. For a simple example:

```
(implies
 ( (MeaningInSystemFn SENSUS-Information1997
   "LESS-THAN-COMPARISON")
   ?X
   ?Y)
 (greaterThan ?Y ?X)).
```

This says that what SENSUS means by "LESS-THAN-COMPARISON" (for any X and Y) implies the superficially different assertion in Cyc: (greaterThan ?Y ?X), which has its arguments reversed. Cyc does not maintain inverse versions of a relation; instead it just reverses the order of arguments when necessary.

Similarly, this function can be used when the corresponding term is more general or more specific than the term in question:

```
(genls
   VisceralPleura
   (MeaningInSystemFn
      MeSH-Information1997
      "pleura | A4.716"))
```

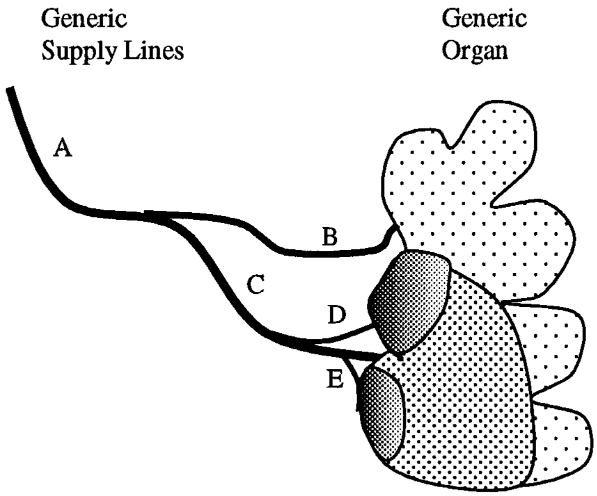
where "genls " means generalizations. Thus the Cyc concept VisceralPleura has as a generalization whatever MeSH means by string "pleura | A4.716".

Another way to link Cyc concepts to terms in external systems, beyond simply citing the relevant parts, is to use Cycorp's Multi-Thesaurus Correlator/Manager product. This is software that manages an unlimited number of large thesauri simultaneously, with an unlimited number of relation types, output filters, integrity constraints, etc. Our "correlator" automatically suggests candidate terms for correlation of a term in one thesaurus to preferred terms in another thesaurus. It uses the underlying Cyc ontology to suggest candidate terms based on semantic meaning rather than just word-resemblance. The enormous work of integrating two large thesauri properly is drastically cut, since the user does not have to hunt around aimlessly trying to find the best "matching term" (if there is one at all). We loaded a combination of the UMLS Metathesaurus and the Semantic Network into our Multi-Thesaurus Correlator/Manager to demonstrate that it could handle large thesauri with complicated structures.

Issue 7. Directness of functional service

Cyc's anatomy project developed the predicate "directlyServesBodyParts".

The need for the "directly" is that too many parts depend indirectly on too many others. For example the spleen depends on the aorta for blood, but it more directly depends on the splenic artery. Arguably all parts depend on the heart, liver, or brain. As you move in the distal direction along some nerve, artery, vein or bronchus, you find that the area served is smaller and smaller. This applies generically at several levels, to several body systems.



Here line A directlyServes the whole organ, line C directlyServes the lower portion, and line E directlyServes only one particular lobe. A test would be: "What part is immediately affected if the line is severed?"

Cyc considers the most fine-grained nerve/artery/vein/ lymph duct/... defined in the system to "directlyServe" a given body part. If and when more detailed work is done on a given region of the body, the directlyServes assertions may be modified and made more specific.

Issue 8. Body Systems Versus Locations in the Body

Body systems are a common way of classifying body parts, as when an artery is said to be part of the cardiovascular system. Most body parts are ascribable to some system, although there are some, like the spleen, greater omentum, or pituitary, for which several or no systems may be appropriate.

A completely separate way of classifying parts is by their gross physical location. Body systems overlay and spatially intersect the system of body regions, like the left forearm, the thorax, the neck, the head, etc. Almost every organ or bone will have positions in both.

Issue 9. Specialized Prototypes and the Veterinary Taxonomy Problem

Using the "prototype" approach as described in Issue 1, we had to create a prototypical creature with certain body parts. Some parts are common to more than just human beings, and occur in other species, other classes, phyla, etc. It was tempting, but practically quite burdensome, to make a prototype for every biological taxonomic level. We settled on the following three simplified Microtheory (Mt) levels for medical purposes: VertebratePhysiologyMt and VertebrateBodyStructureMt contain facts true of almost all vertebrates. HumanPhysiologyMt and HumanBodyStructureMt contain facts true of almost all normal mature human beings. The final specialization is by sex. MaleHumanBodyStructureMt and FemaleHuman-BodyStructureMt deal with gender-linked body parts of the prototypical man and woman, respectively.

We chose not to use every taxon above *homo sapiens* to create a new structure microtheory for the individuals in that taxon. We avoided creating prototype microtheories for Hominid, Primate, Mammal, Chordate, and Animal. For the time being, if we want to note that a typical snake has one lung, we state that as an exception to the rule that vertebrates have two lungs rather than creating a whole Microtheory for snake anatomical structure. [Fish and immature amphibians are exceptions that have no lungs.] We do not have human characteristics spread out among Family, Order, Class, Phylum, etc. taxon Microtheories. In theory we could assign each bodily feature to the most general taxon in which it is common. We would do this if we were to tackle general zoological anatomy in depth.

Related Work

We are aware of several medical vocabulary-servers, thesauri, and taxonomies. The UMLS Metathesaurus (UMLS 1998), SNOMED (Côté 1993), MeSH (MeSH 1997), Read Codes (Read 1997) , NANDA (NANDA 1994), the VA Clinical Lexicon (DVA 1994), ICD-9/10 (ICD-9 1997, ICD-10 1992), ICIDH-2 (ICIDH 1998), CPT (CPT 1996), and the Alcohol and Other Drugs Thesaurus (AODT 1993) are systems that have some hierarchical structure of medical terms, either explicitly or implicitly in the syntax of assigned codes. The medical parts of more general works like Wilkins' An Essay Toward A Real Character and Philosophical Language (Wilkins 1668), Roget's Thesaurus (Roget 1852-1995), Word Menu (Glazier 1992), WordNet (Miller et al., 1993-1998) and the Macquarie Thesaurus (Bernard 1986) have similar structures.

A taxonomy of terms (especially if it is strictly enforced that a child-node is a genuine subclass of its parent-node(s)) forms what we call the "bones" of a medical ontology. The flesh would be the information stored on each term, including ground facts, relation to other terms, rules, axioms, formal logical definitions and constraints. In fact, though, the systems just listed are not strict in requiring that a child-term represents a true subclass of the parent term(s); often the relation is merely a vague "Narrower Term" (NT) relation, (as is used in Information Science thesauri), and the actual relation may really be subclass, sub-part, sub-region, tools used, typical participants, typical causes and effects, symptoms, etc.

Although the GALEN system (Rector et al. 1994, GALEN 1998) includes a terminology server, it is more than just "bones" in that it aims to include definitional information for its concepts. It was implemented using GRAIL, a member of the KL-ONE family of knowledge representation languages (some are sometimes called "description logics"). These languages have deliberately limited expressiveness (which precludes including certain kinds of medical information, like negative information or exception information), and they emphasize classifying concepts given their definitions. GRAIL may be extended to allow greater expressiveness. (Goble et al., 1995).

Some user of UMLS have extended it by representing UMLS facts as Conceptual Graphs (a graphic representation of first order predicate logic). In AQUA (Johnson et al. 1993), the type constraints on relations in the UMLS Semantic Network were used to provide type-restrictions on the conceptual graphs; similarly, in Cyc, every relation has type restrictions on its arguments.

The ON-9 Ontology Group at ITBM-CNR (the Biomedical Technologies Institute (ITBM) of the Italian National Research Council (CNR)) has begun to build a medical ontology in the KIF and Ontolingua languages. We don't know how ON-9 handles the issues we have mentioned in this article, but since the KIF language is based on first order predicate logic, some of the same problems can be expected to arise. ON-9 does not have the advantage of a large, pre-existing ontology of real-world classes with inheritable axioms; however, it is able to make use of a number of pre-named concepts in Ontolingua modules created by others for other purposes.

There are a couple of dozen currently fielded medical expert systems. Most of them use their own representations of just the data needed to give their answers, and are not concerned with integration into any larger or general medical knowledge base. The work on ILIAD by Bouhaddou and his colleagues may be an exception (Lincoln et al. 1991).

Current Status

We have encoded normal adult human gross anatomy (excluding the lymphatic system, which had not been entered as of April 1998). We have represented more detailed levels of organs, bones, etc. than MeSH does, but not quite the level of detail covered in SNOMED.

All major organs and their significant parts have been entered into Cyc. The "main" vessels and nerves are done, -- all human bones, 85 muscles, about 60 veins (including venous sinuses), 80 arteries, and 150 nerves. These figures do not count right and left versions of named body parts separately. It will be straightforward to specialize any area in greater detail, as needed.

For all defined body parts, the orientation, specific end-connections, and general location are described (for left and right sides, where applicable), as well as the "parts directly served" by each. We have entered very little physiology or functionality of the organs, other than what serves what, so far. In the lung, we have defined named parts to the level of segment of lobe (and the segmental bronchi), and to the alveoli as described multifarious entities. Bones are classified into shape classes, joints into connection types. In some case we have notable parts of specific bones.

The total time needed to create the whole gross anatomy system so far was about one staff-year of work. This includes the time spent resolving the issues discussed in this paper.

Future work

Our planned medical "ontologizing" covers histology, physiology, molecular biology, diseases, treatments, and pharmacology. This work will include detailed axioms and rules as does the anatomical work already entered. The order and rate at which this work will be done will depend upon internal priorities, the desires of our corporate clients, availability of funding, and relevance to other projects within Cycorp. There is no known technical obstacle to entering such information and relating it properly to the rest of the Knowledge Base.

Acknowledgements

We are grateful to Glaxo-Wellcome, United Health Care, and Pfizer for supporting this effort.

References

AODT 1993. *Alcohol and Other Drugs Thesaurus,* A Guide to Concepts and Terminology in Substance Abuse and Addiction, First Edition, Public Health Service, U.S. Department of Health And Human Services, Washington, D.C.

Bernard, J. R. L., General Editor, 1986. *The Macquarie Thesaurus,* The Macquarie Library Pty. Ltd., Macquarie University, North Ryde, New South Wales, Australia.

Côté, R.; Rothwell, J.; Palotay, J. L.; Beckett, R. S. and Brochu, Louise, Editors 1993. *The Systematized Nomenclature of Human and Veterinary Medicine (SNOMED International),* Vol. I, College of American Pathologists, Northfield, Illinois.

CPT-96 1996. *Physicians' Current Procedural Terminology (CPT-96),* American Medical Association, Chicago.

DVA 1994. *Clinical Lexicon Technical Manual,* Version 1, Sep. 1994, Salt Lake Information Systems Center , Medical Information Resource Management Office, Veterans Health Services, Department of Veterans Affairs, Salt Lake City, Utah, USA.

GALEN 1998. *GALEN* World Wide Web pages, accessed May, 1998:
http://www.cs.man.ac.uk/mig/galen/brochure.html
http://www.cs.man.ac.uk/mig/galen/overview.html

Glazier, S. 1992, *Word Menu,* Random House, New York.

Goble, C.; Bechofer, S.; Solomon, D.; Rector, A.; Nowlan, A. and Glowinski, A., 1995. Conceptual, Semantic and Information Models for Medicine, in *Proceedings of the 4th European-Japanese Seminar on Information Modelling and Knowledge Bases,* Reprinted *in Information Modelling and Knowledge Bases VI,* IOS Press, Amsterdam.

ICD-9/CM 1997. *The International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM),* World Health Organization, version printed by Context Software Systems, McGraw-Hill, New York, 1997; see U.S. Dept. of Health & Human Services Publication No. (PHS) 91-1260.

ICD-10 1992. *International Classification of Diseases and Related Health Problems, 10th Revision (ICD-10),* World Health Organization, Geneva.

ICIDH 1998. *ICIDH-2: International Classification of Impairments, Activities and Participation: A Manual of Dimensions of Disablement and Functioning,* Beta Version, World Health Organization, Geneva. Accessed 1998 at World Wide Web Page http://who-hq-policy.who.ch/cgi-bin/folioisa.dll/icidh/

Johnson, S.B.; Aguirre, A.; Peng, P. and Cimino, J. 1993. Interpreting natural language queries using the UMLS. In *Proceedings of the Annual Symposium on Computers in Applied Medical Care,* 294-298. This paper describes AQUA (A QUery Analyzer).

Lenat, D. B. 1995, Artificial Intelligence, *Scientific American,* September 1995.

Lenat, D. B.; Guha, R. V.; Pittman, K.; Pratt, D. and Shepherd, M. 1990, Cyc: Towards Programs with Common Sense, Communications of the ACM, 33(8):30-49.

Lincoln, M.J.; Turner, C.W.; Haug, P.J.; Warner, H.R.; Williamson, J.W.; Bouhaddou , O.; Jessen, S.G.; Sorenson, D.; Cundick, R.C. and Grant, M. 1991 Iliad training enhances medical students' diagnostic skills. *Journal of Medical Systems* 14(1):93-110.

MeSH 1997. *Medical Subject Headings (MeSH) -- Tree Structures,* National Library of Medicine, Bethesda, Maryland, USA.

Miller, G. A.; Beckwith, R.; Fellbaum, C.; Gross, D. and Miller, K. 1993/1998. Introduction to WordNet: An On-Line Lexical Database, August 1993; *Five Papers on WordNet,* accessed in 1998 at the WordNet World Wide Web page http://www.cogsci.princeton.edu/~wn/.

NANDA 1994. *Classification of Nursing Diagnoses*, North American Nursing Diagnosis Association, Lippincott, Philadelphia.

Read 1997. *The Read Codes, Version 3* CD-ROM, NHS Centre for Coding and Classification, Information Management Group, National Health Service, Loughborough, UK.

Rector, A.; Gangemi, A.; Galeazzi, E.; Glowinski, A.; and Rossi Mori, A. 1994. The GALEN CORE Model Schemata for Anatomy: Towards a Re-Usable Application-Independent Model of Medical Concepts. In *Proceedings of the 12th International Congress of the European Federation for Medical Informatics (MIE94)*

Roget, P. M., 1852-1995. *Thesaurus of English Words, Phrases, classified and arranged so as to facilitate the Expression of Ideas and assist in Literary Composition, (Roget's Thesaurus)*, many editions, many editors, many publishers.

UMLS 1998. *UMLS Knowledge Sources*, book and CD-ROM, National Library of Medicine, Bethesda, Maryland, USA. See the 1998 NLM UMLS World Wide Web page: http://www.nlm.nih.gov/research/umls/UMLSDOC.HTML.

Wilkins, J. 1668, *An Essay Towards a Real Character and a Philosophical Language*, Royal Society, Gellibrand and Martin, London.