To be published in AIME '95

Coordinating Taxonomies: Key To Re-usable Concept Representations

AL Rector
Medical Informatics Group, Department of Computer Science
University of Manchester, Manchester M13 9PL, England
Tel +44-161-275-6133/6188, FAX: +44-161-275-6932, Internet: rector@cs.man.ac.uk

Abstract: A unified controlled medical vocabulary has been cited as one of the grand challenges facing Medical Informatics. We would restate this challenge as 'achieving a re-usable and application-independent representation of medical concepts.' Achieving a re-usable representation of medical concepts is a pre-requisite for meeting two key strategic goals of the next decade of the development in medical informatics: interoperability and cumulative development. A key strategy for achieving re-usability is to separate concepts into their component parts, organise those parts in nearly pure hierarchies, and then recombine into composite representations which can be classified flexibly and automatically. This paper explores the means and consequences of this strategy as implemented in the GALEN project. It discusses both the strengths — providing greater detail, greater computer support, and avoiding many arguments which are endemic in discussions of classification systems — and the limitations intrinsic in such a formal approach.

1. Introduction

1.1. A Re-usable representation of medical concepts

Achieving a unified controlled medical vocabulary has been cited as one of the grand challenges facing Medical Informatics. We would restate this challenge as 'achieving a re-usable and application-independent representation of medical concepts.' Achieving such a re-usable representation of medical concepts is a pre-requisite to meeting two key strategic goals of the next decade of the development in medical informatics: interoperability and cumulative development.

Complete application independence is, of course, a chimera. Any representation will only be re-usable within a limited area and for limited purposes. In this paper we will concern ourselves primarily with concepts being used for 'patient centred information systems' — systems used directly by clinicians as a routine part of clinical care to capture information which may then be re-used in other secondary application. We will discuss a specific strategy for achieving important aspects of re-use — separating taxonomies cleanly and then re-coordinating them by defining composite concepts and classifying those composite concepts automatically — and we will explore a number of issues raised by this strategy.

1.2. The Barriers to Re-use and application independence

Why is it so difficult to achieve a re-usable representation of medical concepts — or of the concepts used in any other field of endeavour? The aspiration to re-use and application independence is not new. Attempts to produce standard classifications and to formalise their logical foundations go back at least to Aristotle. The first attempts at standardised classification and coding systems for medicine date back to the middle of the nineteenth century.

In this paper we argue that one key problem limiting re-use and application-independence is that the formalisms used force application-specific choices on developers of vocabularies and classifications. We argue that one of the key application-specific decisions which formalisms force on users is the organisation of a fixed taxonomy based on indivisible representations of concepts. Inevitably, indivisible representations of complex concepts conflate several ideas, and since any fixed organisation of such complex ideas is fundamentally arbitrary, choices can only be made on the basis of the current application. We contend that avoiding such decisions is an important technique for achieving re-use. The general technique is:

a) To separate the concepts into their apparently elementary parts for this representation;

b) To form clean taxonomies of the parts

c) To recombine the parts by defining the original concepts in terms of the elementary parts

d) To classify the composite concepts according to formal rules.

Our contention is that it should only very rarely be necessary for an elementary entity to have more than one parent within a taxonomy, though composite entities may have indefinitely many parents formally identified on the basis of their definitions. Our experience is that arguments about where something should be classified almost invariably indicate that the concept has not been adequately decomposed. When decomposed and defined in terms of more elementary concepts, the arguments usually disappear.

The problem of conflating several ideas into a unitary representation is particularly serious in standard mono-axial classification systems such as ICD-10, ACRNEMA or the Read Clinical Classification, but is common in other representations. Consider for example the nomenclatures for Internist/QMR and HELP [1].

A second related mechanism for avoiding application-specific decisions — bridging the levels of detail needed in different applications — is mentioned briefly but will be the subject of a separate paper.

These ideas are not necessarily new. Other authors have suggested using compositional terminological systems as an interlingua for reconciling clinical terminologies, *e.g.* [1-4] or for mediating between databases, *e.g.* [5-7]. The purpose of this paper is to emphasise the specific issue implicit but often not emphasised in these writings — that the use of compositional systems allows clean separation of taxonomies and principled construction of multiple classifications — and to explore some of the specific consequences of this policy raised by GALEN's implementation.

1.3. Background

This paper presents techniques and a point of view developed during the GALEN project[1] which is developing 'Terminology servers' based on compositional models of

---

1 General Architecture for Languages Encyclopædias and Nomenclatures in Medicine. The members of the GALEN consortium are: University of Manchester (UK, Coordinator), Hewlett-Packard UK Ltd, Hôpital Cantonal Universitaire de Genève (Switzerland), Consiglio Nazionale delle Ricerche (Italy), University of Liverpool (UK), Katholieke Universiteit Nijmegen (Netherlands), University of Linköping

medical concepts to act as a means for integration and interoperability of clinical applications. The examples are drawn from the GALEN CORE Model, version 5, in its form as of autumn 1994 [8]. The work has been used as the basis for clinical user interfaces and for conversion amongst existing coding systems. Work to test the wider re-usability of the representation is in progress.

GALEN is based on the assumption that there is a distinctive 'concept level' or 'terminology level' which can usefully be represented separate from both the 'linguistic level' and more 'general inference' or 'assertional level'. GALEN is an attempt to realise a strong notion of such a concept level — a single concept model which can be re-used by different linguistic systems and different general inference systems. This is a strong sense of re-use — not merely that one representation can be transformed into another, but that a common model can be established which can serve as a single basis for many different applications.

GALEN uses the GALEN Representation and Integration Language (GRAIL) Kernel [9, 10], a 'description logic' or 'knowledge representation language' related to KL-ONE [11], CLASSIC [12], or BACK [13] and to Conceptual Graphs [14] but with special features to support the requirements of coordinating taxonomies, particularly for coordinating the 'kind-of' (subsumption) relation with other transitive relations. Most of the discussion is equally applicable to other related systems. A formulation of the relationship between part-whole relations and subsumption analogous, though not identical, to that presented here has been developed for conceptual graphs by Bernauer [15]. Related formulations for representing SNOMED and other medical records have also been developed using Conceptual Graphs by Campbell [3] and a series of models with some features in common have been developed by the CANON group and presented as a series of papers [2, 4, 16]. GRAIL is strictly terminological, it does not, itself, provide any 'assertional' or 'general inferential' component — in the language of the KL-ONE community [17, 18] it provides only a 'T-Box', albeit an extended one, but no 'A-Box'. The goal is that the same 'T-Box' should be re-usable by various inferential mechanisms.

2. Fundamental Ideas

2.1. Taxonomies: Kinds, parts, and other relations

A 'taxonomy' is a hierarchical structure. Most coding and classification systems are taxonomies, but taxonomies are also widely found in biology and elsewhere in engineering and science. Thesauri, which are one kind of taxonomy, are almost universally used as means of accessing bibliographic literature.

A taxonomy may be organised according to a single relationship or a combination of relationships. Thesauri typically avoid defining the precise relationship between different levels of the taxonomy by describing them simply as 'broader than' and 'narrower than'. Other systems are more explicit and specify the specific relationship used in a taxonomy.

(Sweden), The Association of Finnish Local Authorities (Finland), The Finnish Technical Research Centre (Finland), GSF-Medis Institut, (Germany), Conser Sistemi Avanzati (Italy)

The two most important types of relationship used to organise taxonomies are 'kind of' and 'part of' — sometimes called 'generic' and 'partitive' relations respectively. For example, 'automobile' is a kind of 'vehicle' but 'wheel' is a part of 'automobile'. Other taxonomies may be based on other relationships, *e.g.* causation. The 'kind of' relation is often referred to as 'subsumption' and we may say that 'vehicle subsumes automobile'. In most knowledge representation systems, subsumption plays a special role — anything which is true of a concept is also true of any other concept which it subsumes.

2.2. Description and Classification

Most traditional classification or coding systems consist of a set of indivisible entities explicitly classified in 'taxonomies'. In many cases a single entity can only be placed in only one place in the taxonomy, *i.e.* any one code can have only one 'parent'. The arrangement of the taxonomy is typically related to the original use of the classification — hence aetiology and infections play a major role in the International Classification of Diseases because of their importance in traditional epidemiology. Similarly, anatomical structure plays a more important role in ACRNEMA because of the importance of anatomical structures to radiographic findings. Even if more than one parent is allowed, few systems record the reason for different parents, hence it is difficult or impossible to follow a particular alternative view through consistently. Cross references are provided in SNOMED and to a lesser extent in ICD, but they must be maintained laboriously by hand, and there is no way to check that they are complete.

The fundamental idea of the GALEN representation is that most concepts are represented by descriptions as 'composite entities' rather than by elementary indivisible entities. For example, "pulmonary tuberculosis" might be represented as "an inflammation of the lungs caused by an infection by tubercle bacillus".

Composite entities can be classified according to their compositional structure in whatever way is appropriate for a given application. For example, pulmonary tuberculosis can be classified either as a "disease of the lung" or a "inflammation caused by an infection by tubercle bacillus" or both as required.

The classifications need not be specified in advance. Classification can be performed as required — *e.g.* according to anatomy, morphology or aetiology. For any system representing a wide range of concepts, the total number of possible classifications is indefinitely large.

2.3. Natural Kinds

While the idea of GALEN is that most concepts be represented by defining descriptions, not everything can be defined. The system must start with elementary entities; Furthermore, many concepts are best treated as 'natural kinds' — sometimes referred to as 'polythetic categories'. Roughly speaking, natural kinds are recognised rather than inferred. For example, "Rheumatoid Arthritis" is not well enough understood to be defined, but is a recognised syndrome. Within a model, natural kinds are represented as elementary entities rather than composite entities. However, the decision to represent a given concept by an elementary entity does not imply any philosophical commitment to view it as a natural kind, merely the pragmatic decision not to define it within this model — *i.e.* not to identify a set of *sufficient* conditions for its recognition.

Note that even though there is no set of sufficient conditions in the model, there may still be necessary conditions, *e.g.* that “Rheumatoid Arthritis is caused by an autoimmune process.” From this it would follow that “rheumatoid arthritis” would be classified as an “autoimmune disease” but being an auto-immune arthritis would not be sufficient for some other form of arthritis to be classified as “rheumatoid arthritis”. The ability to distinguish between representations of necessary and sufficient criteria for concepts is one of the distinctive features of the GRAIL Kernel. Necessary conditions affect how an entity is classified but do not affect what further entities are classified under it.

The taxonomy of natural kinds forms the skeleton of the model — the elementary entities from which composite descriptions are built up.

2.4. Transitive relationships

Part-whole and causal relationships play a key role in medical concepts. Any system hoping to represent medical concepts adequately must deal with such transitive relations in a consistent way. Transitive relations present special problems to a programme of representing most concepts by composite entities in a model because they interact in specific ways with the kind-of or ‘subsumption’ taxonomy.

For example, the “shaft of the femur” is a part of the “femur” rather than a kind of the femur. However, a “fracture of the shaft of the femur” is a kind of “fracture of the femur”. Conversely, if we focus on the femur rather than the fracture, a “Femur which has a fractured shaft” is a kind of “fractured femur”. Similar patterns hold for direct causation, though not necessarily for loose causal associations. An important and novel features of the GRAIL Kernel is that it implements coordination of transitive and kind-of taxonomies in this way.

2.5. Managing granularity and bridging level of abstraction

Any one application abstracts away many details as irrelevant and emphasises others which are particularly relevant. For example for purposes of most applications, “Gastric Ulcer" is sufficient detail. However, for a pathology system, one might wish to organise a classification of lesions according to the tissues affected, in which it is critical to represent the fact that ulcers actually affect the “mucosa of the wall of the Stomach”. To be re-usable, a representation must allow applications to ignore details which are irrelevant to them and allow other applications to extend the representation where they require more detail without disturbing existing applications.

3. Special Issues: Roles, Signs & Symptoms, Systems, Dualities, and Proper Names

3.1. Roles

What is a “vitamin”? a “hormone”? a “steroid”? a “protein”? Clearly, “vitamin” and “hormone” are functionally defined whereas “steroid” and “protein” are structurally defined. There are many analogous situations in which the detail of the functional processes are beyond the scope of the model but where a clear separation of taxonomies is necessary. Such distinctions are represented in the GALEN CORE model by ‘roles’. For example:

(ChemicalSubstance <u>which</u> playsPhysiologicalRole-VitaminRole) <u>name</u> Vitamin

ChemicalSubstance which playsPhysiologicalRole-HormoneRole) name Hormone

This use of 'roles' has become one of the fundamental mechanisms of the model. It has virtually replaced explicit statements that an entity has more than one parent. It is almost always preferable to represent the subsumption formally using a role as above, because the use of the role explains *why* the one entity is subsumed by the other whereas the use of an explicit subsumption link to two parents does not.

3.2. Signs, Symptoms and Diagnoses — method of observation.

The distinction between "sign", "symptom" and "diagnosis" plagues many medical classification schemes. Frequently the same phenomenon can play more than one role — "wheeze" may be reported by the patient and hence be a symptom or be seen by the doctor and hence be a sign. Some types of "seizure" may be observed and treated as signs or reported as symptoms; in other situations it is only after long investigation that a particular phenomenon is diagnosed as being a seizure. The GALEN CORE Model surmounts this problem by separating the taxonomy of methods of observation from the taxonomy of pathophysiological conditions. For example:

(Disorder which isShownBy PatientReport) name Symptom.
(Disorder which isShownBy MedicalObservation) name Sign
etc.

This approach has the advantage that it extends naturally to other related distinctions, such as whether a wheeze was heard with a stethoscope, heard without a stethoscope by the doctor, or reported by the patient.

3.3. Functional parts and Functional systems

There are at least two uses of the word system in common usage: those based on function such as "the digestive system", "the circulatory system", or the "endocrine system" and those based on anatomical structure such as "the skeletal system" or the "gastrointestinal tract". The CORE Model distinguishes these two uses clearly but forms both compositionally. Functional systems are composed of those entities with the requisite function; anatomical systems are treated simply as collections.

3.4. Dualities

Many medical phenomena come in pairs. The most common pairing is process and lesion, *e.g.* 'ulceration' and 'ulcer' . Often the same word is used for both members of the pair, *e.g.* 'erosion' (a process) and 'erosion' (a lesion). Often common usage blurs the distinction, for example we speak both of 'gastritis' as if it were a process — 'worsening', 'progressive', 'chronic', etc. — and of 'gastritis' as something which can be seen through a gastroscope — clearly a physical lesion. At least two cases exist, those where the physical object is the outcome of a process and those where one physical object or process is caused by another.

The phenomenon is quite general and not limited to pathological lesions, *e.g.* consider the process of 'secretion' and the 'secretion' which is its outcome. Language blurs these distinctions in some cases and emphasises them in others — *e.g.* there is a clear distinction in most languages between 'viral hepatitis' and 'hepatitis virus'. A formal model must either mirror this pattern blurring and distinguishing or provide formal transformations between the different forms. If it is to be independent of surface language, it must do so in a way which does not commit it to the usage in any one linguistic community.

3.5. Proper Names and the Definite Article

There are many anatomical structures named by the combination of a general type and a specific organ or location — *e.g.* "the hepatic artery", "the common ileac vein", or "the islets of Langerhans". These concepts present special problems. The first two examples are fundamentally different from the third. For example, "the hepatic artery" is not the only artery supplying the liver, hence it is a different concept from the category "artery which supplies the liver" which subsumes "the hepatic artery" along with a number of other less important arteries. By contrast, the "islets of Langerhans" is a simple eponym — "Langerhans" has no other meaning in the system. The phrase "islets of Langerhans" can be treated as special string or a pseudo-attribute can be created to link "islet" and "Langerhans" without danger of misunderstanding.

In an early version of the CORE Model, to emphasise this distinction, all general entities were labelled with plurals — "arteries which supply the liver", "bones", "hands" etc. This convention avoided the confusion with specific named entities such as "hepatic artery" but was felt to be awkward and at variance with standard practice in the English speaking community.

There is a fundamental conflict in these cases between the needs of linguistic systems and the needs of concept systems. To the concept system proper, both "Hepatic Artery" and "Islets of Langerhans" are specific named entities best treated as natural kinds and represented as elementary entities. However, to ignore the common compositional structure of the phrases in most languages reduces the value of a compositional model to parsing systems. The current model compromises. A special set of subattributes, *servesSpecifically, hasSpecificCause, hasSpecificOutcome,* etc. are provided which are convey the strong partly linguistic linkage while still retaining the compositional structure.

4. An Extended Example

4.1. Anatomy, morphology and process

The most obvious area for use of compositional models coordinating several independent clean taxonomies is in the basic description of lesions, their location, form and cause.

For example, classifications of burns can be made by whether they are thermal or chemical, their location, their penetration, their extent, and their circumstances. There is little *a priori* reason to choose one ordering of these features over another — different applications require different information. An occupational health system is likely to be concerned with the circumstances whereas an intensive care unit is much more likely to be concerned with the extent and depth of the burn. Furthermore, the occupational health unit may be more concerned with 'circumstances in which burns occur' than the 'burns occurring in particular circumstances'.

The GALEN CORE allows statements such as such as:

BurnLesion which <
hasLocation Arm
hasDepth halfThickness
hasExtent 4cm2
hasCircumstances KitchenAccident
hasCause Heat>

There are separate classifications for anatomical entities such as *Arm*, for energy such as *Heat, and* for the depth of burns, areas, and types of circumstances in which accidents may occur. Any of the dimensions may be refined further, *e.g.*:

BurnLesion which <
hasLocation (AnteriorSurface which isDivisionOf LowerArm)
hasCircumstances (KitchenAccident which involves FatOrOil)
hasCause Heat>

Given these compositional structures we can classify them as required in more abstract categories such as:

BurnLesion which hasLocation UpperExtremity

i.e. “burns of the upper extremity”, or

BurnLesion which <
hasCircumstances HouseholdAccident
hasCause Heat>

i.e. “thermal burns due to household accidents”. Since the classification of burns does not have to be established in advance, data entry need be in no specific order. Rather than look up and down a fixed hierarchy, the user may be presented with an overall ‘predictive data entry’ form using and the ability to expand any item as discussed in detail in [19-21]

BurnLesion which hasCause Chemical
BurnLesion which <hasCause Chemical hasLocation UpperExtremity>
BurnLesion which <hasCause Chemical hasLocation Hand>
BurnLesion which <hasCause Chemical hasLocation PalmarSurfaceOfHand>
...
...
BurnLesion which hasCause Acid
BurnLesion which <hasCause Acid hasLocation UpperExtremity>
BurnLesion which <hasCause Acid hasLocation Hand>
BurnLesion which <hasCause Acid hasLocation PalmarSurfaceOfHand>
...
...
BurnLesion which hasCause Alkali
...
...
...
BurnLesion which hasCause Heat
BurnLesion which hasCause Heat
BurnLesion which <hasCause Heat hasLocation UpperExtremity>
BurnLesion which <hasCause Heat hasLocation Hand>
BurnLesion which <hasCause Heat hasLocation PalmarSurfaceOfHand>
...
...

Figure 1: Part of taxonomy derived by composition and classification

We may further generate fixed classifications along any of the potential axes, for example by circumstances then depth then anatomy as shown in Figure 1. .

Note that even in this simple example, depending on interest, there are at least 2^4=16 possible lesions and 4!=24 possible orderings. If the internal structure of the axes such as ‘involving OilOrFat’ are taken into account there are many more. Any one of these orderings may be particular useful for a specific application, but there is no fundamental reason for choosing one over the other in general. Any choice amongst them biases a representation in favour of one group of applications and makes it awkward for use by others.

Note that in this example we have chosen to treat *BurnLesion* as elementary, as a natural kind. *BurnProcess* is defined in the GALEN models as the composite entity *Process which hasOutcome BurnLesion*. We have chosen to make lesions rather than processes elementary consistently throughout the model as it seems to produce a slightly simpler structure. However, this choice is fundamentally arbitrary.

5. Paradoxes and Limitations

5.1. Choices concerning dualities

The GRAIL Kernel is strongly typed and lacks disjunction. Given a duality such as *BurnLesion* and *BurnProcess* it is always necessary to choose one or the other. Yet many statements in natural language do not make this distinction clear. In many situations the distinction is unnecessary because it makes no difference. For example, “The burn was caused by acid” makes no distinction between process and lesion — acid is the cause of both the process and the lesion. By contrast, “The erosion lasted for three months” is ambiguous — was it the erosive process which lasted for three months, possibly resulting in an erosive lesion at the end of that time or has there been

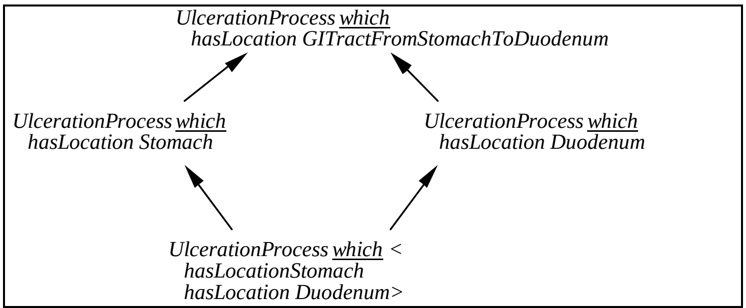

Figure 2: Paradoxical subsumption involving part-whole relations.

an erosive lesion for three months? Without context, and possibly even with it, there is no way of being confident of the answer.

On the one hand the ability to make such distinctions can be valuable. On the other it leads to problems. The first problem is that there are many situations in which we want to retrieve all cases or categories pertaining to either half of a duality — *e.g.* all cases of or all kinds of burns, erosions, etc. The second, more serious problem, is that it may lead to recording ‘over interpretations’ of doctors’ intended meaning. If there is no means of disambiguating the natural language phrase, then forcing a disambiguation risks distorting meaning. No complete answer is available at the moment within GRAIL for this problem, although several ‘work arounds’ are in use and extensions are being considered.

5.2. Paradoxes with conjunctions and regions

The rules for the interaction of part-whole relations and subsumption produce the taxonomy shown in Figure 2.:

This occurs because the stomach and duodenum are parts of the “GI tract from stomach to duodenum”. Ulceration of either organ is a kind of ulceration of that section of the GI tract. However, ulceration of both individually is a kind of each, *i.e.* “ulceration of both the duodenum and stomach” is a kind of “ulceration of the stomach”. The authors were surprised when one of the members of the consortium requested almost precisely this behaviour. They were gratified to see the formalism correspond to users’ intuitions, but remained concerned that the construction seemed ‘odd’.

The explanation of the behaviour lies in the difference in meanings of the phrases as commonly used in coding systems and medical records and a potential ambiguity in at least English usage. As used in coding systems, “Ulceration of the GI Tract in the region of the stomach through to the duodenum” almost certainly means *any* such ulceration and is properly seen as a disjunction. As used in a medial record, it would probably indicate "a continuous area of ulceration from the stomach through to the duodenum" and would correspond to a conjunction. The expression:

UlcerationProcess <u>which</u>
<hasLoation-Stomach
hasLocation-Duodenum>

captures the idea of conjunction but not the idea of the continuity and the intervening structures. The example demonstrates both the subtlety of language usage in different contexts and a previously unrecognised difference between the behaviour of discrete and diffuse lesions not catered for in GRAIL or, as far as we are aware, any related formalism.

6. Discussion

The mechanism of dividing the representation into separate taxonomies of elementary entities and then recombining these entities through composition has proven fruitful. The extent to which it has proved possible to separate taxonomies into simple hierarchies has been one of the major surprises of the GALEN project. Originally, the requirement that any entity, whether elementary or composite, should be able to have multiple parents in the subsumption hierarchy was fundamental to the design of the representation language. In early models twenty-five to thirty per cent of all elementary entities had more than one parent in the subsumption hierarchy. This was consistent with our earlier experience in the PEN&PAD project and with results from examining independently compiled hierarchies in the Oxford System of Medicine project.

In the current representation only one or two percent of all elementary entities have more than one parent. Multiple parents occur almost solely amongst the very abstract concepts at the top of the taxonomy where they serve specific purposes such as allowing a single abstraction notion of “disease” to cover pathological “lesions”, “processes” and “states”. All other multiple classification is now done formally through classification of composite concepts.

However, the power of the approach is not without cost. Issues are raised which do not arise in traditional representations, such as forced distinction between process and structure in natural dualities and problems in dealing with proper nouns which appear descriptive such as “the hepatic artery”. Amongst the most powerful ideas are the use of ‘roles’ and the flexible handling of "sign", "symptom" and "diagnosis" which follows from separating the observation method from the phenomena observed.

Of the broader claims at the beginning of this paper, the usefulness of a separate concept layer seems to us increasingly well established. The examples given here of its use to clarify the classification of concepts from independent taxonomies is only one important area where this separation, empirically, provides a manageable level of complexity. Subjectively, it has provided important for focusing for attention of the knowledge engineers compiling the model and for providing a general test for resolving arguments. The GALEN model strives to capture the expected behaviour of the concepts in applications for data entry and retrieval. It attempts to capture neither the detailed diagnostic criteria nor the idiosyncrasies of linguistic usage.

It must be emphasised that the strategy of ‘coordinating taxonomies’ is not, in itself, sufficient to achieve re-use and sharing of terminology. We have touched briefly, for example, on the issue of bridging the levels of detail needed in different applications which must also be addressed to achieve effective terminological re-use. Nor is terminological re-use sufficient to achieve complete sharing of knowledge and interoperability — for example terminological methods are based on the structural composition and definition of terms and do not deal with issues requiring numerical calculations such as transformations amongst different systems of units. Furthermore, while a sharing of knowledge and interoperability at this level are necessary to achieve comparability of data, they are not in themselves sufficient, since clinical diagnostic criteria are explicitly excluded from this terminological model. If this seems paradoxical, consider the fact that two groups of clinicians can talk to each other about a condition without prior agreement on precise diagnostic criteria.

The contention of this paper is, however, that ‘coordinating taxonomies’ provides one key strategy for achieving re-use of a central core of terminological knowledge and a foundation on which to build for further layers of inference and information management for decision support and medical records. In practice, this approach has allowed a representation which has served a range of applications in data entry, medical records, drug interaction [22], and the authoring of decision support systems. So far these applications have been relatively isolated from each other, so that the fundamental claim to re-usability has only been tested for the general abstract layers of the model rather than for specific subdomains. Such tests in specific domains are just now beginning.

References

1. Masarie Jr F, Miller R, Bouhaddou O, Giuse N, Warner H. An interlingua for electronic interchange of medical information: using frames to map between clinical vocabularies. Computers in Biomedical Research 1991;24(4):379-400.
2. Evans DA, Cimino J, Hersh WR, Huff SM, Bell DS, The Canon Group. Position Statement: Towards a Medical Concept Representation Language. Journal of the American Medical Informatics Association 1994;1(3):207-217.
3. Campbell KE, Das AK, Musen MA. A logical foundation for representation of clinical data. JAMIA 1994;1(3):218-232.
4. Cimino J. Controlled Medical Vocabulary Construction: Methods from the Canon Group. Journal of the American Medical Informatics Association 1994;1(3):296-197.
5. Bergamachi S, Sartori C. On taxonomic reasoning in conceptual design. ACM Transactions on Database Systems 1992;17(3):385-421.
6. Maida AS. Knowledge representation requirements for description-based communication. in: Nebel B, Rich C, Swartout W, (ed). Third International Conference on Principles of Knowledge Representation and Reasoning, KR '92. Cambridge MA: Morgan Kaufmann, 1992: 232-243`.
7. Borgida A. A new look at the foundations and utility of Description Logics (or Terminology Logics are not just for the Flightless Birds). Internal report. Rutgers University Department of Computer Science, 1992
8. Rector A, Gangemi A, Galeazzi E, Glowinski A, Rossi-Mori A. The GALEN CORE Model Schemata for Anatomy: Towards a re-usable application-independent model of medical concepts. in: Barahona P, Veloso M, Bryant J, (ed). Twelfth International Congress of the European Federation for Medical Informatics, MIE-94. Lisbon, Portugal, 1994: 229-233.
9. Rector A, Nowlan W, Glowinski A. Goals for Concept Representation in the GALEN project. 17th Annual Symposium on Computer Applications in Medical Care (SCAMC-93). McGraw Hill, 1993: 414-418.
10. Rector A, Nowlan W. The GALEN Representation and Integration Language (GRAIL) Kernel, Version 1. The GALEN Consortium for the EC AIM Programme. (Available from Medical Informatics Group, University of Manchester), 1993

11. Brachman R, JB S. An overview of the KL-ONE knowledge representation system. Cognitive Science 1985;9(2):171-216.
12. Borgida A, Brachman RJ, McGuiness DL, Resnick LA. CLASSIC: A Structural Data Model for Objects. ACM SIGMOD International Conference on Management of Data. ACM, 1989: 58-67.
13. Nebel B. Computational Complexity of Terminological Reasoning in Back. Artificial Intelligence 1988;34:371-383.
14. Sowa J. Conceptual Structures: Knowledge Representation in Mind and Machine. New York: John Wiley & Sons, 1985
15. Bernauer J, Goldberg H. Compositional classification based on conceptual graphs. Artificial Intelligence in Medicine Europe (AIME-93). Munich: , 1993: .
16. Tuttle MS. The position of the canon group: A reality check. Journal of the American Medical Informatics Association 1994;1(3):298-299.
17. MacGregor R. The evolving technology of classification-based knowledge representation systems. in: Sowa J, ed. Principles of Semantic Networks: Explorations in the representation of knowledge. San Mateo, CA: Morgan Kaufmann, 1991: 385-400.
18. Brachman R, Fikes R, Levesque H. An essential hybrid reasoning system; knowledge and symbol level accounts of KRYPTON. International Joint Conference on Artificial Intelligence (IJCAI-85). Morgan Kaufman, 1985: 532-539.
19. Nowlan W, Rector A. Medical Knowledge Representation and Predictive Data Entry. in: Stefanelli M, Hasman A, Fiesch M, Talmon J, (ed). Artificial Intelligence in Medicine Europe (AIME-91). Maastricht: Springer-Verlag, 1991: 105-116.
20. Nowlan W, Rector A, Kay S, Horan B, Wilson A. A Patient Care Workstation Based on a User Centred Design and a Formal Theory of Medical Terminology: PEN&PAD and the SMK Formalism. in: Clayton P, (ed). Fifteenth Annual Symposium on Computer Applications in Medical Care. SCAMC-91. Washington DC: McGraw- Hill, Inc, 1991: 855-857.
21. Horan B, Rector A, Sneath E, et al. Supporting a Humanly Impossible Task: The Clinical Human-Computer Environment,. in: Diaper D, (ed). Interact 90. Elsevier Science Publishers, B.V.North-Holland, 1990: 247-252.
22. Solomon W, Heathfield H. Conceptual modelling used to represent drug interactions. in: Barahona P, Veloso M, Bryant J, (ed). Twelfth International Congress of the European Federation for Medical Informatics, MIE-94. Lisbon, Portugal: , 1994: 186-190.