Medical Informatics Europe '96
J. Brender et al. (Eds.)
IOS Press, 1996

The GALEN High Level Ontology

AL Rector JE Rogers P Pole
Medical Informatics Group, Department of Computer Science,
University of Manchester, Manchester, UK, M13 9PL

Abstract. This paper summarises the top level schema of the GALEN Common Reference model. The top levels of a structure tested in a range of uses is described. This includes a taxonomy and an associated set of semantic link types. The current status of the project and its goals is discussed.

1. Introduction

GALEN is a large project developing terminology servers and data entry systems based on a Common Reference, or CORE, model for medical terminology [1,2]. In previous papers we have described the schema used for modelling anatomical structures [3], and the overall modelling style [4]. This paper summarises the highest levels of the GALEN schema for 'ontology'. Further detail on this material can be found at the URL http://www.cs.man.ac.uk/mig/galen/ or by e-mail from galen@cs.man.ac.uk.

We use the term 'ontology' here to indicate the model of the categories within the universe of discourse, plus sufficient information about those categories to allow them to be classified automatically. We take 'ontologies' to be language independent, using the broader term 'terminology' for an ontology linked to linguistic information. We also assume the ontology to be separate from more general inference systems dealing with time, causality and so on.

Despite the philosophical connotations of the word 'ontology', our work in modelling an ontology for computer applications has required a practical and pragmatic approach. Although agreement on a high level schema is a prerequisite for effective co-operation between different groups of modellers, it is not practically possible either to build, or construct an implementation to support, a philosophically perfect model. What follows, therefore, is not claimed as an ultimate or perfect solution. It is, rather, the current stage in a series of successive approximations to practical tools which will:

- Fit closely enough with accepted medical usage to enable intuitive user interaction, especially in applications for clinical data entry or the retrieval and classification of medical concepts
- Be re-usable, i.e. be sufficiently general to be used by a range of different medical applications
- Support practical collaboration between different groups, being easily understood, taught, and used

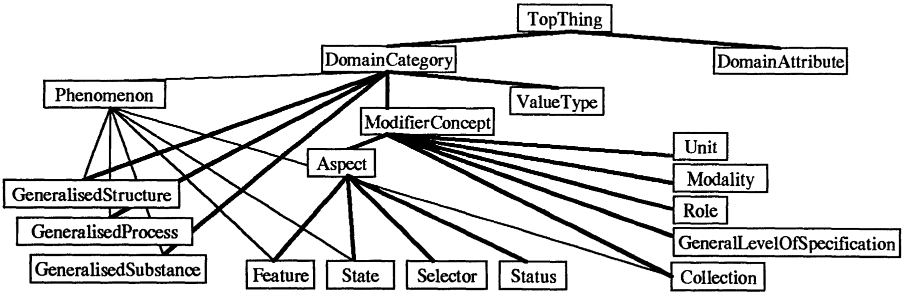

Figure 1: GALEN CORE Model top level category space

2. Summary of the GALEN Schema

2.1 Categories: Structure, Substance, Process and Modifiers

The initial division of a taxonomy - physical or abstract, structure or process etc. - is arbitrary and inevitably results in a 'tangle'. In GALEN, the primary breakdown is into:

- GeneralisedStructures — abstract or physical things with parts independent of time
- GeneralisedSubstances — continuous abstract or physical things independent of time
- GeneralisedProcesses — changes which occur over time
- Modifiers — a heterogeneous grouping broken down as shown in 2.2

A secondary structure is superimposed over the primary, aiming to capture the medical intuition of 'disease' or 'disorder'. Figure 1 shows the resulting overall schema. The category labelled *Phenomenon* represents the disjunction of those categories which can be observed. Disease, in GALEN, is defined as:

Phenomenon which hasPathologicalStatus pathological.

The choice of a label, 'Phenomenon', for this category is potentially controversial, but no better alternative has yet been proposed.

As a medically oriented ontology, GALEN further divides structures, substances, and processes into organic and inorganic or biological and non-biological (Table 1), the labels 'organic' and 'biological' co-existing for historical reasons only and being interchangeable. BodyStructure, BodySubstance and BodyProcess are used to group together the primary categories for human anatomy and patho-physiology.

Table1: Taxonomy of major elementary Categories

Entity	Example	Entity	Example
GeneralisedProcess		GeneralisedSubstance	
SpecificProcess		Energy	Radiation, SoundEnergy
BiologicalProcess		Substance	
BodyProcess	Peristalsis, Breathing, Clotting	BodySubstance	
Behaviour	VolitionalAct, ClinicalAct	Tissue	MuscleTissue, BoneTissue
NonBiologicalProcess		NAMEDBodySubstance	Urine, Bile , Sputum etc.
PhysicalProcess	Irradiation	ChemicalSubstance	Drugs, Sodium etc.
ChemicalProcess	Histological Staining	NAMEDSubstance	Air, Wood
GenericProcess	Transport, Opening, Closing		
GeneralisedStructure		ModifierConcept	
AbstractStructure		Aspect	
PsychosocialConstruct	Clinic, Hospital	Feature	Sex, Chronicity, Shape, Malignancy
LogicalStructure	Protocol, Plan	State	male/female, acute/chronic, round/square
PhysicalStructure		Selector	leftSelector/rightSelector
LinearStructure	Displacement	Status	pathological/pysiological normal/nonNormal
PlanarStructure	Triangle, Square	Collection	(Multiple which isMultipleOf Spot).=Rash
SolidStructure		Modality	FamilyHistory, presence/absence
MicroscopicStructure	Cell, Microorganism	Role	DrugRole, HormoneRole, PatientRole
InertSolidStructure	Building, Device	Unit	second, metre, kilogram
OrganicSolidStructure		LevelOfSpecification	uniquelySpecified
BodyStructure			
BodyPart	Heart,Leg, Head, Femur, AdrenalGland, Sacrum, PisiformBone, etc etc		
GenericBodyStructure	Cusp, Horn, Promontory, Artery, Lump, Bursa, Orifice, Rim, Ridge etc		
Organism	Bacterium. Protozoan, Virus, Fungus, Dog, Bird		
SolidRegion	e.g. (SolidRegion which isSolidDivisionOf Liver) name PieceOfLiver.		
Space	Cavity, Potential Space, PathologicalCavity		

Some primitives in the category hierarchy are deliberate artefacts, identifiable by the prefix 'NAMED' in their label. The child categories of such entities share a common characteristic, but are otherwise heterogeneous. The artefacts break up what would otherwise be very large flat lists, to assist the modellers in managing early models. They may also provide a useful common parent from which a single constraint may be inherited. The prefix 'NAMED' is to distinguish the artefact from the corresponding abstraction. For example, all *NAMEDPathologicalProcesses* (Trauma, Inflammation, Ulceration, Neoplasia, Degeneration etc) are *PathologicalProcesses* but not vice versa. This is expressed in GRAIL by:

(GeneralisedProcess which hasPathologicalStatus pathological) name PathologicalProcess.
NAMEDPathologicalProcess necessarily hasPathologicalStatus pathological.

2.2 Modifiers

In GALEN, the many modifiers of medical terms are divided principally between:

- Aspects which refine the meaning of a category — e.g. location, severity, shape etc. where a 'severe diabetes' is a kind of 'diabetes'
- *Modalities* (such as *FamilyHistory* or *Risk)* and *Collections* which take their meaning from the category but which are fundamentally different from the original category, e.g. 'family history of diabetes' is not a kind of 'diabetes', nor is a 'set of polyps' itself a kind of 'polyp'

In addition, GALEN recognises:

- Roles such as PatientRole, DoctorRole, HormoneRole etc. as discussed in detail in [4]
- Levels of specification whose use is discussed in [3]
- Units such as metre, kilogram, minute, etc. which are used as part of quantities

2.2.1 Aspects: Feature-State pairs, Selectors, and Statuses

Aspects generally correspond to categories used as adjectives or adjectival expressions. In GALEN, most properties which can describe phenomena are found beneath *Feature* (e.g. *Temperature*), and the corresponding values beneath *State* (e.g. *Hot/Cold*). However, many single *Features* can take several values. An example would be the body temperature of a patient, which could be raised, increasing, lower than the last reading and higher than expected (for the condition). The GALEN model must allow a *Feature* to take only one value from a mutually exclusive set of *States* (e.g. *rising, constant, falling*) but simultaneously to take multiple *States* from any number of such sets (e.g. *rising* as well as *low*). The schema for *Aspects*, therefore is:

Category-hasXFeature-Feature-hasXState-State

for example (in GRAIL notation):

Patient which
hasTemperatureFeature (Temperature which < hasAbsoluteState raised
hasTrendInState decreasing
hasQuantity (Quantity which < hasMagnitude 39.8
hasUnits degreesCentigrade>)>).

However, it should not be possible to simultaneously link to (*Temperature - high*) and (*Temperature - low*). To prevent linking to more than one kind of a single *Feature*, all attributes linking to a *Feature-State* pair - such as *hasTemperatureFeature* above - have single-valued cardinality. However, the cardinality of GRAIL attributes is constant, being determined at the point of their initial declaration. To allow linking to more than one different *Feature-State* pair (e.g to *Shape* as well as *Temperature*), a dedicated, single-valued linking attribute must exist for every *Feature*: *Frequency* may only link via *hasFrequencyFeature*, *Shape* via *hasShapeFeature* and so on. This work-around may be improved upon in the future.

By contrast, there are other modifiers which only serve to select which of two or more options pertains, *e.g.* 'left' in 'left hand'. There is nothing else to say about the 'laterality' of the hand and, hence, no reason to create a *Feature*. Such modifiers are termed *Selectors,* which are linked by a single attribute in the simpler schema:

Category which hasXSelector Selection.

for example:

Hand which hasLeftRightSelector leftSelection.

Many selectors such as left/right, upper/lower, etc. interact with the part-whole structure in important ways, *e.g.* the 'right hand' can only be part of the 'right arm'. The simpler schema above makes the implementation of these patterns more efficient using GRAIL constructs for transitivity [5].

Finally, *normal* and *nonNormal*, *physiological* and *pathological* and a few similar modifiers with special usage are placed together under *Status*. Like *Selectors*, *Statuses* are linked to their modifiers by a single attribute.

2.2.2 *Modalities: FamilyHistory, presence and absence.*

A *FamilyHistory* can be of almost any phenomenon, e.g. 'family history of diabetes', 'family history of elevated blood pressure', 'family history of tumours'. Similarly, any phenomenon can be absent or present, currently or in the past and with a variable degree of certainty. Such categories are constructed using modalities:

Modality which isModalityOf Phenomenon.

For example:

FamilyHistory which isFamilyHistoryOf Diabetes.

In addition, collections and sets are created using modalities, for example

(Multiple which isMultipleOf Spots) name Rash.

2.3 *Attribute Hierarchy*

The taxonomy of 'attributes' (or 'semantic link types') is influenced by, and supports the category taxonomy already outlined. The primary distinction is between:

- *ConstructiveAttributes* – linking processes, structures and substances together
- *ModifierAttributes* – linking processes, structures and substances to modifiers.

The modifier attributes and modifier categories are intimately tied, as discussed above. The constructive attributes are broken down initially as shown in Table 2.

Table 2: Summary of constructive attributes

Attribute	Examples
ConstructiveAttribute	
PartitiveAttribute	
StructuralPartitiveAttribute	*hasSolidDivision, hasSurfaceDivision, hasLinearDivision, hasStructuralComponent, hasLayer, hasFunctionalComponent*
ProcessPartitiveAttribute	*hasSubprocess*
StructuralAttribute	*connects, serves (e.g. Artery which serves Liver), isAnatomicallyRelatedTo*
DelimitingAttribute	*hasBranch, definesSpace, contains, passesThrough,*
LocativeAttribute	
AnatomicalLocativeAttribute	*hasLocation, isContainedIn*
FunctionalLocativeAttribute	*actsOn, isFunctionOf*
FunctionalAttribute	*actsOn, hasGoal, isFunctionOf, hasExposureTo, hasFunctionalComponent*
hasCausalLinkTo	*hasAssociation, hasConsequence*

3. Discussion and Conclusion

This paper has outlined the high level structure of the GALEN Common Reference Model as at January 1996. It shows how a basic structure of substances, processes and structures, coupled with a system of modifiers and modalities, can be used to model most medical concepts intuitively. Although heavily influenced by the UMLS Semantic Network [6], the GALEN ontology has required finer granularity and a somewhat more general taxonomy. A detailed comparison is forthcoming. The structure has evolved iteratively through experience in representing segments of coding systems and practical experience in building clinical systems within the GALEN project [7] and in a commercial collaboration. Work with coding systems has focused on SNOMED-International and ICD-9-CM, covering most of their cardiovascular and respiratory sections, plus segments including Urology, Gastroenterology and Orthopaedic Surgery. The 'Clinical Act' section is heavily influenced by RICHE/NUCLEUS [8,9] and CEN [10] models. About 6000 primitive or composite categories are present in the model, providing many points to which constraints are attached. These constraints imply an extended category space running to several billion entities, which can potentially be automatically generated by the existing implementation.

In terms of the goals identified in the introduction, the model is being used in both experimental and prototype commercial data entry systems to support intuitive clinical data entry. Re-use has been demonstrated between data entry and auto-encoding applications, and experimental work with decision support systems is underway and encouraging.

From our experience so far, we are confident of our understanding of the choices in the current model, and that we have a firm foundation for both practical and theoretical development. The need for an efficient implementation, together with the need to cope with the mismatch between human cognition and formal modelling tools, mean that the schema must continue to evolve. Above all, the high level schema must enable co-operative working if it is to be truly scaleable. The key test of this will be the GALEN-IN-USE project, begun in January 1996, which involves nine centres building or using the model.

4. References:

1. Rector, A. (1994). Compositional models of medical concepts: towards re-usable application-independent medical terminologies. in Knowledge and Decisions in Health Telematics P. Barahona and J. Christensen (ed). IOS Press. 133-142.
2. Rector, A., W. Solomon, W. Nowlan and T. Rush (1995). A Terminology Server for Medical Language and Medical Information Systems. Methods of Information in Medicine, Vol. 34, 147-157
3. Rector, A., A. Gangemi, E. Galeazzi, A. Glowinski and A. Rossi-Mori (1994). The GALEN CORE Model Schemata for Anatomy: Towards a re-usable application-independent model of medical concepts. Twelfth International Congress of the European Federation for Medical Informatics, MIE-94, Lisbon, Portugal, 229-233.
4. Rector, A. (1995). Coordinating taxonomies: Key to re-usable concept representations. Fifth conference on Artificial Intelligence in Medicine Europe (AIME '95), Pavia, Italy, Springer. 17-28.
5. Rector, A. and W.A. Nowlan (1993). 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).
6. McCray, A. and W. Hole (1990). The Scope and Structure of the First Version of the UMLS Semantic Net. Fourtheenth Symposium on Computer Applications in Primary Care (SCAMC-90), Washington DC, IEEE Computer Society. 126-130.
7. Wigertz, O., G. Hripscak, M. Shasavar, P. Bågenholm, H. Åhlfeldt and H. Gill (1994). Data-driven medical knowledge based systems based on Arden Syntax. in Knowledge and Decisions in Health Telematics P. Barahona and J. Christianson (ed). IOS Press. 126-131.
8. Nicklin, P. and B. Frandji. (1995). Act Management and Clinical Guidelines, in Health Telematics for Clinical Guidelines and Protocols, Eds. C Gordon and JP Christiansen, IOS Press pp117-124.
9. Frandji, B., J. Schot, M. Joubert, I. Soady. (1994). The RICHE Reference Architecture, in Medical Informatics Vol. 19 no. 1 Jan-Mar 1994
10. CEN PT 251:002S (1995). Structure of Coding Systems of Medical Devices. Draft ENV, CEN.