Experience building a Large, Re-usable Medical Ontology using a Description Logic with Transitivity and Concept Inclusions

Alan L. Rector and Ian R. Horrocks
Medical Informatics Group, Department of Computer Science
University of Manchester, Manchester M13 9PL, England
Tel: +44-161-275-6133 FAX: +44-161-275-6932
email rector/horrocks@cs.man.ac.uk URL http://www.cs.man.ac.uk/mig/galen

Abstract

The European GALEN project is developing terminology services based on a large, re-usable medical ontology. The ontology is being built using GRAIL, a description logic with transitivity and general concept inclusions.

Introduction

The GALEN Project and GRAIL Language

GALEN is a European Union funded programme developing terminology services based on a Common Reference Model for medical terminology. It is seen as a key element in an overall architecture for integrating clinical information systems (Rector *et al.* 1995). The project is currently in its second phase, GALEN-IN-USE, which is developing terminologies for medical procedures in cooperation with the European Federation of Classification Centres and demonstrating the GALEN technology in data entry and natural language modules for commercial clinical systems. The current basic model contains roughly 8,000 concepts; extensions for applications range up to 10,000 predefined concepts, and by the end of the project several extensions will contain over 20,000 predefined concepts.

GALEN's long term goal is a terminology server based on a re- usable medical ontology for:

- Mediation—to act as an interlingua between medical terminologies, between medical records and decision support systems, and between different database schemata.
- Authoring—to support and maintain the terminologies themselves and to use the terminologies as a common skeleton for authoring knowledge based and data entry systems.
- Human computer interaction—to support quick, intuitive data entry and query formulation by end users. Data entry has been found to be the key bottle neck in clinical systems. GALEN arose from user centred design projects on clinical data entry (Nowlan *et al.* 1991; Nowlan 1994) and its first commercial products are data entry modules for clinical systems (Kirby & Rector 1996).
- Natural Language Processing—to support semantic inference for multi-lingual language generation and analysis (Wagner *et al.* 1994; Baud *et al.* 1993; Rasinoux *et al.* 1995).
- Managing multiple viewpoints—to bridge the differences in granularity and viewpoint between the needs of direct clinical care and the abstractions required for quality assurance and decision support systems (Rector 1995).

The GALEN Terminology server architecture is based on a strict separation of Concept, Language, Coding System, and General Inference modules (Rector *et al.* 1995). This paper is concerned only with the Concept Module which manages the GALEN ontology, particularly the motivation for transitivity and concept inclusions.

Background: GRAIL and the Requirements of Medical Terminology

The Common Reference Model is formulated in the GALEN Representation and Integration Language (GRAIL) which is a description logic with support for transitivity and concept inclusions but lacking a number of common constructs (Rector *et al.* 1996). GRAIL's predecessor, SMK (Nowlan *et al.* 1991; Rector, Nowlan, & Kay 1990) was developed starting in 1988 as part of a series of investigations of human computer interaction for clinical systems and of the structure of medical records. Doyle, Patil and Haimowitz had demonstrated the limitations of highly restricted description logics as a basis for clinical systems (Haimowitz, Patil, & Szolovits 1988;

Doyle & Patil 1989; 1991), but the structure of the applications which emerged from the user centred design studies cried out for a rigorous frame-like approach. By experimentation and principled investigation of existing medical terminologies, the project developed a description logic which is extremely restricted in some respects but which includes several unusual features. It lacks detailed cardinality constraints, negation, disjunction, conjunction of primitives, a general form of restriction, and any construct such as 'SAME-AS' for creating cyclical structures. On the other hand it includes:

- A general construct for transitive roles and their interaction with related roles, marked in GRAIL by the `specialisedBy` operator (see).
- A restricted form of general concept inclusions, known in GRAIL as 'necessary statements' (see).
- The presumption of disjoint but non-exhaustive partitions in the hierarchy of primitive types.
- The use of 'sanctions' to constrain composition. Sanctions are analogous to Sowa's canonical graphs which constrain permitted graph structures. However, two levels of sanctions are provided: a high level for one for abstractions and queries, and a lower more restrictive level for statements of facts suitable for a data entry interface. The ultimate (unattainable) goal is that all and only 'sensible' medical statements are sanctioned at the lower level.

Both the restrictions and the extensions were motivated by careful empirical examination of existing medical classifications and nomenclatures. For example, despite their prominence as examples in textbooks on knowledge representation, concepts requiring detailed cardinality constraints such as 'bigamist' are rare in medical nomenclatures ($< 1\%$). On the other hand, classifications which require that 'disorders of parts' are kinds of 'disorders of the whole' are ubiquitous. By and large, the choices made have been justified by subsequent experience, although the limitations are now becoming irksome, particularly the lack of a construct equivalent to SAME-AS. The resulting system is sufficiently tractable, empirically, for the present scale (10,000-20,000 concepts) of applications (although there are some compromises so that the implementation is not complete). Fundamental research reported elsewhere aims at optimising complete algorithms based on tableaux calculus which avoid these restrictions while retaining the use of transitivity and concept inclusions and improving, if possible, the empirical tractability (Horrocks, Rector, & Goble 1996; Horrocks & Rector 1996). This paper reports primarily on the use of the constructs for medical applications and hence the motivation for the inclusion of potentially intractable features. The same constructs and representational style have also been used for describing the functioning of protein and DNA sequences (Goble, Paton, & Bechhofer 1996), images for art history, and is being investigated as the basis for representation the structure of multi-media applications (Bechhofer & Goble 1996).

Summary of GRAIL Semantics

One of the features of GRAIL which makes it more accessible to domain experts is a non-standard notation, the critical subset of which is summarised and defined in terms of the more usual description logic notation in table 1. As usual CN is a concept[1] name, C a concept term, R and S are role names and K is a cardinality keyword, one of *oneOne*, *oneMany*, *manyOne* or *manyMany*. Terminological subsumption inferences are justified by the usual Tarski style model theoretic semantics (Tarski 1956; Baader *et al.* 1991).

GRAIL Statement	**Abstract form**
C `which` $\langle \mathsf{R_1 C_1 \ldots R_n C_n} \rangle$	$\mathsf{C} \sqcap \exists \mathsf{R_1.C_1} \sqcap \ldots \ldots \sqcap \exists \mathsf{R_n.C_n}$
C `newSub` CN	$\mathsf{CN} \sqsubseteq \mathsf{C}$
C `name` CN	$\mathsf{CN} \doteq \mathsf{C}$
C `topicNecessarily` $\langle \mathsf{R_1 C_1 \ldots R_n C_n} \rangle$	$\mathsf{C} \sqsubseteq \exists \mathsf{R_1.C_1} \sqcap \ldots \ldots \sqcap \exists \mathsf{R_n.C_n}$
R `newAttribute` $\mathsf{S_1}$ $\mathsf{S_2}$ K	$\mathsf{S_1} \sqsubseteq \mathsf{R}$, $\mathsf{S_2} \doteq \mathsf{S_1^{-1}}$
R `addSub` S	$\mathsf{S} \sqsubseteq \mathsf{R}$
R `specialisedBy` S	$\mathsf{R} \circ \mathsf{S} \sqsubseteq \mathsf{R}$

Table 1: GRAIL statements and equivalent abstract forms

The GRAIL `which` statement provides a very restricted mechanism for forming concept terms, limited to a base concept conjoined to an arbitrary number of existentially quantified restrictions. GRAIL's power lies in providing axiom schemas for a restricted form of general concept inclusion and in providing an operator which allows the definition of transitively closed roles and the interaction of transitive roles with other roles. The keyword K, in the `newAttribute` statement also allows roles to be declared functional (single valued).

[1]'Concept' and 'type' are used synonymously in this paper.

Key Constructs and Usage

Modelling style and strategy

The GALEN modelling style is based on three principles aimed at maximising re-use:

- Primitive type hierarchies are separated into strict single hierarchies with only one primitive parent for each primitive type. At each level of the hierarchy the primitive types are disjoint but do not form an exhaustive partition (with well controlled exceptions at the very top and bottom of the hierarchy). All other classification is performed by the classifier based on descriptions. For example, the complex mixed hierarchy of chemical substances found in many existing terminologies is split into two clean hierarchies: one based on structure and the other on function. The composite entities constructed by combining categories from each hierarchy can be classified along either or both axes.

- Levels of granularity are bridged by general concept inclusions which treat fine details as 'understood'. For example, 'ulcers of the stomach' are understood as occurring to the 'lining of the wall of the stomach'. Hence we wish to treat 'ulcer of stomach' and 'ulcer of lining of wall of stomach' as the same concept.

- All classification must be done at the level of types. There are no supplementary rules for classifying individuals. GALEN is designed to be used to develop classifications where the type hierarchy can be separated from the storage mechanism for individuals. GALEN assumes that the historical pattern will continue whereby different groups develop ontologies (terminologies and classifications), medical record schemata and decision support systems. The demand from the medical informatics community is for common 'controlled vocabularies' (Sittig 1994) or concept systems (Evans *et al.* 1994). Implicitly the demand is that navigation and inference be at the level of the types in the classification. Therefore, a separate 'A-Box' with supplementary rules for classifying individuals is unacceptable as part of the 're-usable ontology'.

The major features of the modelling style are motivated by the desire to abide by these three principles while at the same time adhering to structures which doctors find intuitive. It is a fundamental requirement that we can reconstruct, at least to a good approximation, the structures found in the thesauri and classifications which to which doctors are accustomed.

Transitive roles and role inclusion axioms

GRAIL does not support specialised reasoning about part-whole relations (Smart & Roux 1995; Bernauer & Goldberg 1993), compositional inclusion (Padgham & Lambrix 1994) or the interaction of different compositional relations (Sattler 1995). Instead, it provides a general terminological mechanism, the 'specialised by' construct, which is similar to the assertional mechanism provided by the $\texttt{transfers} - \texttt{thro}$ construct in CycL (Lenat & Guha 1989). The 'specialised by' construct indicates that one characteristic is inherited across relations other than subsumption: R $\texttt{specialisedBy}$ S leads to the inference that for any objects x, y and z, $x\mathsf{R}y \wedge y\mathsf{S}z \Rightarrow x\mathsf{R}z$. Note that the special case where a role is specialised by itself results in a transitively closed role. Although there is no role forming operator for transitive closure, the combination of specialisation and the role hierarchy does allow transitive orbits of roles (Sattler 1996) to be defined.

The range of use of transitive roles and 'specialisation' in GALEN's ontology can be classified under two main headings, but the ramifications are surprisingly wide.

- To express partitive and locative relationships in anatomy and causal and functional relationships in patho-physiology.
- To express recursive encapsulation and constructs which are propagated along the axis of encapsulation.

Transitive roles and refinement allow the thesauri to which doctors are accustomed to be reconstructed from a principled classification using subsumption inferences. Consider the following (simplified) example:

actsOn `specialisedBy` isComponentOf

therefore

Disorder `which` actsOn Heart

subsumes

Disorder `which` actsOn (Valve `which` isComponentOf Heart)

Or more concisely for purposes of this example:

Disorder of Heart

subsumes

Disorder of Valve in Heart

Typical traditional thesauri contain hierarchies which mix subtype and part-whole roles indiscriminately because they are designed, implicitly, for a particular use. For example, a typical hierarchy mixing partitive and subtype relations as found in traditional thesauri is:

Organ
> sub-type

Heart
> part

Heart Valve
> sub-type

Aortic Valve
> part

Aortic Valve Cusp

In GRAIL, the individual types can be defined appropriately—'Valve' and 'Cusp' as kinds of 'Generic Parts'; 'Aortic Valve' as a specific kind of 'Valve' which is a component of the 'Heart'; the 'Heart' as an 'Organ'. The original use-specific hierarchy can then be reconstructed using the fact that disorders of parts are disorders of the whole:

Disorder of Organ
  Disorder of Heart
    Disorder of Valve in Heart
      Disorder of Aortic Valve in Heart
        Disorder of Cusp in Aortic Valve in Heart

However, the separate types can be re-used in other contexts in which they do not form a subtype hierarchy, *e.g.*

Surface on Heart
Surface on Valve in Heart

This style of modelling forces the modeller to make explicit the use-specific reasoning implicit in the original thesaurus structure. The modeller must separate the use-specific composites from the underlying hierarchy of primitive types. For example, since the valve of the heart is an internal component of the heart 'Cleaning the heart surface' does not subsume or include in any other way 'Cleaning of the heart valve surface'.

Concept Inclusions

Most description logics restrict concept axioms to unique and acyclic concept introductions (type definitions)—*i.e.* any given type name may appear only once on the left hand side of an axiom and the type definition may not refer, either directly or indirectly to the type being introduced. Given these restrictions, type definitions can be expanded until they contain only primitives, and the subsumption of fully expanded terms can be evaluated independently of the remainder of the model.

General concept inclusions are more general axioms of the form $\mathsf{A} \sqsubseteq \mathsf{B}$, where both A and B are arbitrary type definitions. The constraints which they put on types may lead to additional subsumption inferences, so that subsumption in a description logic which supports general concept inclusions can only be evaluated relative to the model as a whole.

Concept inclusions occur whenever it is necessary to further describe a defined type—*i.e.* whenever there are necessary conditions for a type which are not part of any set of sufficient defining conditions for its recognition. Concept inclusions are roughly equivalent to rules and are used in several different ways in the GALEN ontology:

- To propagate specific values along transitive roles; for example being 'abnormal' propagates along locative and partitive relations but being 'normal' does not—*i.e.* that the part is 'abnormal' implies that the whole is 'abnormal' but not conversely.
- To represent multiple viewpoints
  - By bridging of levels of detail, usually in conjunction with transitivity, for example, to express the fact that 'ulcers located in the lining of the wall of the stomach' are not merely a kind of 'ulcer located in the stomach' but are precisely the same thing as 'ulcer located in the stomach'.
  - By representing views not implicit in the definition, for example, many operations, procedures, and therapy are only done for specific purposes. Often the purpose and the mechanics of the procedure are used interchangeably—*e.g.* vascular shunts are inserted to 'revascularise' the affected anatomy, and any classification of 'operations for revascularisation' must retrieve vascular shunts, even though there is nothing in the literal meaning of the definition to imply this usage. Similarly 'pins' are inserted in bones to 'fixate' them, and the operations are equally likely to be described as 'fixations' or 'insertions'.
  - By representing metonymy—two definitions for the same thing. The 'sole of the foot' is the same as 'plantar surface of the foot'. (In practice the current implementation cannot express proper metonymy and must content itself with inclusion).
- To promote parsimony—for example with respect to the use of entities composed of a generic construct and a specific structure or function, *e.g.* 'the region of the heart'. The potential number of such constructs is vast. A modest number of them require special description. Defining some as primitive concepts and some as composite concepts makes the modelling style inconsistent.
- To deal with topological and spatial relationships—*e.g.* that anything which is hollow defines a space.

Within the existing model, the largest number of concept inclusions exist for managing the key medical concepts of abnormality and pathology. To provide an adequate model which satisfies our users we have had to separate the concept of what is 'noteworthy' in a medical record—whether or not it is 'normal' or 'non-Normal'—from whether something is considered in need of medical management—whether or not it is 'pathological' or 'physiological'. The next largest group concern propagation of other specific values and selectors; and the other large group within the model are those coping with the parsimonious description of entities using left, right, ordinal position, or being defined in terms of a generic construct and a specific location or function. Within the model of procedures, the ability to express implied use is considered essential and, on present experience, will rapidly grow to outweigh all other uses combined.

Experience

Knowledge Acquisition

The GALEN ontology is complex. Experience has shown that it takes three to six months for a modeller to become fully conversant with the formalism and ontology and able to model reliably. The project requires that several dozen individuals, working in different countries and with limited training, be able to compile models collaboratively. This stark contradiction is at least partly resolved by the use of an intermediate form, which drastically simplifies the ontology for the repeating patterns which are the main focus of the collaborative effort, such as surgical procedures. The overall modelling style is easier to enforce at the level of the intermediate representation without the distractions of the detailed formal model. Many issues can be dealt with in the transformation of the intermediate representation to the GRAIL formalism. Where necessary, final decisions about the best formal representation of key constructs can be deferred, awaiting further evidence, without delaying the main thrust of the collaborative effort.

Difficulties

Congenital abnormalities Congenital abnormalities can severely distort normal anatomical relations. For example, the heart may appear on the right side and all of the normal relations of the heart, heart chambers, and lungs be the mirror image of their usual position. Worse, intermediate states can present with all manner of oddities and structures which may have no counterpart in normal anatomy. On the one hand, normal use of the model for intuitive user interfaces requires that the extra complexities of congenital abnormalities do not clutter the system in routine use; on the other hand such anomalies must be provided for. Several experimental reorganisations of the ontology have been tested to resolve this problem, but so far none has proved completely satisfactory.

Contiguous structures Spatial reasoning in GRAIL is confined to part-whole relations. There are numerous situations in which this is not entirely adequate. The most common of these is the lack of support for the idea of contiguous structures, *e.g.* the stretch of the intestinal tract from the mouth to the first part of the small intestine (duodenum) inclusive. Research on integrating more spatial reasoning into the description logic framework is under way.

Lack of any SAME-AS construct Although GRAIL is able to express the idea that 'Stomach ulcers occur in the lining of the wall of the stomach', it is not able to express the more general statement that 'Ulcers in GI-Tract-Organs occur in the lining of the wall of the same organ'. This example can be partially overcome by generating all of the necessary statements semi-automatically, but this leads to difficulties in maintenance and cannot be applied in situations where the organ types are constructed compositionally, *e.g.* for left and right hands, first through ninth ribs, etc.

Connectivity A major feature of the anatomical model is that, ultimately, everything is connected to everything else in a complicated cyclical way. Although most of the connections are transitive roles, the problem of connectivity would exist even if the roles were not transitive. Potentially, any classification requires taking into account all of anatomy, even though most of it is 'obviously' irrelevant. Strategies for coping with highly connected networks are an important part of the future research programme.

Performance

Improvements in performance of hardware and translation of the underlying classification engine into C++ from the original Smalltalk have kept the performance at the upper limits of acceptable. The commercial data entry application uses extensive caching to achieve sub-one-second response for the construction of complex forms requiring numerous queries to the model. Total compilation time is now a significant problem for large models, and ranges for the 8,000-10,000 type central model from 2 hours for the fastest version to over 30 hours for the Smalltalk version which is still used to support some tools. Scaling has been roughly linear with the size of the model (Horrocks 1995). Currently, optimised tableaux calculus algorithms appear highly promising for achieving improved performance and better scaling on a much more expressive formalism (Horrocks, Rector, & Goble 1996).

Discussion and Future Developments

The GALEN ontology and GRAIL language have allowed the development of a realistic scale medical ontology which is in use in practical applications in both research and commercial settings. It is providing a means of overcoming some of the serious difficulties which plague the developers of medical classifications and nomenclatures—currently a key problem for the development of medical systems[2]. It is also proving a powerful tool for developing ontologies in non-medical fields for mediation and information retrieval.

The GRAIL language includes both general concept inclusions and transitive roles, both known to give rise to intractable systems in the worst case. Research on the circumstances contributing to worst case behaviour and optimisations is under way and reported elsewhere(Horrocks, Rector, & Goble 1996; Horrocks & Rector 1996). However, the more fundamental questions is whether or not these constructs are necessary to achieve the stated goals.

The case for transitive roles seems unanswerable for medical applications. No sensible representation of anatomy is possible which excludes transitive partitive relationships. Once the transitive closure of roles is available[3], the worst case complexity is equivalent to that for concept inclusions (Baader 1991), but concept inclusions still add a major practical burden.

There are two uses of concept inclusions which seem unavoidable given the overall goals of the project: supporting multiple viewpoints through the expression of 'implicit' information and the propagation of values along transitive roles. Other uses, such as the use for parsimonious descriptions of anatomy, *e.g.* asymmetry in 'right lung' and 'left lung', reflect a conflict between good software engineering practice—putting a given piece of information in exactly one place and maintaining consistency of style—and formal complexity. The possibility of 'compiling out' such cases between the user representation and the internal representation is being investigated. Finally, there are uses of concept inclusions, for example to express topological relationships, which appear to represent statements which are stronger than required for the applications envisaged—it would probably be sufficient to say that hollow objects may define a space (*i.e.* are 'sanctioned' to define a space) rather than that they necessarily define a space. Such statements might be usefully dropped from the ontology.

As to other issues of modelling style, the limitation of the primitive type hierarchy to a strict hierarchy of disjoint types (except for well controlled cases at the very top and very bottom) has proved highly successful in practice in enforcing constructs which lend themselves to re-use. Our experience is that the usual reason for modellers wishing to have multiple parents is that two or more notions are being conflated to satisfy a particular application. Enforcing disjoint categories forces modellers to separate the conflated notions. The effect of the predominance of disjoint types on optimisations is not yet clear. On the one hand it severely restricts the search space and provides a quick test for failure of subsumption. On the other hand, it is additional information which increases the number of possible inferences. No current implementation makes extensive use of this feature in its algorithms, so results are pending.

Finally, all of these requirements presuppose GALEN's fundamental approach to a common ontology expressed as a description logic and executable by a terminology server. Approaches which involve semi-automatic translations via an interlingua such as KIF, which is not fully executable, face related but different problems (see *e.g.* (McGuire *et al.* 1993; Fikes *et al.* 1991; Patil *et al.* 1992; Fikes, Farquhar, & Pratt 1996; Farquhar *et al.* 1996)).

There remain many practical issues in representing medical diseases, procedures, and actions. Current research is proceeding in two distinct streams—one extending the existing models and applications using the existing formalism, the other developing the algorithms and structure for a successor formalism which satisfies the requirements of those tasks and applications more effectively.

Acknowledgements

This work supported in part by the European Commission, grants AIM 2012 and HcT 1018 and by the EPSRC grant GR/J98820 (PAEPR).

The GALEN-IN-USE consortium includes: University of Manchester (Coordinator), VAMP Health Ltd, UK; National Research Council, Italy; University of Saint Etienne, France; University of Nijmegen, Dutch National Classification Centre (WCC), Netherlands; European Federation of Classification Centres; Hôpital Universitaire de Geneva, LNAT Associates, Switzerland; Technical Research Centre of Finland (VTT), Oulu University Hospital, Medici Data Ltd, National Research & Development Centre for Welfare

[2] The development of a 'common controlled vocabulary' has been put forward as the number one 'grand challenge' of Medical Informatics (Sittig 1994).

[3] Or even the slightly less expressive transitive orbits of roles (Sattler 1996).

& Health, Finland; Swedish Institute for Health Services Development (SPRI), University of Linköping, Sweden; Office Line Engineering, RAMIT vzw, Datasoft, University of Louvain, Belgium; GSF Medis Institut, University of Hildesheim, ID Gesellschaft für Information und Dokumentation im Gesundheitswesen mbH, Germany; University of Athens, Greece; IASIST, Barcelona, Spain. Hewlett Packard and the Bavarian State Government are sponsoring partners.

References

Baader, F.; Heinsohn, H.-J.; Hollunder, B.; Muller, J.; Nebel, B.; Nutt, W.; and Profitlich, H.-J. 1991. Terminological knowledge representation: A proposal for a terminological logic. Technical Memo TM-90-04, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI).

Baader, F. 1991. Augmenting concept languages by transitive closure of roles: An alternative to terminological cycles. In *Proceedings of IJCAI'91*, 446–451.

Baud, R.; Lovis, C.; Alpay, L.; Rassinoux, A.-M.; Scherrer, J.-R.; Nowlan, A. W.; and Rector, A. L. 1993. Modelling for natural language understanding. In Safran, C., ed., *Seventeenth Annual Symposium on Computer Applications in Medical Care (SCAMC-93)*, 289–93. Washington DC: McGraw Hill.

Bechhofer, S., and Goble, C. 1996. Description logics and multimedia—applying lessons learnt from the GALEN project. In *Proceedings of the workshop on Knowledge Representation for Interactive Multimedia Systems (KRIMS'96), at ECAI'96*. To appear.

Bernauer, J., and Goldberg, H. 1993. Compositional classification based on conceptual graphs. *Artificial Intelligence in Medicine Europe (AIME-93)*.

Doyle, J., and Patil, R. S. 1989. Two dogmas of knowledge representation: language restrictions, taxonomic classification, and the utility of representation services. Technical Report MIT/LCS/TM-387, Massachusetts Institute of Technology.

Doyle, J., and Patil, R. S. 1991. Two theses of knowledge representation: Language restrictions, taxonomic classification and the utility of representation services. *Artificial Intelligence* 48:261–297.

Evans, D. A.; Cimino, J.; Hersh, W. R.; Huff, S. M.; Bell, D. S.; and The Canon Group. 1994. Position statement: Towards a medical concept representation language. *Journal of the American Medical Informatics Association* 1(3):207–217.

Farquhar, A.; Fikes, R.; Pratt, W.; and Rice, J. 1996. Collaborative ontology construction for information integration. Technical Report KSL-95-63, Knowledge Systems Laboratory, Department of Computer Science, Stanford University.

Fikes, R.; Cutkosky, M.; Gruber, T.; and Van Baalen, J. 1991. Knowledge sharing technology – project overview. Technical Report KSL-91-71, Knowledge Systems Laboratory, Department of Computer Science, Stanford University.

Fikes, R.; Farquhar, A.; and Pratt, W. 1996. Information brokers: Gathering information from heterogeneous information sources. In Stewman, J. H., ed., *Ninth Florida Artificial Intelligence Research Symposium (FLAIRS-96)*, 192–197.

Goble, C.; Paton, N.; and Bechhofer, S. 1996. http://www.cs.man.ac.uk/mig/tambis/.

Haimowitz, I. J.; Patil, S. P.; and Szolovits, P. 1988. Representing medical knowledge in a terminological language is difficult. In *Twelfth Annual Symposium on Computer Applications in Medical Care (SCAMC-88)*, 101–105. Washington, DC: Computer Society Press.

Horrocks, I., and Rector, A. 1996. Using a description logic with concept inclusions. In Padgham, L.; Franconi, E.; Gehrke, M.; McGuinness, D. L.; and Patel-Schneider, P. F., eds., *Collected Papers from the International Description Logics Workshop (DL'96)*, number WS-96-05 in AAAI Technical Report, 132–135. AAAI Press, Menlo Park, California.

Horrocks, I.; Rector, A.; and Goble, C. 1996. A description logic based schema for the classification of medical data. In Baader, F.; Buchheit, M.; Jeusfeld, M.; and Nutt, W., eds., *Reasoning about structured objects—knowledge representation meets databases. Proceedings of the 3rd Workshop KRDB'96*, 24–28.

Horrocks, I. 1995. A comparison of two terminological knowledge representation systems. Master's thesis, University of Manchester.

Kirby, J., and Rector, A. L. 1996. The PEN&PAD data entry system: From prototype to practical system. In *AMIA Fall Symposium*, in press.

Lenat, D. B., and Guha, R. V. 1989. *Building Large Knowledge-Based Systems: Representation and inference in the Cyc Project*. Reading, MA: Addison-Wesley.

McGuire, J. G.; Kuokka, D.; Weber, J. C.; Tenenbaum, J. M.; Gruber, T. R.; and Olsen, G. R. 1993. SHADE: Technology for knowledge based collaborative engineering. *Jounral of Concurrent Engineering: Applications and Research (CERA)* 1(2).

Nowlan, W. A.; Rector, A. L.; Kay, S.; Horan, B.; and Wilson, A. 1991. 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. D., ed., *Fifteenth Annual Symposium on Computer Applications in Medical Care. SCAMC-91*, 855–857. Washington DC: McGraw- Hill, Inc.

Nowlan, W. A. 1994. Clinical workstation: Identifying clinical requirements and understanding clinical information. *International Journal of Bio-Medical Computing* 34:85–94.

Padgham, L., and Lambrix, P. 1994. A framework for part–of hierarchies in terminological logics. In Doyle, J.; Sandewall, E.; and Torasso, P., eds., *Principals of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference (KR'94)*, 485–496. Morgan-Kaufmann.

Patil, R. S.; Fikes, R. E.; Patel-Schneider, P. F.; McKay, D.; Finin, T.; Gruber, T.; and Neches, R. 1992. The DARPA knowledge sharing effort: Progress report. In Nebel, B.; Rich, C.; and Swartout, W., eds., *Principals of Knowledge Representation and Reasoning: Proceedings of the Third International Conference (KR'92)*, 777–788. Morgan-Kaufmann.

Rasinoux, A.-M.; Juge, C.; Michel, P.-A.; Baud, R. H.; Lemaitre, D.; Jean, F.-C.; Degoulet, P.; and Scherrer, J.-R. 1995. Analysis of medical jargon: The RECIT system. In Barahona, P.; Stefanelli, M.; Wyatt, J.; and Sickmann, J., eds., *Fifth conference on Artificial Intelligence in Medicine Europe (AIME '95)*, number 934 in Lecture Notes in Artificial Intelligence, 42–42. Pavia, Italy: Springer.

Rector, A. L.; Solomon, W. D.; Nowlan, W. A.; and Rush, T. 1995. A terminology server for medical language and medical information systems. *Methods of Information in Medicine* 34:147–157.

Rector, A. L.; Bechhofer, S.; Goble, C. A.; Horrocks, I.; Nowlan, W. A.; and Solomon, W. D. 1996. The GRAIL concept modelling language for medical terminology. *Artificial Intelligence in Medicine.* To appear.

Rector, A. L.; Nowlan, W. A.; and Kay, S. 1990. Unifying medical information using an architecture based on descriptions. In Miller, R. A., ed., *4th Annual Symposium on Computer Applications in Medical Care , SCAMC-90*, 190– 194. Washington DC: IEEE Computer Society Press.

Rector, A. L. 1995. Coordinating taxonomies: Key to re-usable concept representations. In Barahona, P.; Stefanelli, M.; Wyatt, J.; and Sickmann, J., eds., *Fifth conference on Artificial Intelligence in Medicine Europe (AIME '95)*, number 934 in Lecture Notes in Artificial Intelligence, 17–28. Pavia, Italy: Springer.

Sattler, U. 1995. A concept language for engineering applications with part–whole relations. In *Proceedings of the International Conference on Description Logics—DL'95*, 119–123.

Sattler, U. 1996. The complexity of concept languages with different kinds of transitive roles. In Görz, G., and Hölldobler, S., eds., *20. Deutsche Jahrestagung für Künstliche Intelligenz*, number 1137 in Lecture Notes in Artificial Intelligence. Springer Verlag.

Sittig, D. F. 1994. Grand challenges in medical informatics. *Journal of the American Medical Informatics Association* 1:412–413.

Smart, J. F., and Roux, M. 1995. Medical knowledge representation for medical report analysis. In Barahona, P.; Stefanelli, M.; Wyatt, J.; and Sickmann, J., eds., *Fifth conference on Artificial Intelligence in Medicine Europe (AIME '95)*, number 934 in Lecture Notes in Artificial Intelligence, 53–66. Pavia, Italy: Springer.

Tarski, A. 1956. *Logic, Semantics, Mathemetics: Papers from 1923 to 1938.* Oxford University Press.

Wagner, J. C.; Rassinoux, A.-M.; Baud, R. H.; and Scherrer, J.-R. 1994. Generating noun phrases from a medical knowledge representation. *Twelfth International Congress of the European Federation for Medical Informatics, MIE-94* 218–223.