Boosting the Medical Knowledge Infrastructure — A Feasibility Study on Very Large Terminological Knowledge Bases

Udo Hahn
Text Knowledge Engineering Lab
Albert-Ludwigs-Universität Freiburg
D-79085 Freiburg, Germany
hahn@coling.uni-freiburg.de

Stefan Schulz
Medical Informatics Department
Universitätsklinikum Freiburg
D-79104 Freiburg, Germany
stschulz@uni-freiburg.de

ABSTRACT

We conducted a feasibility study in which conceptual knowledge was extracted from an informal medical thesaurus (UMLS) and automatically converted into a formally sound description logics system. Our approach consists of four steps: concept definitions are automatically generated from the UMLS source, integrity checking of taxonomic and partonomic hierarchies is performed by the terminological classifier, cycles and inconsistencies are eliminated, and incremental refinement of the evolving knowledge base is performed by a domain expert. We report on experiments with a very large terminological knowledge base composed of 164,000 concepts and 76,000 relations.

1. INTRODUCTION

Over several decades, an enormous body of medical knowledge, e.g, disease taxonomies, medical procedures, anatomical terms etc., has been assembled in a wide variety of medical terminologies, thesauri and classification systems. The conceptual structuring of a domain they allow is typically restricted to the provision of broader and narrower terms, related terms or (quasi-)synonymous terms. This is most evident in the UMLS, the *Unified Medical Language System* [10], an umbrella system which covers more than 60 medical thesauri and classifications. Its metathesaurus component contains more than 800,000 concepts which are structured in hierarchies and classified by 134 semantic types (provided by the UMLS *semantic network*). Their semantics is shallow and entirely intuitive, which is due to the fact that their usage was primarily intended for humans as a backbone for various forms of clinical knowledge management, e.g., cross-mapping between different terminologies, medical information retrieval, disease and procedure encoding, etc.

Given the size, the evolutionary diversity and inherent heterogeneity of the UMLS, it is no surprise that the lack of a formal semantic foundation leads to inconsistencies, circular definitions, etc. [2]. This may not cause utterly severe problems when humans are in the loop, but anticipating its use for more knowledge-intensive applications such as natural language understanding of medical narratives (e.g.,discharge summaries,admission or X-ray reports), medical decision support systems, etc., those shortcomings might lead to an impasse.

As a consequence, formal models for dealing with medical knowledge have been proposed, using representation mechanisms based on conceptual graphs, semantic networks or description logics [3, 9, 12, 17, 5]. No doubt, there is a price to be paid for more expressiveness and formal rigor, *viz.* increasing modeling efforts and, hence, increasing maintenance costs. Operational systems making full use of this rigid approach, especially those which employ high-end knowledge representation languages, are usually restricted to rather small subdomains. The most comprehensive of these sources we know of is the GRAIL-encoded GALEN knowledge base which covers up to 9,800 concepts [12]. The small coverage then hampers their routine usage, an issue which is always highly rewarded in the medical informatics community.

Almost all of the knowledge bases developed on the basis of formal representation languages have been designed from scratch – without making systematic use of the large body of knowledge contained in widely spread medical terminologies. Hence, it would be an intriguing approach to join the *massive coverage* offered by informal medical terminologies with the high level of *expressiveness and deductive reasoning* capabilities supported by state-of-the-art knowledge representation systems in order to develop formally solid medical knowledge bases on a larger scale. This idea has already been fostered by Pisanelli et al. [11] who extracted knowledge from the UMLS SN and from parts of the Metathesaurus, and merged them with logic-based top-level ontologies from various sources. Another example is the re-engineering of SNOMED [4] from a multi-axial coding system into a formally well-founded ontology [16]. These efforts, however, are entirely focused on generalization-based reasoning along taxonomies and lack the coverage of partonomies, another crucial part of medical knowledge.

2. PART-WHOLE REASONING

As far as medical knowledge is concerned, two main hierarchy-building relationships can be distinguished, *viz.* is-a (taxonomic) and part-whole (partonomic) relations. Hence, the need arises to have formally solid inference mechanisms for taxonomic (generalization hierarchies), as well as partonomic reasoning (part-whole hierarchies) available within a uniform representation model. Second, it is not sufficient for us to have a sound formal platform of knowledge representation, but we also require an inference engine which performs this style of advanced reasoning on large data sets

.

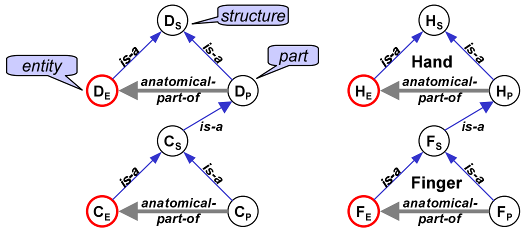

Figure 1: SEP Triplets: Partitive Relations within Taxonomies

($\gg$ 10,000 items). Therefore, we consider descriptions logics [1], at the formal representation level, and LOOM's classification-based inference machine [7, 8], at the system level, the most convenient match of our requirements and the current state of the art.

Motivated by informal approaches sketched by Schmolze & Mark [14] and Schulz *et al.* [15], we formalized a model of part-whole reasoning [6] that incorporates the above requirements and also does not exceed the expressiveness of the well-understood, parsimonious concept language $\mathcal{ALC}$ [13].[1]

Our proposal is centered around a particular data structure, so-called *SEP triplets*, especially designed for part-whole reasoning (cf. Figure 1). They define a characteristic pattern of IS-A hierarchies which support the emulation of inferences typical of transitive PART-OF relations. In this formalism, the relation ANATOMICAL-PART-OF describes the partitive relation between physical parts of an organism.

A triplet consists, first of all, of a composite '**s**tructure' concept, the so-called **S-node** (e.g., HAND-STRUCTURE or HAND$_S$). Each *structure* concept subsumes both an anatomical *entity* and each of the anatomical *part*s of this entity. Unlike entities and their parts, structures have no physical correlate in the real world — they constitute a representational artifact required for the formal reconstruction of systematic patterns of part-whole reasoning. The two direct subsumees of an S-node are the corresponding **E-node** ('**e**ntity') and **P-node** ('**p**art'), e.g., HAND$_E$ and HAND$_P$, respectively. Unlike an **S-node**, these nodes refer to specific ontological objects. The E-node denotes the whole anatomical entity to be modeled, whereas the P-node is the common subsumer of those concepts that have their role ANATOMICAL-PART-OF filled by the corresponding E-node concept. A reconstruction of some basic anatomical relations in terms of SEP triplets is illustrated in Figure 2. As an example, with partonomic reasoning we may infer that a THUMB-NAIL (TN$_S$) *is-a* THUMB-PART (T$_P$) which *is-a* FINGER-PART (F$_P$). Alternatively, taxonomic reasoning allows us to infer that a THUMBNAIL (TN$_S$) *is-a* FINGERNAIL (FN$_S$) which *is-a* NAIL (N$_S$), or a FINGERNAIL (FN$_S$) *is-a* FINGER-PART (F$_P$) which, via F$_S$, *is-a* HAND-PART (H$_P$).

The reconstruction of the relation ANATOMICAL-PART-OF by taxonomic reasoning proceeds as follows. Let us assume that C_E and D_E denote E-nodes, C_S and D_S denote the S-nodes that subsume C_E and D_E, respectively, and C_P and D_P denote the P-nodes related to C_E and D_E, respectively, via the role ANATOMICAL-PART-OF (cf. Figure 1). These conventions can be captured by the

[1] $\mathcal{ALC}$ allows for the construction of hierarchies of concepts and relations, where $\sqsubseteq$ denotes subsumption and $\doteq$ definitional equivalence. Existential ($\exists$) and universal ($\forall$) quantification, negation ($\neg$), disjunction ($\sqcap$) and conjunction ($\sqcup$) are supported. Role filler constraints (e.g., typing by C) are linked to the relation name R by a dot, $\exists R.C$.

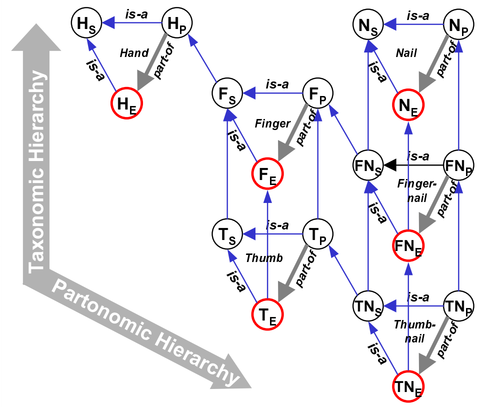

Figure 2: SEP Triplet Model of a Partonomic Hierarchy of the Concept HAND

following terminological expressions:

$$C_E \sqsubseteq C_S \sqsubseteq D_P \sqsubseteq D_S \quad (1)$$

$$D_E \sqsubseteq D_S \quad (2)$$

The P-node is defined as follows (note the disjointness between D_E and D_P, i.e., no anatomical concept can be ANATOMICAL-PART-OF itself):

$$D_P \doteq D_S \sqcap \neg D_E \sqcap \exists anatomical\text{-}part\text{-}of.D_E \quad (3)$$

Since C_E is subsumed by D_P (according to (1)), we infer that the relation ANATOMICAL-PART-OF holds between C_E and D_E, too:

$$C_E \sqsubseteq \exists anatomical\text{-}part\text{-}of.D_E \quad (4)$$

3. KNOWLEDGE ENGINEERING

Our goal is to extract conceptual knowledge from two relevant subdomains of the UMLS, *viz.* anatomy and pathology, in order to construct a formally sound knowledge base using a terminological knowledge representation language. This task can be divided into four steps: (1) the automated generation of terminological expressions, (2) their submission to a terminological classifier for consistency checking, (3) the manual restitution of formal consistency in case of inconsistencies, and, finally, (4) the manual rectification and refinement of the formal representation structures. These four steps are illustrated by the workflow diagram depicted in Figure 3.

Step 1: Automated Generation of Terminological Expressions. Sources for concepts and relations are the UMLS semantic network and the *mrrel*, *mrcon* and *mrsty* tables of the 1999 release of the UMLS metathesaurus. The *mrrel* table which contains approximately 7,5 million records (for a fragment, cf. Figure 4) exhibits the semantic links between two UMLS CUIs (concept unique identifier),[2] the *mrcon* table contains the concept names and *mrsty* keeps the semantic type(s) assigned to each CUI. These tables,

[2] As a convention in UMLS, any two CUIs must be connected by at least a shallow relation (in Figure 4, CHilD relations in the column REL are assumed between CUIs). Shallow relations may be refined in the column RELA, if a thesaurus is available which contains more specific information. Some CUIs are linked either by *part-of* or *is-a*. In any case, the source thesaurus for the relations and the CUIs involved is specified in the columns X and Y (e.g., MeSH 1999, SNOMED International 1998).

		Step 1	Step 2	Step 3	Step 4
UMLS relation	**number of links**	**Automatic generation of Loom definitions, augmented by P-Loom language elements** ***;;; = comment line***	**Submission to Loom classifier.** **Validation for formal consistency by Loom**	**Manual restitution of formal consistency**	**Manual rectification and refinement of the resulting knowledge base**
Anatomy Concepts Linked to Anatomy Concepts					
sibling_of	267.218	*;;; SIB*			add negations in order to express taxonomic or partitive disjointness
child_of	59.808	*;;; CHDRN*			include related concepts into :is-primitive or :part-of clause where plausible
narrower_term	24.223	*;;; CHDRN*			
isa	9.755	:is-primitive	check for definitional cycles	remove taxonomic parent concepts	substitute of primitive links by non-primitive ones where possible
location_of	4.803	*;;; LOCATION_OF*			include related concepts into :has-part clause where plausible
has_location	4.803	*;;; HAS_LOCATION*			include related concepts into :part-of clause, where plausible
has_part	4.321	:has-part			check whether this part is mandatory (under "real-anatomy" assumption)
has_conceptual_part	126				
part_of	4.321	:part-of	1. check for partonomic cycles 2. check for disjointness between E and P node	1. remove partonomic or taxonomic parent concepts 2. redefine triplet as single concept	check for plausibility and completeness
conceptual_part_of	126				
parent	59.808	*;;; PARRB*			include related concepts into :has-part clause where plausible
broader_term	24.223	*;;; PARRB*			
inverse_isa	9,755	<do nothing>			
associated_with	14				
mapped_from	2.643				
other_relation	10.908				
qualified_by	1.864				
allowed_ qualifier	1.864				
mapped_to	2643				
<other named relations>	11.886	(:some x)	check for inherited constraints	remove constraints	remove or add constraints
Pathology Concepts Linked to Pathology Concepts					
sibling_of	457.542	;;; SIB			add negations in order to express taxonomic disjointness
child_of	72.426	:is-primitive	check for definitional cycles	remove parent concepts	substitute primitive links by non-primitive ones whenever possible
narrower_term	26.972				
isa	3.635				
inverse_isa	3.635	<do nothing>			
associated_with	13.902				
mapped_to	15.024				
mapped_from	15.024				
part_of	1				
has_part	1				
parent	72.426				
broader_term	28.972				
other_relation	25.796				
qualified_by	6.255				
allowed_qualifier	6.255				
<other named relations>	4.162	(:some x)	check for inherited constraints	remove constraints	remove or add constraints
Pathology Concepts Linked to Anatomy Concepts					
CUIpat = CUIana	2.247	(:some has_anatomic_correlate)			plausibility check of concept "duplication" (assignment to both domains)
<missing>		<do nothing>			add pathology-anatomy links
associated_with	2.314	(:some associated_with <anatomy_concept>_S)		check for consistency	render links complete, link to E-node instead of S-node when role propagation has to be disabled
has_location	9,230	(:some has_location <anatomy_concept>_S)			
<other>		<do nothing>			

Figure 3: **Workflow Diagram for the Construction of a LOOM Knowledge Base from the UMLS**

CUI1	REL	CUI2	RELA	x	y
C0005847	CHD	C0014261	part_of	MSH99	MSH99
C0005847	CHD	C0014261		CSP98	CSP98
C0005847	CHD	C0025962	isa	MSH99	MSH99
C0005847	CHD	C0026844	part_of	MSH99	MSH99
C0005847	CHD	C0026844		CSP98	CSP98
C0005847	CHD	C0034052		SNMI98	SNMI98
C0005847	CHD	C0035330	isa	MSH99	MSH99
C0005847	CHD	C0042366	part_of	MSH99	MSH99
C0005847	CHD	C0042367	part_of	MSH99	MSH99
C0005847	CHD	C0042367		SNM2	SNM2
C0005847	CHD	C0042449	isa	MSH99	MSH99

Figure 4: Semantic Relations in the UMLS Metathesaurus

available as ASCII files, were imported into a Microsoft Access relational database and manipulated using SQL embedded in the VBA programming language. For each CUI in the *mrrel* subset its alphanumeric code was substituted by the English preferred term found in *mrcon*.

After a manual remodeling of the 135 top-level concepts and 247 relations of the UMLS semantic network, we extracted, from a total of 85,899 concepts, 38,059 anatomy and 50,087 pathology concepts from the metathesaurus. The criterion for the inclusion into one of these sets is the assignment to predefined semantic types. Also, 2,247 concepts appeared in both sets, anatomy and pathology. Since we wanted to keep the two subdomains strictly disjoint, we maintained these 2,247 concepts duplicated, and prefixed all concepts by ANA- or PAT- according to their respective subdomain. This can be justified by the observation that these hybrid concepts exhibit, indeed, multiple meanings. For instance, TUMOR has the meaning of a malignant disease on the one hand, and of an anatomical structure on the other hand. The same applies to congenital and acquired malformations, e.g., *claw foot*, etc.

As target structures for the anatomy domain we chose SEP triplets. These are expressed in the terminological language LOOM which we had previously extended by a special DEFTRIPLET macro (cf. Table 1 for an example). Only UMLS *part-of*, *has-part* and *is-a* relation attributes are considered for the construction of taxonomic and partonomic hierarchies (cf. Figure 3). Hence, for each anatomy concept, one SEP triplet is created. The result is a mixed IS-A and PART-WHOLE hierarchy a straightforward example of which is depicted in Figure 2.

For the pathology domain, we treated *CHD* (child) and *RN* (narrower relation) from the UMLS as indicating taxonomic links. No part-whole relations were considered, since this category does not apply to the pathology domain.

As a fundamental semantic assumption all roles generated in this process were considered as existentially quantified. This means that any relation r (PART-OF, HAS-LOCATION,..) which holds between two concepts A and B is mapped to a role $R.B$ which is a

```
(deftriplet HEART
   :is-primitive HOLLOW-VISCUS
   :has-part (:p-and
      FIBROUS-SKELETON-OF-HEART
      WALL-OF-HEART
      CAVITY-OF-HEART
      CARDIAC-CHAMBER-NOS
      LEFT-SIDE-OF-HEART
      RIGHT-SIDE-OF-HEART
      AORTIC-VALVE
      PULMONARY-VALVE ))
```

Table 1: Generated Triplets in LOOM Format

necessary condition in the definition of the concept A. All conceptual constraints for a concept definition are mapped to a conjunction of constraints.

In both subdomains, shallow relations, such as the extremely frequent sibling *SIB* relation, were included as comments into the code to give some heuristic guidance for the manual refinement phase.

Step 2: Automatic Consistency Checking by the LOOM Classifier. The import of UMLS anatomy concepts resulted in 38,059 DEFTRIPLET expressions for anatomical concepts and 50,087 DEFCONCEPT expressions for pathological concepts. Each DEFTRIPLET was expanded into three DEFCONCEPT (S-, E-, and P-nodes), and two DEFRELATION (ANATOMICAL-PART-OF-X, INV-ANATOMICAL-PART-OF-X) expressions, summing up to 114,177 concepts. This yielded (together with the concepts from the UMLS semantic network) a total of 240,764 definitory LOOM expressions.

From 38,059 anatomy triplets, 1,219 DEFTRIPLET statements exhibited a :HAS-PART clause followed by a list of a variable number of triplets, containing more than one argument in 823 cases (average cardinality: 3.3). 4,043 DEFTRIPLET statements contained a :PART-OF clause, only in 332 cases followed by more than one argument (average cardinality: 1.1). The resulting knowledge base was then submitted to the terminological classifier and checked for terminological cycles and coherence. In the anatomy subdomain, one terminological cycle and 2,328 incoherent concepts were found, in the pathology subdomain 355 terminological cycles though not a single incoherent concept were determined (cf. Table 2).

	Anatomy	Pathology
Triplets	38,059	—
defconcept statements	114,177	50,087
cycles	1	355
inconsistencies	2,328	0

Table 2: Classification Results for the Concept Import

Step 3: Manual Restitution of Consistency. The inconsistencies in the anatomy part of the knowledge base identified by the classifier could all be traced back to the simultaneous linkage of two triplets by both *is-a* and *part-of* links, an encoding that raises a conflict due to the disjointness required for corresponding P- and E-nodes. In most of these cases the affected parents belonged to a class of concepts that obviously cannot be appropriately modeled as SEP triplets, e.g., SUBDIVISION-OF-ASCENDING-AORTA or ORGAN-PART. The meaning of each of these concepts almost paraphrases that of a P-node, so that in these cases the violation of the SEP-internal disjointness condition could be accounted for by substituting those triplets with simple LOOM concepts, by matching them with already existing P-nodes or by disabling IS-A or PART-OF links.

In the pathology part of the knowledge base, we expected a large number of terminological cycles, as a consequence of interpreting the thesaurus-style *narrower term* and *child* relations through taxonomic subsumption (IS-A). Bearing in mind the size of the knowledge base, we consider 355 cycles a decent number. Those cycles were primarily due to very similar concepts, e.g., ARTERIOSCLEROSIS *vs.* ATHEROSCLEROSIS, AMAUROSIS *vs.* BLINDNESS, and residual categories ("other", "NOS" = *not otherwise specified*). These were directly inherited from the source terminologies and are notoriously difficult to interpret out of their definitional context, e.g., OTHER-MALIGNANT-NEOPLASM-OF-SKIN *vs.* MALIGNANT-NEOPLASM-OF-SKIN-NOS.

The cycles were analyzed and a negative list which consisted of 630 concept pairs was manually derived. In a subsequent extraction cycle we incorporated this list in the automated construction of the LOOM concept definitions, and given these new constraints, a fully consistent knowledge base was generated.

Step 4: Manual Rectification and Refinement of the Knowledge Base. To set up this high-volume knowledge base including the aforementioned working steps required three months of work for a single person, in total. The fourth step – when performed for the whole knowledge base – is very time-consuming and requires broad and in-depth medical expertise. An analysis of random samples from both subdomains is currently being performed by the second author, a domain expert. From the experience we gained in the anatomy and pathology subdomains so far, the following workflow can be derived:

- *Checking the correctness of both the taxonomic and partitive hierarchies.* Taxonomic and partitive links are manually added or removed. Primitive subsumption is substituted by non-primitive subsumption whenever possible. This is a crucial point, because the automatically generated hierarchies contain only information about the parent concepts and necessary conditions. As an example, the automatically generated definition of DERMATITIS includes the information that it is an INFLAMMATION, and that the role HAS-LOCATION must be filled by the concept SKIN. An INFLAMMATION that HAS-LOCATION SKIN, however, cannot automatically be classified as DERMATITIS.

- *Check of the* :has-part *arguments assuming 'real anatomy'.* In the UMLS sources *part-of* and *has-part* relations are considered as symmetric. According to our transformation rules, the attachment of a role HAS-ANATOMICAL-PART to an E-node B_E, with its range restricted to A_E implies the existence of a concept *A* for the definition of a concept *B*. On the other hand, the classification of A_E as being subsumed by the P-node B_P, the latter being defined via the role ANATOMICAL-PART-OF restricted to B_E, implies the existence of B_E given the existence of A_E. These constraints do not always conform to 'real' anatomy, i.e., anatomical concepts that may exhibit pathological modifications or result from surgical interventions, e.g., a large intestine without an appendix, or an oral cavity without teeth.

- *Analysis of the sibling relations and defining concepts as being disjoint.* In UMLS, *SIB* relates concepts that share the same parent in a taxonomic or partonomic hierarchy. Pairs of sibling concepts may have common descendants or not. If not, they constitute the root of two disjoint subtrees. In a taxonomic hierarchy, this means that one concept implies the negation of the other (e.g., a benignant tumor cannot be a malignant one, *et vice versa*). In a partitive hierarchy, this can be interpreted as *spatial* disjointness, *viz.* one concept does not spatially overlap with another one. As an example, ESOPHAGUS and DUODENUM are spatially disjoint, whereas STOMACH and DUODENUM are not (they share a common transition structure, called PYLORUS), such as all neighbor structures that have a surface or region in common.

- *Completion and modification of anatomy–pathology relations.* Surprisingly, only few pathology concepts contained an explicit reference to a corresponding anatomy concept. These relations must, therefore, be added by a domain expert.

4. CONCLUSIONS

There is a growing demand for high-quality terminology services and their embedding in functionally advanced health information systems. Instead of developing sophisticated medical knowledge bases from scratch, we here propose a 'conservative' approach — reuse existing large-scale resources, but refine the data from these resources so that advanced representational requirements imposed by more expressive knowledge representation languages are met. The resulting knowledge bases can then be used for sophisticated applications requiring formally sound medical reasoning.

The knowledge engineering methodology we have proposed in this paper does exactly this. It provides a formally solid description logics framework with a modeling extension by SEP triplets so that both taxonomic and partonomic reasoning are supported equally well. While purely automatic conversion from semi-formal to formal environments causes problems of adequacy of the emerging representation structures, the refinement methodology we propose already inherits its power from the terminological reasoning framework. In our concrete work, we found the implications of using the terminological classifier, the inference engine which computes subsumption relations, of utmost importance and of outstanding heuristic value. Hence, the knowledge refinement cycles are truly semi-automatic, fed by medical expertise on the side of the human knowledge engineer, but also driven by the terminological reasoning system which makes explicit the consequences of (im)proper concept definitions.

5. REFERENCES

[1] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider, editors. *The Description Logic Handbook. Theory, Implementation and Applications*. Cambridge, U.K.: Cambridge University Press, 2003.

[2] James J. Cimino. Auditing the Unified Medical Language System with semantic methods. *Journal of the American Medical Informatics Association*, 5(1):41–45, 1998.

[3] James J. Cimino, Paul D. Clayton, George Hripsack, and Stephen B. Johnson. Knowledge-based approaches to the maintenance of a large controlled medical terminology. *Journal of the American Medical Informatics Association*, 1(1):35–50, 1994.

[4] Roger Côté, David J. Rothwell, Ronald S. Beckett, James L. Palotay, and Louise Brochu. *The Systemised Nomenclature of Medicine:* SNOMED *International*. Northfield, IL: College of American Pathologists, 1993.

[5] Aldo Gangemi, Domenico M. Pisanelli, and Geri Steve. An overview of the ONION project: Applying ontologies to the integration of medical terminologies. *Data & Knowledge Engineering*, 31(2):183–220, 1999.

[6] Udo Hahn, Stefan Schulz, and Martin Romacker. Part-whole reasoning: A case study in medical ontology engineering. *IEEE Intelligent Systems & Their Applications*, 14(5):59–67, 1999.

[7] Robert MacGregor and Raymond Bates. The LOOM knowledge representation language. Technical Report RS-87-188, Information Sciences Institute, University of Southern California, 1987.

[8] Robert M. MacGregor. A description classifier for the predicate calculus. In *AAAI'94 – Proceedings of the 12th National Conference on Artificial Intelligence*, volume 1, pages 213–220. Seattle, WA, USA, July 31 - August 4, 1994. Menlo Park, CA: AAAI Press & MIT Press, 1994.

[9] Eric Mays, Robert Weida, Robert Dionne, Meir Laker, Brian White, Chihong Liang, and Frank J. Oles. Scalable and expressive medical terminologies. In J. J. Cimino, editor, *AMIA'96 – Proceedings of the 1996 AMIA Annual Fall Symposium (formerly SCAMC). Beyond the Superhighway: Exploiting the Internet with Medical Informatics*, pages 259–263. Washington, D.C., October 26-30, 1996. Philadelphia, PA: Hanley & Belfus, 1996.

[10] Alexa T. McCray and Stuart J. Nelson. The representation of meaning in the UMLS. *Methods of Information in Medicine*, 34(1/2):193–201, 1995.

[11] Domenico M. Pisanelli, Aldo Gangemi, and Geri Steve. An ontological analysis of the UMLS metathesaurus. In C. G. Chute, editor, *AMIA'98 – Proceedings of the 1998 AMIA Annual Fall Symposium. A Paradigm Shift in Health Care Information Systems: Clinical Infrastructures for the 21st Century*, pages 810–814. Orlando, FL, November 7-11, 1998. Philadelphia, PA: Hanley & Belfus, 1998.

[12] Alan L. Rector, Sean Bechhofer, Carole A. Goble, Ian Horrocks, W. Anthony Nowlan, and W. Danny Solomon. The GRAIL concept modelling language for medical terminology. *Artificial Intelligence in Medicine*, 9:139–171, 1997.

[13] Manfred Schmidt-Schauß and Gerd Smolka. Attributive concept descriptions with complements. *Artificial Intelligence*, 48(1):1–26, 1991.

[14] James G. Schmolze and William S. Mark. The NIKL experience. *Computational Intelligence*, 7(1):48–69, 1991.

[15] Erich B. Schulz, Colin Price, and Philip J. B. Brown. Symbolic anatomic knowledge representation in the READ CODES Version 3: Structure and application. *Journal of the American Medical Informatics Association*, 4(1):38–48, 1997.

[16] Kent A. Spackman. Normal forms for description logic expression of clinical concepts in SNOMED RT. In S. Bakken, editor, *AMIA 2001 – Proceedings of the Annual Symposium of the American Medical Informatics Association. A Medical Informatics Odyssey: Visions of the Future and Lessons from the Past*, pages 627–631. Washington, D.C., November 3-7, 2001. Philadelphia, PA: Hanley & Belfus, 2001.

[17] Françoise Volot, Michel Joubert, and Marius Fieschi. Review of biomedical knowledge and data representation with Conceptual Graphs. *Methods of Information in Medicine*, 37(1):86–96, 1998.