Part-Whole Reasoning in Medical Ontologies Revisited – Introducing SEP Triplets into Classification-Based Description Logics

Stefan Schulz [a,b] Martin Romacker [a,b] Udo Hahn [a]

[a]Freiburg University, Computational Linguistics Lab (http://www.coling.uni-freiburg.de)

[b]Freiburg University Hospital, Department of Medical Informatics (http://www.imbi.uni-freiburg.de/medinf)

The development of powerful and comprehensive medical ontologies that support formal reasoning on a large scale is one of the key requirements for clinical computing in the next millennium. Taxonomic medical knowledge, a major portion of these ontologies, is mainly characterized by generalization and part-whole relations between concepts. While reasoning in generalization hierarchies is quite well understood, no fully conclusive mechanism as yet exists for part-whole reasoning. The approach we take emulates part-whole reasoning via classification-based reasoning using SEP triplets, a special data structure for encoding part-whole relations that is fully embedded in the formal framework of standard description logics.

INTRODUCTION

The provision of controlled medical terminology services [1, 2, 3] within clinical information systems constitutes a major challenge in medical informatics research in order to facilitate semantic interoperability [4, 5]. The desideratum of standardized, ubiquitous and logically consistent terminological knowledge repositories for clinical communication and information management is, however, not accomplished by most existing large terminologies such as MeSH, SNOMED, the Read thesaurus and UMLS. This is due to a semantic underspecification of concepts and relations, e.g., the lack of an explicit distinction between generalization (*is-a*) and partitive (*part-whole*) relations [6]. Especially noteworthy is the frequent mixture of these two relations that often occur at the same hierarchical level. In MeSH, for instance, "blood" can be found as parent of both "blood plasma" (*part-whole*) and "fetal blood" (*is-a*). This prevents the use of these terminological systems for formal reasoning procedures, since both types of knowledge require different encodings in order to achieve valid deductions, an idea that will be worked out in this paper.

Our model builds on recent efforts that have been made to overcome this lack of a clean semantics. Prominent examples are MED [7], K-Rep[8] and GALEN [9, 1], by which controlled clinical vocabularies can be encoded, using either a semantic network (MED) or a description logics approach (K-Rep, GALEN).

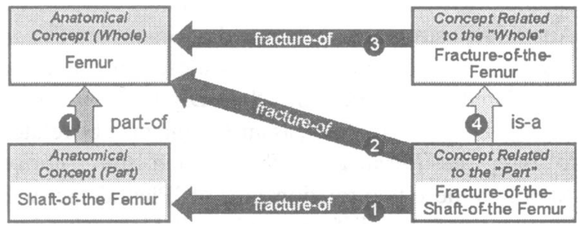

Figure 1: Part-Whole Specialization

THE PART-WHOLE REASONING PROBLEM

Many controversial discussions of part-whole (or meronymic) reasoning focus on the *transitivity* and *specialization* of the underlying PART-OF relations.

Transitivity. Transitivity issues concerning part-whole reasoning have largely been discussed in the literature, cf. the overview in [10]. A particularly strong claim is made by Winston et al. [11] who argue that part-whole relations can be considered transitive as long as "a single sense of part" is kept. Considering the medical domain, this means that if an anatomical object is part of the physical structure of another one (e.g., APPENDIX is a PART-OF the COLON), which itself is included in a larger structure (e.g., INTESTINE), then the first one is also a PART-OF this larger structure. So, we assume that transitivity generally holds for the ANATOMICAL-PART-OF relation.[1]

Part-whole specialization. Horrocks et al. [12] discuss *part-whole specialization* in depth, a reasoning pattern[2] which is defined by the inheritance of roles other than IS-A (!) along part-whole taxonomies. Fig. 1 illustrates a typical example following [12]. A concept x related via some relation **R** (here: FRACTURE-OF ❶) to a "part" concept y (here: SHAFT-OF-THE-FEMUR) of a whole z (here: FEMUR) is also related to the corresponding "whole" by **R** (❷) provided both, y and z, are related via a partitive relation **S** (here: PART-OF ❶) that "specializes" **R** just in terms of the PART-OF relation. Furthermore, defining a concept w (here:

[1]We are nevertheless aware of the fact that for certain subrelations of the ANATOMICAL-PART-OF relation the transitivity assumption is questionable or may even be rejected. Examples are LAYER-OF and LINEAR-DIVISION-OF. Hence, transitivity seems not to be an inheritable property.

[2]Given two relations, **R** and **S**, the implication $x\mathbf{R}y \wedge y\mathbf{S}z \Rightarrow x\mathbf{R}z$ holds, if **R** is PART-OF-specialized by **S**.

FRACTURE-OF-THE-FEMUR) related to the "whole" z by the same relation (❸), a subsumption relation (❹) can be deduced such that x IS-A w.

To account for this kind of reasoning, special concept representation languages, such as GRAIL [13], have been developed. Here, part-whole specialization is modeled as a property of certain conceptual relations in the form **R** *specializedBy* **S**, where **S** $\sqsubseteq$ PART-OF. This implies that the relation **R** is *always* propagated along hierarchies based on **S**, i.e. the inheritance mechanism is invariably associated with the relation **S**.

As a result of our experience with the construction of a pathology knowledge base and based on shared medical expertise we claim that part-whole specialization does *not generally* hold for all relations linking the same concept. For example, a PERFORATION-OF the APPENDIX can be classified as an INTESTINAL-PERFORATION, whereas an INFLAMMATION-OF the APPENDIX (APPENDICITIS) is certainly not an (INFLAMMATION-OF the INTESTINE) ENTERITIS. Also, the *same* relation between different concepts (e.g., INFLAMMATION-OF) may support part-whole specialization in one case, but not in the other. In contradistinction to the fact that APPENDICITIS is not an ENTERITIS, INFLAMMATION-OF, applied to another organ, e.g. the KIDNEY, exhibits a different behavior: PYELONEPHRITIS, an INFLAMMATION-OF the PYELON is consistently subsumed by NEPHRITIS. Thus, part-whole specialization is clearly not a property of the relation itself.

Currently, neither established large-scale terminologies nor dedicated medical knowledge representation languages are able to properly account for the above-mentioned, regular as well as irregular, phenomena typical of part-whole hierarchies.

TEXT UNDERSTANDING FRAMEWORK

The context of our research is determined by medSYNDIKATE [14, 15], a natural language text knowledge acquisition system that processes pathology reports. The expressiveness of its knowledge representation layer corresponds to the concept language $\mathcal{ALC}$ [16]. The actual implementation of $\mathcal{ALC}$, however, is done in LOOM [17], a KL-ONE-style terminological representation language [18].[3]

[3] Although LOOM is more expressive than $\mathcal{ALC}$, we will only refer to the latter, since its expressiveness is sufficient for our knowledge engineering requirements. $\mathcal{ALC}$ 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 fillers are linked to the relation name by a dot, e.g., $\exists R.C$. Note that neither $\mathcal{ALC}$ nor LOOM support the definition of transitive roles [18].

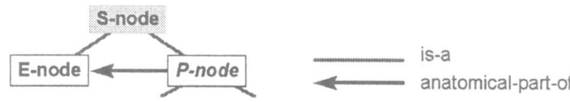

Figure 2: Basic Construct of Part-Whole Hierarchies

In an empirical study on text cohesion structures within pathology reports we have shown the importance of incorporating part-whole reasoning into proper text understanding processes [15]. In order to adapt the reasoning capabilities to these specific requirements, one possible solution might have been to introduce particular transitivity operators, such as proposed by Baader [19] and Sattler [20]. We refrain from such a solution, since we want to preserve the description logics as simple (and tractable) as possible in order to employ off-the-shelves knowledge representation systems (such as LOOM). Instead, we developed an alternative solution using just the expressiveness of $\mathcal{ALC}$ in order to embed part-whole reasoning fully into the standard terminological classification process. This proposal incorporates both previous work on description logics as applied to medicine [21] and large-scale medical coding systems [22].

TERMINOLOGICAL PART-WHOLE REASONING – AN ALTERNATIVE

Part-Whole Hierarchies and SEP Triplets. By using plain $\mathcal{ALC}$ without any language extensions, and by introducing a special data structure for part-whole encoding, we build up specifically structured IS-A hierarchies which support the emulation of inferences typical of transitive PART-OF relations. The same formalism also allows for conditioned part-whole specialization, i.e., mechanisms by which this reasoning mode can be enabled and disabled on demand.

In our domain model, the ANATOMICAL-PART-OF relation describes the partitive relation between physical parts of an organism. It is a constituent element of a specific SEP triplet data structure by which anatomical concepts are generally modeled (cf. Fig. 2). Each of these SEP triplets are dominated by a "structure" concept, the so-called **S-node** (e.g., INTESTINE-STRUCTURE). Each S-node subsumes a pair of concept siblings, namely an **E-node** and a **P-node**, both of which are conceptually related by the relation ANATOMICAL-PART-OF. The E-node denotes the *whole* anatomical entity to be modeled (e.g. INTESTINE), whereas the P-node (e.g. INTESTINE-PART) stands for any *part* of the corresponding E-node. Fig. 3 illustrates a fragment of the gastro-intestinal anatomy subdomain. Note that the formalism supports the definition of concepts as conjunctions of more than one P-node concept, as illustrated by the concept CAECUM-EPITHELIUM.

Let A, C and D be E-nodes (e.g., ORGANISM, CAECUM and APPENDIX), and $AStr$ be the top-level con-

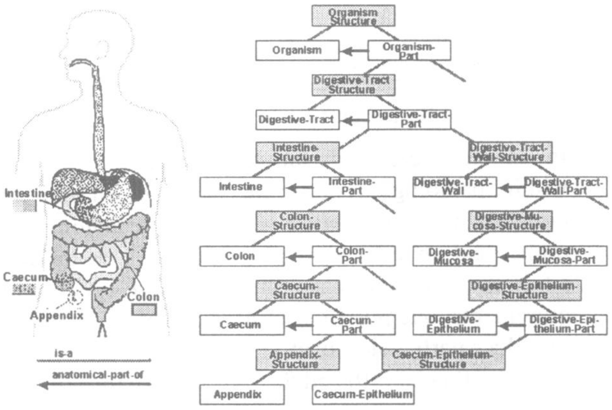

Figure 3: Fragment of the Part-Whole Taxonomy of the Gastrointestinal Tract Using Triplets.

cept of a domain subgraph (e.g., ORGANISM-STRUCTURE). $CStr$ and $DStr$ are the S-nodes that subsume C and D, respectively, just as CPt and DPt are the P-nodes related to C and D, respectively, via the role ANATOMICAL-PART-OF. All these concepts are embedded in a generalization hierarchy:

$$D \sqsubseteq DStr \sqsubseteq CPt \sqsubseteq CStr \sqsubseteq .. \sqsubseteq APt \sqsubseteq AStr \quad (1)$$

$$C \sqsubseteq CStr \sqsubseteq .. \sqsubseteq APt \sqsubseteq AStr \quad (2)$$

The P-node is defined as follows:

$$CPt \doteq CStr \sqcap \exists anatomical\text{-}part\text{-}of.C \quad (3)$$

Since D is subsumed by CPt (1), we infer that D is an ANATOMICAL-PART-OF the organ C :

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

It is obvious that this pattern holds at any level of the *part-whole* hierarchy. In our example (cf. Fig. 3), this may be illustrated by identifying D with APPENDIX, C with CAECUM and A with INTESTINE. APPENDIX is, consequently, an ANATOMICAL-PART-OF CAECUM and of INTESTINE. Single "proto nodes" (*viz.* S-nodes) are introduced as a means to make feasible *transitive* reasoning about partonomies *within* common *is-a* taxonomies. Hence, SEP triplets augment standard terminological inference engines by capabilities for part-whole reasoning without incurring unwarranted computational costs. These are likely to occur when terminological knowledge representation languages are extended by additional language constructs such as transitive roles [19, 20].
We are also able to model part-whole *specialization* using the same triplet structures. Whenever, e.g., a disease concept is related to an anatomical concept, one must explicitly determine whether part-whole specialization is supported or not. Part-whole specialization

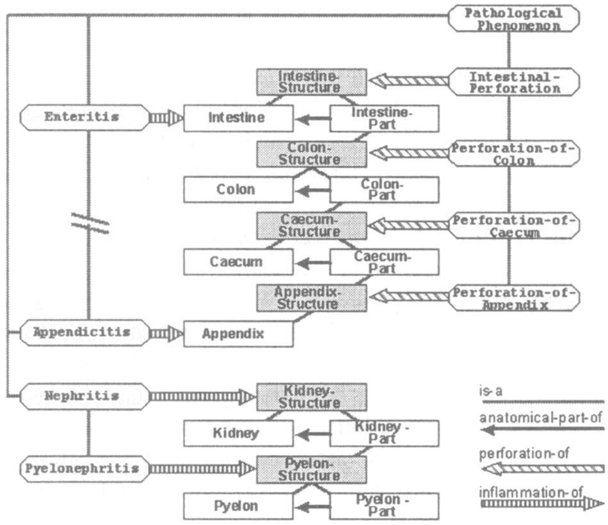

Figure 4: Conditioned Part-Whole Specialization in a Part-Whole Hierarchy. *Upper Part, Right Side*: "S-node Pattern" – Enabled Part-Whole Specialization (applied to the role PERFORATION-OF in the same subdomain). *Upper Part, Left Side*: "E-node Pattern" – Disabled Part-Whole Specialization (applied to the role INFLAMMATION-OF in the digestive tract subdomain). *Lower Part*: "S-node Pattern" – Enabled Part-Whole Specialization (the same role INFLAMMATION-OF as applied to the kidney subdomain)

is always inferred when a disease concept is linked to an S-node, while it is not, if a disease concept is connected to an E-node. An example is shown in Fig. 4 (upper part, right side). The concept INTESTINAL-PERFORATION is linked via the PERFORATION-OF relation to INTESTINE-STRUCTURE – an **S-Node**. This way, PERFORATION-OF-APPENDIX, PERFORATION-OF-CAECUM and PERFORATION-OF-COLON are all classified as INTESTINAL-PERFORATION. In contrast to this encoding, on the left side ENTERITIS is linked via the INFLAMMATION-OF relation to the **E-node** INTESTINE. So, APPENDICITIS as an INFLAMMATION-OF the APPENDIX is not classified as ENTERITIS.
Let us consider the same taxonomy as already described in the terminological expressions (1) to (4). Let R and S be relations that link a PATHOLOGICAL-PHENOMENON concept to an anatomical concept, and let W, X, Y and Z be concepts that stand for a PATHOLOGICAL-PHENOMENON. From

$$W \doteq \exists S.CStr \quad (5)$$

$$X \doteq \exists S.DStr \quad (6)$$

$$DStr \sqsubseteq CStr \quad (7)$$

we conclude that

$$X \sqsubseteq W \quad (8)$$

This "S-node pattern" realizes part-whole specialization, whereas the following "E-node pattern" does not:

$$Y \doteq \exists R.C \qquad (9)$$

$$Z \doteq \exists R.D \qquad (10)$$

The conclusion

$$Z \sqsubseteq Y \qquad (11)$$

cannot be drawn, since the extension of D is not a subset of the extension of C.

It is therefore only the difference in the ontology engineering patterns (linkage to S-nodes *vs.* linkage to E-nodes) that liberates or obviates part-whole specialization. Moreover, this effect is completely independent from the relations R and S. Thus, even the use of the same relation in both patterns is possible, as it is sometimes required by the structure of domain knowledge. An example is the relation INFLAMMATION-OF, comparing its use in two subgraphs as illustrated by Fig. 4. Here, the S-node pattern (formulae 5 – 8) is applied to the KIDNEY subgraph in order to define NEPHRITIS as well as PYELONEPHRITIS, whereas the definition of ENTERITIS and APPENDICITIS obeys the E-node pattern in the INTESTINE subgraph.

IMPLEMENTATION

Our previous work [14, 15] was based on a pathology knowledge base (267 concepts, 156 relations) implemented in LOOM without special constructs for part-whole reasoning. Early evaluations of medSYNDIKATE motivated the elaboration of the original representation model and led to the modification of a subset of the original knowledge base, by providing SEP triplets (covering 81 concepts, 54 relations). We experimentally validated the required reasoning capabilities, as well as the logical consistency of the knowledge base. Fortunately, the increase in the number of concepts and relations, which is due to the triplet structure, does not have a significant impact on the overall system performance. We are now beginning to reengineer the original knowledge base entirely according to the modeling principles described. Since the implementation of SEP triplets leads to a considerable amount of structurally identical LOOM code, we have developed a LISP code generator that allows for automated transformation of triplets from simple concept nodes, and, hence, facilitates knowledge base maintenance.

RELATED WORK

Considering the ontological structure of medical domain knowledge, Haimowitz et al. [23] first raised the claim for a dedicated representation formalism for part-whole relations and corresponding reasoning capabilities as an extension to terminological logics. Baader [19] and Sattler [20] discuss such extensions in terms of the transitive closure of role definitions and the implications this extension has on the computational complexity of the underlying language.

Schmolze and Marks [21] worked out a solution that also relies on subsumption to obtain inferences similar to that of transitive roles or transitive closure of roles. Their approach was criticized by Artale et al. [10] due to its inherent potential for a "proliferation of artificial concepts". We argue, however, that many of these additional concepts are by no means "artificial". On the contrary, we even claim that they reflect ontologically valid distinctions, as the necessity for conditioned part-whole specialization reveals (cf. Fig. 4). Nevertheless, our model implies to trade off the number of "proto nodes" (the structural S-nodes) and the additional complexity due to the internal structure of the triplets against other forms of complexity, e.g., the supply of special-purpose, i.e., part-whole-specific reasoning procedures. The latter is the case with the GRAIL formalism [13]. This framework constitutes the most far-reaching alternative approach to serve the needs of part-whole reasoning, and to incorporate the specialization feature into a medical ontology.

CONCLUSION

We introduced a general representation construct for part-whole modeling in the medical domain. It is fully embedded in the framework of a parsimonious variant of description logics, using the constructs of $\mathcal{ALC}$ only. This allows us to rely upon the built-in terminological classifier when emulating reasoning across part-whole hierarchies. Emphasis was laid on the simulation of transitivity and the construction of part-whole specialization. In contrast to approaches that extend description logics by transitive roles, we do not consider transitivity a property of the ANATOMICAL-PART-OF relation. In a strict sense, our model does not implement transitivity, but it emulates inferences typical of transitive part-whole reasoning through the taxonomy itself. This way, a serious limitation of GRAIL is overcome, *viz.* that part-whole specialization is invariably linked to a conceptual relation [24]. We avoid this dependency by conditioning (i.e., enabling and disabling) specialization through range constraints of the respective conceptual role. If the type of its range is an S-node, specialization is enabled, if it is an E-node, specialization is "switched off". Therefore, the specialization property is not contained in the conceptual relation itself, but in the structure of the ontology.

Acknowledgements. We want to thank our colleagues in the CLIF group and from the Department of Medical Informatics for fruitful discussions. S. Schulz and M. Romacker are supported by a grant from DFG (Ha 2097/5-1).

References

[1] A. Rector, W. Solomon, W. Nowlan, and T. Rush. A terminology server for medical language and medical information systems. *Methods of Information in Medicine*, 34(2):147–157, 1995.

[2] S. Falasconi, G. Lanzola, and M. Stefanelli. Ontology and terminology servers in agent-based health-care information systems. *Methods of Information in Medicine*, 36(1):30–43, 1997.

[3] A. Burgun, P. Denier, O. Bodenreider, G. Botti, O. Delamarre, B. Pouliquien, P. Oberlin, M. Lévêque, B. Lukacs, F. Kohler, M. Fieschi, and P. Le Beux. A Web terminology server using UMLS for the description of medical procedures. *JAMIA*, 4(5):356–363, 1997.

[4] D. Evans, J. Cimino, W. Hersh, S. Huff, and D. Bell. Toward a medical-concept representation language. *JAMIA*, 1(3):207–217, 1994.

[5] C. Friedman, S. Huff, W. Hersh, E. Pattison-Gordon, and J. Cimino. The Canon group's effort. Working towards a merged model. *JAMIA*, 2(1):4–18, 1995.

[6] A. T. McCray and S. Nelson. The representation of meaning in the UMLS. *Methods of Information in Medicine*, 24(1–2):193–201, 1995.

[7] J. J. Cimino, P. D. Clayton, G. Hripsack, and S. B. Johnson. Knowledge-based approaches to the maintenance of a large controlled medical terminology. *JAMIA*, 1(1):35–50, 1994.

[8] E. Mays, R. Weida, R. Dionne, L. Meir, B. White, and F. Oles. Scalable and expressive medical terminologies. In *Proc. of the 1996 AMIA Annual Fall Symposium*, pages 259–263. Philadelphia, PA: Hanley & Belfus, 1996.

[9] A. Rector and I. R. Horrocks. Experience building a large, re-usable medical ontology using a description logic with transitivity and concept inclusions. In *AAAI Spring Symposium on Ontological Engineering*. Menlo Park, CA: AAAI Press, 1997.

[10] A. Artale, E. Franconi, N. Guarino, and L. Pazzi. Part-whole relations in object-centered systems: an overview. *Data and Knowledge Engineering*, 20(3):347–383, 1996.

[11] M. Winston, R. Chaffin, and D. Herrmann. A taxonomy of part-whole relationships. *Cognitive Science*, 11:417–444, 1987.

[12] I. Horrocks, A. Rector, and C. Goble. A description logic based schema for the classification of medical data. In *KRDB'96 – Proc. of the 3rd Workshop on Knowledge Representation Meets Databases*, pages 24–28, 1996.

[13] A. L. Rector, S. Bechhofer, C. A. Goble, I. Horrocks, W. A. Nowlan, and W. D. Solomon. The GRAIL concept modelling language for medical terminology. *Artificial Intelligence in Medicine*, 9:139–171, 1997.

[14] U. Hahn, K. Schnattinger, and M. Romacker. Automatic knowledge acquisition from medical texts. In *Proc. of the 1996 AMIA Annual Fall Symposium*, pages 383–387. Philadelphia, PA: Hanley & Belfus, 1996.

[15] U. Hahn and M. Romacker. Text structures in medical text processing: empirical evidence and a text understanding prototype. In *Proc. of the 1997 AMIA Annual Fall Symposium*, pages 819–823. Philadelphia, PA: Hanley & Belfus, 1997.

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

[17] R. MacGregor. A description classifier for the predicate calculus. In *AAAI'94 – Proc. of the 12th National Conference on Artificial Intelligence*, pages 213–220. Menlo Park, CA: AAAI Press, 1994.

[18] W. Woods and J. Schmolze. The KL-ONE family. *Computers & Mathematics with Applications*, 23(2-5):133–177, 1992.

[19] F. Baader. Augmenting concept languages by transitive closure of roles: an alternative to terminological cycles. In *IJCAI'91 – Proc. of the 12th International Joint Conference on Artificial Intelligence*, pages 446–451. San Mateo, CA: Morgan Kaufmann, 1991.

[20] U. Sattler. A concept language extended with different kinds of transitive roles. In *KI'96 – Proc. of the 20th Annual German Conference on Artificial Intelligence*, pages 333–345. Berlin: Springer, 1996.

[21] J. Schmolze and W. Marks. The NIKL experience. *Computational Intelligence*, 6:48–69, 1991.

[22] E. B. Schulz, C. Price, and P. J. B. Brown. Symbolic anatomic knowledge representation in the Read Codes Version 3: structure and application. *JAMIA*, 4(1):38–48, 1997.

[23] I. J. Haimowitz, R. S. Patil, and P. Szolovits. Representing medical knowledge in a terminological language is difficult. In *SCAMC'88 – Proc. of the 12th Annual Symposium on Computer Applications in Medical Care*, pages 101–105. New York: IEEE Computer Society Press, 1988.

[24] J. Bernauer. Analysis of part-whole relation and subsumption in the medical domain. *Data and Knowledge Engineering*, 20(3):405–415, 1996.