Mereological Semantics for Bio-Ontologies

Udo Hahn[1] **Stefan Schulz**[2] **Kornél Markó**[1,2]

[1]Text Knowledge Engineering Lab, Freiburg University
Werthmannplatz 1, D-79085 Freiburg, Germany
http://www.coling.uni-freiburg.de

[2]Department of Medical Informatics, Freiburg University Hospital
Stefan-Meier-Str. 26, D-79104 Freiburg, Germany

Abstract

Biomedical ontologies are typically structured in a biaxial way, reflecting both a taxonomic and a mereological order. Common examples such as the *Gene Ontology* and the *Unified Medical Language System (UMLS)* excel in terms of coverage but lack an adequate semantics of the mereological relations they incorporate. This shortcoming is particularly evident as far as the (non-)mandatory existence of parts for their wholes is concerned, on the one hand, and the propagation of properties across part-whole hierarchies, on the other hand. We provide a formal specification of the semantic foundations of mereology in the biomedical domain that is closely linked to the paradigm of description logics. In essence, we here propose to emulate mereological reasoning by taxonomic reasoning. In an attempt to capture much of the shared intuition underlying merelogical reasoning in the biomedical domain, we distinguish for each mereologically relevant concept four different classes of parts and wholes which allow for the expression of five different propagation patterns.

Introduction

The notion of 'biological structure' plays a key role in bio-ontologies, since tangible parts of organisms such as organs, tissues, cells, molecules, etc. constitute the location of biological processes and are the targets of experimental, diagnostic and therapeutic interventions. Up until now, most biomedical concept systems encode terminological knowledge in an informal thesaurus-style way (McCray & Nelson 1995) such as the source vocabularies of the *UMLS Metathesaurus* (UMLS 2004) and the *Gene Ontology* (Gene Ontology Consortium 2001). The latter is a nice example of a domain description, which apparently provides "clean", i.e., acyclic *is-a* and *part-of* hierarchies, but refrains from specifying explicitly the semantic nature of its foundational relations (Smith, Williams, & Schulze-Kremer 2003). The *Foundational Model of Anatomy* [FMA] (Rosse *et al.* 1998; Noy *et al.* 2002) has just recently become an exception to this rule as its creators started to develop a proper axiomatic basis (Smith & Rosse 2004). This desideratum has also been formulated after auditing the FMA using formal reasoning devices (Hahn & Schulz 2003; Beck & Schulz 2003).

As an example of semantic underspecification, assertions such as *Cell Has-Part Cell Nucleus* leave open whether *Cell Nucleus* should be interpreted as some kind of sortal restriction (or sanctioning) of the role filler, or as an existential condition for *Cell*. Such existential dependency relationships are, however, crucial for adequate domain models which should *(i)* allow the expression of ontological dependency (e.g., "each *cell nucleus* is part of a *cell*"), *(ii)* permit possible relations ("a *cell* may have *cell nuclei*"), and *(iii)* reject invalid assertions ("there is a *cell nucleus* which is part of a *protein molecule*").

Of equal importance is another reasoning phenomenon, *viz.* the propagation of properties between parts and wholes (Artale *et al.* 1996; Horrocks, Rector, & Goble 1996; Rogers & Rector 2000; Rector 2002; Horrocks & Sattler 2003). The concept *insulin secretion*, e.g., is usually considered to be related to the concept *pancreas* by the role *function-of*, because *pancreatic beta cells* (cells which produce the substance) are considered *part-of* a *pancreas*. In the same way, *contraction* would be classified as a *function-of* a *muscle*, since it is a *function-of* the *actin-myosin complex* which is a *component-of muscle cells*, the latter being *part-of muscle*.

However, there are counterexamples: An *amputation-of toe* cannot be classified as *amputation-of foot* although every *toe* is *part-of* a *foot*. *Mitosis* is a *cell* function, but it is generally not seen as a *pancreas* or *liver* function although these organs have *cells* as parts. Such propagation patterns – attributes propagate from parts to wholes, or from wholes to parts or do not propagate at all – bear subtle intricacies which still have not been sufficiently accounted for in biomedical knowledge engineering although they are relevant in order to enhance the empirical adequacy and expressive power of biomedical ontologies.

In this paper, we will discuss and classify typical reasoning patterns from which we derive our requirements for adequate biomedical domain modeling. Then we propose an ontology engineering approach, based on our previous work on biomedical knowledge representation, in the context of $\mathcal{ALC}$-based description logics (Baader & Nutt 2003). The central idea of this approach (Hahn, Schulz, & Romacker 1999) has already proved useful for the construction of a large terminological biomedical domain knowledge base. (Hahn & Schulz 2003).

Semantic Distinctions for Parts and Wholes

In this section, we discuss and formalize four dependency patterns by which parts and wholes are conceptually related. We draw on common categories for knowledge representation. Concepts are characterized by their attributes or properties, e.g., *slots* in frame-based representations or *roles* in description logic systems (such as *has-location, part-of,* etc.), while role filler restrictions constrain their sortal ranges. For example, the concept *liver* is a reasonable filler of the role *has-location* of *hepatitis*. With C being the concept to be defined (i.e., a unary predicate), r a conceptual relation (i.e., a binary predicate) and F the target concept (role filler), the following distinctions on the role status can be made:

- **Mandatory roles** of a concept C are required to be filled at least once by a filler of the type F, as in the case of *has-location liver* when defining the concept *hepatitis*. This allows one to infer that for each concrete *hepatitis* there *must* be a concrete *liver* where it is located.
- **Forbidden roles**. F is *disallowed* to be a filler of the role r of the concept C. For the concept *dermatitis* and the relation *has-location*, e.g., a filler such as *liver* must be rejected in order to prevent unintended models.
- **Possible roles** r are used to *allow* an instance of C to be related to an instance of F via r, but this is not necessarily an existential condition for C.[1] Possible roles constitute the complement to forbidden ones. For example, *has-part* is a possible role with filler *nucleus* when we define the concept *cell*, as cells with and without nuclei exist (e.g., bacteria).

In $\mathcal{ALC}$[2] notation (Baader & Nutt 2003), mandatory roles are specified in terms of existential quantification, whereas possible roles can be implicitly expressed in terms of conceptual constraints.

We now classify mereologically relevant concepts with regard to their role dependence status. Considering part-whole hierarchies, the notion of mandatory roles gives rise to the distinction of mp *mandatory* parts, on the one hand, and sp *specific* parts, on the other hand. Also taking into consideration *has-part*, the inverse relation of *part-of*, on equal terms, we are able to express in a similar way mi *mandatory* wholes (includers), and si *specific* wholes (includers).[3] This leads, quite naturally, to the following distinction of four kinds of related concepts for each concept in a partonomy. These dependency patterns describe particular 'views' from parts to a whole or wholes to a part, which we formalize as corresponding existential conditions for parts or wholes.

[1] All mandatory roles are possible roles, too.

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

[3] We prefer the term 'includer' as argued for by Schulz & Hahn (2002). This should not be confounded with a solely topological understanding of part-whole relations such as discussed by Schulz & Hahn (2004).

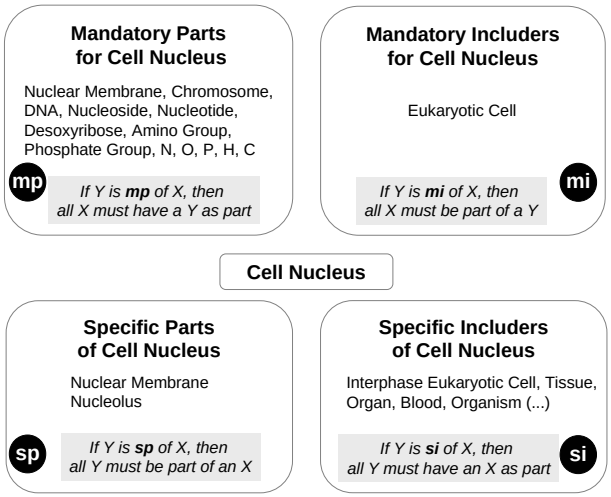

Figure 1: Part-Whole Conceptualizations of *Cell Nucleus*.

mp W_{mp} (**mandatory parts** of a "whole concept" W) subsumes all those "part concepts" P_1, P_2,... P_n, which fulfill the following condition: Each instance w of W *must* be related by *has-part* to at least one instance of each "part concept" p_1, p_2,... p_n of P_1, P_2,... P_n, respectively. Here, the whole depends on the parts, or in other words, the parts are mandatory for the whole.

sp W_{sp} (**specific parts** of a "whole concept" W) subsumes those concepts P_1, P_2, ... P_n, which depend on a given class of wholes W. Each instance p_1, p_2,... p_n of P_1, P_2,... P_n, respectively, *must* be related to w (an instance of W) via *part-of*. Here, the parts depend on the whole, or the whole is mandatory for the parts.

mi P_{mi} (**mandatory includers** of a "part concept" P) subsumes all those "whole concepts" W_1, W_2,... W_n, which fulfill the following condition: Each instance p of P *must* be related via *part-of* to at least one instance of each "whole concept" w_1, w_2,... w_n of W_1, W_2,... W_n, respectively. In this case, the part depends on the wholes, or the wholes are mandatory for the part.

si P_{si} (**specific includers** of a "part concept" P) subsumes those "whole concepts" W_1, W_2, ... W_n, which depend on a given class of parts P. Each instance w_1, w_2,... w_n of W_1, W_2,... W_n, respectively, *must* be related to p (an instance of P) via *has-part*. Here, the wholes depend on the part, or the part is mandatory for the wholes.

Figure 1 illustrates these four concept groups, using *cell* as an example. A cell nucleus has as *mandatory parts* mp nuclear membrane, as well as DNA, which is axiomatically constituted by nucleotides etc., the latter by nucleosides, nucleosides by functional groups, and functional groups by atoms. Nuclear membrane and nucleolus are necessarily parts of a cell nucleus, the latter being here a *specific part* sp. The existence of a cell nucleus implies the existence of a eukaryotic cell as a *mandatory includer* mi. Finally, interphase (i.e., non-replicating) eukaryotic cells, tissues, organs, blood, organisms necessarily have a cell nucleus as part, they are *specific includers* si.

C_{mp} can be defined as the common subsumer of all those concepts, for which C is a specific includer; C_{mi} as the com-

mon subsumer of those for which C is a specific part. Note that $C_{mp}(x)$ and $C_{mi}(y)$ hold true even for individuals x and y which are *not* part of a C or *do not* have C as their part, e.g., *atom* for physical objects or *organism* for body parts. Rather the extension of C_{mp} (C_{mi}) includes all those individuals which – by their sortal kind – are principally allowed to be a part (respectively includer) of C. Subsequently, we will use this 'modal' reading to express the notion of *possible* parts and wholes.

Next, we discuss some algebraic properties of *part-of* and *has-part* (at the instance level). We note that *has-part* is the inverse relation of *part-of*:

$$\forall x, y : part\text{-}of(x, y) :\Leftrightarrow has\text{-}part(y, x) \quad (1)$$

According to Casati & Varzi (1999), the generic parthood relation is reflexive, antisymmetric and transitive. Although the transitivity property is by no means uncontroversial (Winston, Chaffin, & Herrmann 1987), we here claim:

$$\forall x, y, z : (part\text{-}of(x, y) \wedge part\text{-}of(y, z)) \Rightarrow part\text{-}of(x, z) \quad (2)$$

Now, we introduce some new kinds of mereological relations. In contradistinction to *part-of* and *has-part*, they relate concepts (i.e., classes), not individuals.[4] These relations are '*specific part of*' (SP), '*specific includer of*' (SI), '*mandatory part of*' (MP), and '*mandatory includer of*' (MI). Given the transitivity of *part-of* and *has-part*, we can show that these relations are transitive, too ($\pi = \{SP, SI, MP, MI\}$):

$$\forall X, Y, Z : (\pi(X, Y) \wedge \pi(Y, Z)) \Rightarrow \pi(X, Z) \quad (3)$$

According to the above definition of the common subsumer concepts C_{mi} and C_{mp}, we may also conclude:

$$\forall X, Y : SP(X, Y) \Leftrightarrow MI(Y, X) \quad (4)$$

$$\forall X, Y : SI(X, Y) \Leftrightarrow MP(Y, X) \quad (5)$$

The proposed encoding enables us not only to directly address mandatory and specific parts, mandatory includers, and specific includers of a given concept, but also offers an opportunity to capture *possible parts* (PP) as well as *possible includers* (PI):

$$\forall X, Y : (MP(X, Y) \vee SP(X, Y)) \Rightarrow PP(X, Y) \quad (6)$$

$$\forall X, Y : (MI(X, Y) \vee SI(X, Y)) \Rightarrow PI(X, Y) \quad (7)$$

provided that there are no parts left that are neither mandatory nor specific. With (4) and (5), we then derive

$$\forall X, Y : PP(X, Y) \Leftrightarrow PI(Y, X) \quad (8)$$

We now illustrate examples of biomedical reasoning which can be stated using the predicates we introduced above. These examples are not meant to be generally valid inference rules. Rather they stand for particular concept instantiations in the context of biomedicine for which the chosen instantiation yields a valid reasoning result that meets the shared intuitions in the biomedical community. Still, these instantiations hint at another challenging problem we will not get into, *viz.* the interaction of part-whole conceptualizations we propose with spatial, functional, etc. relations in terms of *general* inference schemata.

1. *Insulin synthesis* takes place (*has-location*, HL) in the pancreatic *Beta Cells*. *Beta Cells* are specific parts of *Pancreas*. We infer that *Insulin Synthesis* is also located in the *Pancreas*:

Insulin Synthesis	HL	*Beta Cells*
Beta Cells	SP	*Pancreas*
Insulin Synthesis	HL	*Pancreas*

   Therefore, the role filler of *has-location* needs to be sortally restricted by *Beta Cells* or "whatever mandatorily includes *Beta Cells*", the latter corresponding to *BetaCells*$_{mi}$. This inference pattern seems to be typical of the *location* of biological processes. It also explains why the following inference appears counter-intuitive:

Mitosis	HL	*Eukaryotic Cells*
Eukaryotic Cells	MP	*Human Body*
Mitosis	HL	*Human Body* (**FALSE**)

   In contradistinction to the first example, this part is not specific for the whole: Eukaryotic cells can be found in all plants and animals, they can even constitute organisms of their own. Therefore, the conclusion is not necessarily true.

2. The following example illustrates another case in which an inference across part-whole hierarchies is prohibitive (the relation HT stands for *has-target*):

Amputation of Toe	HT	*Toe*
Toe	SP	*Foot*
Amputation of Toe	HT	*Foot* (**FALSE**)

   Obviously, an amputation of a toe is not an amputation of a foot. HT targets an anatomical structure as a whole and does not include any part or includer into its range.

3. In contrast to the above example of invalid "upstream propagation" from the parts to the whole, here is an example of valid "downstream propagation" from the whole to the parts (IO stands for *ingestion-of*):

Erythrocyte-Phagocytosis	IO	*Erythrocyte*
Erythrocyte	SI	*Iron*
Erythrocyte-Phagocytosis	IO	*Iron*

   The role filler of *ingestion-of* would therefore be sortally restricted by *Erythrocyte* or "whatever mandatorily constitutes *an Erythrocyte* ", the latter corresponding to *Erythrocyte*$_{si}$.

4. Another example of "downstream propagation" (FO stands for *function-of*):

Immunity	FO	*Immune System*
Immune System	MI	γ*Globulin*
Immunity	FO	γ*Globulin*

   The role filler of *function-of* would therefore be sortally restricted by *Immune System* or "any specific part of an *Immune System*", the latter corresponding to *ImmuneSystem*$_{sp}$. The latter does *not* include components of γGlobulin such as amino acids, since these are not *specific* (albeit *mandatory*) ones. This inference scheme may

[4] This difference is highlighted by capital letters for concept-to-concept relations, as well as for concept (class) names.

be restricted to functional systems in biology (such as the immune system, the circulatory system, etc.), and cannot be generalized to any composite biological structure.

We observe that once we use the four auxiliary nodes in concept definitions we get a more precise description which yields inference results that match community-wide shared biomedical intuitions.

Parts and Wholes Recast in Description Logic

Reasoning engines for biomedical applications face large-sized knowledge bases, usually on the order of hundreds of thousands of concepts and relations. Therefore, we intentionally chose $\mathcal{ALC}$ as our formal framework, a rather simple variant of description logics that is supported by most off-the-shelf reasoning engines. Hence, we diverge from most alternative approaches to part-whole reasoning which advocate extensions of the underlying (description) logic (Franconi 1993; Padgham & Lambrix 1994; Bernauer 1996; Rogers & Rector 2000; Horrocks & Sattler 2003). At the methodological level, the gain in expressive flexibility these approaches offer has to be traded against the proliferation of (artificial) concepts and relation hierarchies in our approach. At the system engineering level, however, language extensions always require substantial adaptation or even completely new implementations of the inference engines in order to make these extensions work. Due to the inherent complexity of description classifiers the resulting extended engines usually lack the capability to deal robustly with very large knowledge bases.

Based on the $\mathcal{ALC}$ framework, we will now recast the stipulations from the previous section in terms of what we call the extended **PI** ("**p**arts"/"**i**ncluder") model which updates our previous work on the *SEP* (Hahn, Schulz, & Romacker 1999) and *Extended SEP* (Schulz & Hahn 2002) models. The extended PI model is centered around the relations *part-of* and *has-part* with the semantics introduced above. We emulate transitive mereological reasoning by taxonomic reasoning. This is achieved through the introduction of additional concepts which serve as 'reificators' of the relations *part-of* and *has-part*. For each concept S we introduce two artificial concepts for both the relations *part-of* and *has-part*. These 'proto' concepts are common subsumers for all those concepts which must have, by definition, the role *part-of* filled by S, or the role *has-part* filled by S.

$$S_{sp} \equiv \exists part\text{-}of.S \quad (9) \qquad S_{si} \equiv \exists has\text{-}part.S \quad (10)$$

S_{sp} then subsumes all concepts that have S as a *mandatory includer* (whole), and S_{si} subsumes the ones which have S as a *mandatory part*. As an example, the concept $Hand_{sp}$ subsumes both *Finger* and *Thumb*. Cascading subsumption of concepts by 'part' nodes of their mandatory wholes emulates transitivity of the *part-of* relation. The same applies to the emulation of *has-part* transitivity: In this case, concepts – together with their corresponding 'includer' nodes – are subsumed by 'includer' nodes of their mandatory parts.

Figure 2 depicts a scenario in which mandatory parts correspond to mandatory includers (wholes). Each instance of *Cell Membrane* (M) has its role *part-of* filled by an instance

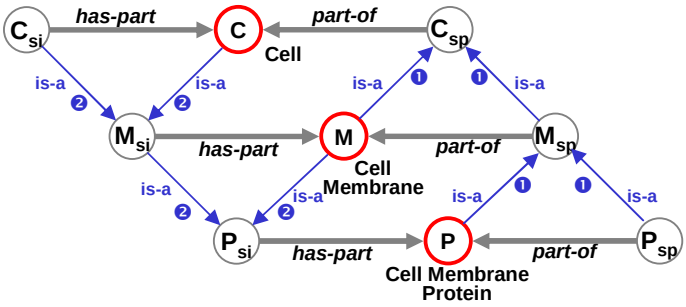

Figure 2: PI Architecture: Emulation of transitivity of both *part-of* and its converse *has-part*. The *is-a* links with label (1) emulate *part-of* hierarchies, those with label (2) emulate *has-part* hierarchies.

of *Cell* (C), because M is subsumed by C_{sp}. Each *Cell* (C) has the role *has-part* filled by an instance of *Cell Membrane*, because C is subsumed by M_{si} (the same applies to *Cell Membrane* (M) and *Cell Membrane Protein* (P)). As P is subsumed by M_{sp} and M_{sp} by C_{sp}, we infer that not only M, but also C is a mandatory whole for P. Accordingly, C is subsumed not only by M_{si}, but also by P_{si}, therefore both M and P are mandatory parts of C (in other words, each instance of *Cell* must have an instance of *Cell Membrane* and an instance of *Cell Membrane Protein* related via *has-part*).

For a formal specification of these considerations, let us assume C, M and P denote the concepts to be modeled, and C_{sp}, M_{sp}, and P_{sp} denote "part" nodes, related to C, M, and P via the role *part-of*, respectively. Accordingly, C_{si}, M_{si}, and P_{si} denote "includer" nodes related to C, M, and P via the role *has-part*, respectively. Translating Figure 2 into (redundant) description logic expressions, we get:

$$M \sqsubseteq C_{sp} \quad (11) \qquad C_{si} \sqsubseteq M_{si} \sqsubseteq P_{si} \quad (18)$$
$$M_{sp} \sqsubseteq C_{sp} \quad (12) \qquad C_{sp} \equiv \exists part\text{-}of.C \quad (19)$$
$$P \sqsubseteq M_{sp} \sqsubseteq C_{sp} \quad (13) \qquad M_{sp} \equiv \exists part\text{-}of.M \quad (20)$$
$$P_{sp} \sqsubseteq M_{sp} \sqsubseteq C_{sp} \quad (14) \qquad P_{sp} \equiv \exists part\text{-}of.P \quad (21)$$
$$M \sqsubseteq P_{si} \quad (15) \qquad C_{si} \equiv \exists has\text{-}part.C \quad (22)$$
$$M_{si} \sqsubseteq P_{si} \quad (16) \qquad M_{si} \equiv \exists has\text{-}part.M \quad (23)$$
$$C \sqsubseteq M_{si} \sqsubseteq P_{si} \quad (17) \qquad P_{si} \equiv \exists has\text{-}part.P \quad (24)$$

It is obvious that, by using this pattern across various physical concepts linked with each other via the *part-of* or the *has-part* relation, we get the same deductions as if *part-of* and *has-part* were really transitive at the level of concepts.

However, using the two auxiliary nodes from expressions (9) and (10), it is still not possible to address mandatory parts and includers, which so far are only available as a filler of the (inherited) roles *has-part* and *part-of*. Therefore, we provide two additional hierarchies, *viz.* the hierarchy of mandatory parts and that of mandatory includers. As all the information is already available in the PI hierarchy the new hierarchies, indeed, contain redundant information. The construction of these hierarchies is straightforward: A C_{mp} node (mandatory part) subsumes all first-level (non-inherited) role fillers of *has-part* at C, and C_{mi} (mandatory includer) all first-level role fillers of *part-of* at C, as well as their associated *mp* and *mi* nodes.

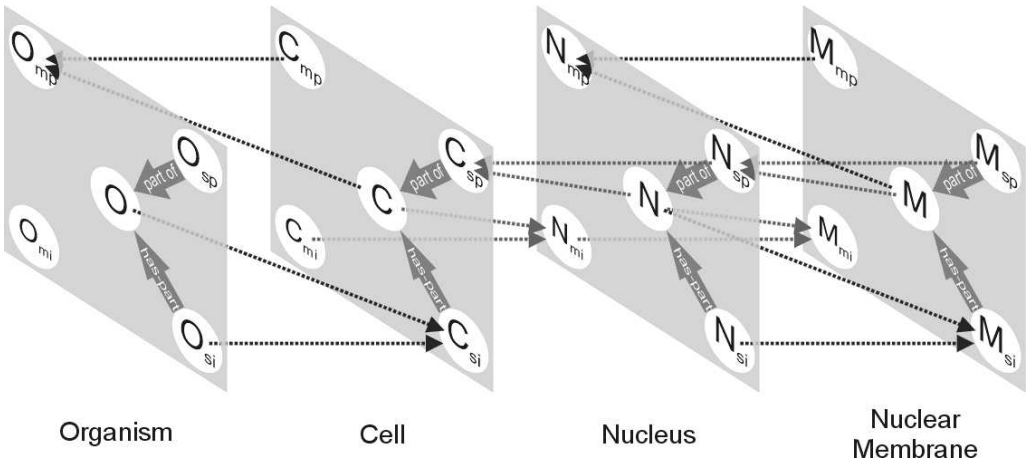

Figure 3: Extended PI Architecture using the quadruple concept encoding pattern. The *si*-nodes are linked to their core concepts by *has-part*, the *sp*-nodes by *part-of* (dotted lines indicate subsumption relations).

In

$$C \sqsubseteq A_{si} \sqcap B_{si} \tag{25}$$

C inherits the roles $has\text{-}part.A$ and $has\text{-}part.B$; the concept C also has A and B as specifi c includers. Hence, the auxiliary node C_{mp} has to subsume both A and B, but also A_{mp} and B_{mp}. The following axioms will have to be added:

$$A \sqsubseteq C_{mp} \quad (26) \qquad A_{mp} \sqsubseteq C_{mp} \quad (28)$$
$$B \sqsubseteq C_{mp} \quad (27) \qquad B_{mp} \sqsubseteq C_{mp} \quad (29)$$

The flexibility of this approach is shown in Figure 3:

(i) An *organism* is mandatorily constituted by a *cell*;
(ii) a *cell* is not necessarily part of an organism;
(iii) a *nucleus* is mandatorily part of a *cell*;
(iv) cells without nuclei exist;
(v) a *nucleus* mandatorily has a *nuclear membrane*;
(vi) a *nuclear membrane* is mandatorily part of a *nucleus*.

(i) is reflected by the subsumption of O and O_{si} by C_{si}, and, consequently, by the subsumption of C and C_{mp} by O_{mp}. Due to (ii), there are no subsumptions by *mi* and *sp* nodes. (iii) is reflected by the subsumptions of N and N_{sp} by C_{sp}, and, consequently, by the subsumption of C and C_{mi} by N_{mi}. With (iv) there is no subsumption of C by N_{si}. (v) and (vi) result from the following subsumptions: $N \sqsubseteq M_{si}$, $N_{si} \sqsubseteq M_{si}$, $N \sqsubseteq M_{mi}$, $N_{mi} \sqsubseteq M_{mi}$, $M \sqsubseteq N_{sp}$, $M_{sp} \sqsubseteq N_{sp}$, and $M \sqsubseteq N_{mp}$, $M_{mp} \sqsubseteq N_{mp}$.

This encoding allows us to introduce a range restriction into concept defi nitions suitable for allowing "possible" parts and wholes, and obviating "forbidden" role fi llers. According to (6) and (7), we constrain the defi nition of the concept *cell* (C) by the following range restrictions:

$$C \sqsubseteq \forall part\text{-}of.(C_{si} \sqcup C_{mi}) \tag{30}$$
$$C \sqsubseteq \forall has\text{-}part.(C_{sp} \sqcup C_{mp}) \tag{31}$$

Thus, an instance of *cell* may have an instance of *nucleus* and *nuclear membrane* as part, as well as an instance of *organism* as includer, but any object not contained in the extension of the two disjunctions[5] would be disallowed to be connected to an instance of C.

Finally, we demonstrate how the proposed part-whole model is capable of accounting for the inferences discussed previously (numbers correspond to the examples).

1. $InsulinSynthesis \sqsubseteq$
   $\exists has\text{-}location.(BetaCells \sqcup BetaCells_{mi})$
   $InsulinSynthesis \sqsubseteq$
   $\forall has\text{-}location.(BetaCells \sqcup BetaCells_{mi} \sqcup BetaCells_{si})$
2. $AmputationOfToe \sqsubseteq \exists has\text{-}target.Toe$
   $AmputationOfToe \sqsubseteq \forall has\text{-}target.Toe$
3. $ErythrocytePhagocytosis \sqsubseteq$
   $\exists ingestion\text{-}of.(Erythrocyte \sqcup Erythrocyte_{mp})$
   $ErythrocytePhagocytosis \sqsubseteq$
   $\forall ingestion\text{-}of.(Erythrocyte \sqcup Erythrocyte_{mp} \sqcup Erythrocyte_{sp})$
4. $Immunity \sqsubseteq$
   $\exists function\text{-}of.(ImmuneSystem \sqcup ImmuneSystem_{sp})$
   $Immunity \sqsubseteq$
   $\forall function\text{-}of.(ImmuneSystem \sqcup ImmuneSystem_{sp} \sqcup ImmuneSystem_{mp})$

The *PI* encoding not only provides an elegant way of emulating *part-of* and *has-part* transitivity through taxonomic subsumption, but also grants more flexibility and precision in adapting the ontology to common conceptualizations in the biomedical domain. By adding four "proto-nodes" for every concept we are able to express four propagation patterns (two downstream and two upstream), as well as one pattern without propagation. The "proto-nodes" also support several control mechanisms for the propagation of roles within mereological hierarchies. Hence, our approach cannot be considered just a data structure 'hack' but rather bears ontological plausiblity in the biomedical domain.

[5] In order to simplify the underlying logic, the disjunctions can be avoided by common subsumer concepts.

Conclusions

Adequate and expressive domain models require not only appropriate ontological categories but also the support of typical reasoning patterns. Our solution is based on a layered modeling strategy for concepts within standard description logics, essentially based upon the introduction of additional concepts such as reifi cators for the relations *part-of* and *has-part*. These artifacts not only enable the propagation of roles across partonomic hierarchies, but also the accommodation of various standard reasoning patterns. Unlike the approaches which advocate language extensions in order to adequately deal with part-whole reasoning, we do not require newly designed (and implemented) terminological inference engines.

In order to deal with the proliferation of concepts we found it imperative to ease the knowledge editing and maintenance phase. We chose to use an intermediate representation format as interface to the knowledge engineer, a frame-based system with a graphical user interface (Noy, Fergerson, & Musen 2000). We combined the ease of knowledge maintenance with the power of a terminological reasoner for our current experiments in which we converted the Foundational Model of Anatomy into a description logic format using the extended PI architecture (Beck & Schulz 2003).

References

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

Baader, F., and Nutt, W. 2003. Basic description logics. In Baader, F.; Calvanese, D.; McGuinness, D.; Nardi, D.; and Patel-Schneider, P., eds., *The Description Logic Handbook. Theory, Implementation and Applications*. Cambridge University Press. 43–95.

Beck, R., and Schulz, S. 2003. Logic-based remodeling of the digital anatomist foundational model. In *AMIA'03 – Proceedings of the 2003 Annual Symposium of the American Medical Informatics Association*, 687–691.

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

Casati, R., and Varzi, A. C. 1999. *Parts and Places. The Structures of Spatial Representation*. MIT Press.

Franconi, E. 1993. A treatment of plurals and plural quantifi cations based on a theory of collections. *Minds and Machines* 3(4):453–474.

Gene Ontology Consortium. 2001. Creating the Gene Ontology resource: Design and implementation. *Genome Research* 11(8):1425–1433.

Hahn, U., and Schulz, S. 2003. Building a very large ontology from medical thesauri. In Staab, S., and Studer, R., eds., *Handbook on Ontologies in Information Systems*. Springer. 133–150.

Hahn, U.; Schulz, S.; and Romacker, M. 1999. Partonomic reasoning as taxonomic reasoning in medicine. In *AAAI'99 – Proceedings of the 16th National Conference on Artificial Intelligence*, 271–276.

Horrocks, I., and Sattler, U. 2003. The effect of adding complex role inclusion axioms in description logics. In *IJCAI'03 – Proceedings of the 18th International Joint Conference on Artificial Intelligence*, 343–348.

Horrocks, I.; Rector, A.; and Goble, C. 1996. A description logic based schema for the classifi cation of medical data. In *KRDB'96 – Proceedings of the 3rd Workshop Knowledge Representation Meets Databases*, 24–28.

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

Noy, F.; Musen, M. A.; Mejino, J. L. V.; and Rosse, C. 2002. Pushing the envelope: Challenges in a frame-based representation of human anatomy. Technical Report SMI-2002-0925, 2002, Stanford University.

Noy, F.; Fergerson, R. W.; and Musen, M. A. 2000. The knowledge model of PROTEGE-2000: Combining interoperability and flexibility. In *Proceedings of the 12th International Conference on Knowledge Engineering and Knowledge Management – EKAW 2000*, 17–32.

Padgham, L., and Lambrix, P. 1994. A framework of part-of hierarchies in terminological logics. In *Principles of Knowledge Representation and Reasoning. Proceedings of the 4th International Conference – KR'94*, 485–496.

Rector, A. L. 2002. Analysis of propagation along transitive roles: Formalisation of the GALEN experience with medical ontologies. In *DL 2002 – Proceedings of the 2002 International Workshop on Description Logics*.

Rogers, J., and Rector, A. 2000. GALEN's model of parts and wholes: Experience and comparisons. In *AMIA 2000 – Proceedings of the Annual Symposium of the American Medical Informatics Association*, 714–718.

Rosse, C.; Mejino, J. L. V.; Modayur, B. R.; Jakobovits, R.; Hinshaw, K. P.; and Brinkley, J. F. 1998. Motivation and organizational principles for anatomical knowledge representation: The DIGITAL ANATOMIST symbolic knowledge base. *Journal of the AMIA* 5(1):17–40.

UMLS. 2004. *Unified Medical Language System*. Bethesda, MD: National Library of Medicine.

Schulz, S., and Hahn, U. 2002. Necessary parts and wholes in bio-ontologies. In *Proceedings of the 8th International Conference – KR 2002*, 387–394.

Schulz, S., and Hahn, U. 2004. Parthood as spatial inclusion: Evidence from biomedical conceptualizations. In *Proceedings of the 9th Intl. Conference – KR 2004*.

Smith, B., and Rosse, C. 2004. The role of foundational relations in the alignment of biomedical ontologies. In *MEDINFO 2004 – Proceedings of the 11th World Congress on Medical Informatics*.

Smith, B.; Williams, J.; and Schulze-Kremer, S. 2003. The ontology of the Gene Ontology. In *AMIA'03 – Proceedings of the 2003 Annual Symposium of the American Medical Informatics Association*, 609–613.

Winston, M.; Chaffi n, R.; and Herrmann, D. J. 1987. A taxonomy of part-whole relationships. *Cognitive Science* 11:417–444.