

# Parthood as Spatial Inclusion – Evidence from Biomedical Conceptualizations

**Stefan Schulz**

Department of Medical Informatics  
Universitätsklinikum Freiburg  
Stefan-Meier-Str. 26, D-79104 Freiburg, Germany  
stschulz@uni-freiburg.de

**Udo Hahn**

Text Knowledge Engineering Lab  
Universität Freiburg  
Werthmannplatz 1, D-79085 Freiburg, Germany  
hahn@coling.uni-freiburg.de

## Abstract

Modeling complex compositional objects in the field of the life sciences requires to solve intricate ontological problems, especially those related to parts of a whole, space and location. For concrete physical entities, we here stipulate that the distinction between parthood and location can be abstracted away. We outline a logic-based reasoning framework in which taxonomic subsumption not only takes care of the propagation of roles across taxonomic hierarchies, but also across nontaxonomic, i.e., partonomic or spatial, ones.

**Keywords:** Ontologies, Knowledge Representation, Formal Reasoning

## Introduction

The rapid increase of research activities in the health sciences (medicine), cell and molecular biology, genomics, etc. has created a surge of descriptive data. The knowledge representation and formal reasoning community, however, is only hesitatingly opening its mind for a new set of challenging issues from this relevant application domain. For instance, biology and medicine both deal with complex physical entities (and processes) whose aggregate structure needs inquiry into the composition of wholes and their associated parts, ranging from molecules and subcellular entities up to tissues, organs and organisms. In the following, we refer to the physical components of organisms using the term *biological structure*.

The pivotal role of biological structure is evidenced by

- the cellular component branch of GO<sup>1</sup> (Gene Ontology Consortium 2001), a symbolic model of species-independent relations between cell components;
- FMA<sup>2</sup>, the Foundational Model of Anatomy (Rosse *et al.* 1998; Rosse & Mejino 2004), which deals with the canonical anatomy of the adult human, with a focus on macroscopic anatomy;
- the animal and plant anatomies hosted by OBO, an open access umbrella system of structured biological vocabu-

laries for, e.g., the anatomy of the mouse<sup>3</sup>, the drosophila fly<sup>4</sup>, and grains<sup>5</sup>;

- the anatomy schemata of the GALEN CORE model (Rector *et al.* 1994);
- the SNOMED CT (2004) anatomy branch;
- the hierarchy of top-level anatomy concepts in the Semantic Network of the Unified Medical Language System (UMLS 2003).

In spite of their high degree of domain heterogeneity, nearly all of these controlled vocabularies use a bipartite hierarchical structure, with two kinds of partial orders, viz. *taxonomies* (characterized by the *is-a* relation, which associates specific with general concepts), and *partonomies* (associating parts and wholes by mereologic relations, such as *part-of* and *has-part*).

While reasoning in taxonomic hierarchies is quite well understood – each instance of a specific concept class is also an instance of any more general class, and, thus, inherits all its properties (Patel-Schneider 1991) – we lack an equal form of consensus for part/whole-related reasoning (for a survey, cf. Artale *et al.* (1996)). The embedding of part/whole-related reasoning into more general mereological reasoning, as well as the interactions between mereological and topological reasoning (so-called mereotopological reasoning), i.e., the relation between parts and wholes and the space they occupy, are pressing and difficult issues from the perspective of knowledge representation (Casati & Varzi 1999). Still the phenomenon of property inheritance can be found in mereotopological hierarchies as well. We naturally classify a “fracture of the neck of the femur” as being a “fracture of the femur”, simply because the neck of the femur is a part of the femur (Horrocks, Rector & Goble 1996).

The underlying general principles, however, are still under scrutiny and need to be carefully worked out, both at the epistemological and formal levels of knowledge representation.

Another phenomenon to be observed in biomedical domain descriptions is that the same description formalism is

Copyright © 2004, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

<sup>1</sup><http://www.godatabase.org>

<sup>2</sup><http://sig.biostr.washington.edu/projects/fm>

<sup>3</sup><http://www.informatics.jax.org>

<sup>4</sup><http://www.flybase.org>

<sup>5</sup><http://www.gramene.org>

used for both taxonomies and paronomies. Whereas it is methodically well-founded to relate two concepts by an *is-a* link (which is interpreted as class inclusion or taxonomic subsumption), such as *Femur is-a Bone*, it remains quite unclear how to interpret a mereological relation which holds between two concept classes, such as *Neck-of-Femur part-of Femur*, or *Bone has-part Calcium*.

Up until now, none of the present models of biological structure has been able to convey an ontologically founded, semantically precise and uncontroversial account of both taxonomy and paronomy (Aitken, Webber & Bard 2004). This poses severe problems for inference engines, in which both reasoning patterns have to be combined.

In our previous research on biomedical knowledge representation (Hahn, Schulz & Romacker 1999; Schulz & Hahn 2001; 2002a; 2002b) we have already addressed some of these issues in isolation. In this paper, we provide a coherent account of modeling parthood in biological structure in terms of spatial inclusion. Based on the discussion of the semantics underlying part-whole relations for concrete physical entities, we stipulate that the relations between wholes and their associated parts can be mirrored to spatial relations and that subsumption-based reasoning patterns can be reused to propagate concept roles across mereotopological hierarchies.

## Semantics of Part-Whole Relations

For the conceptualization of generalized biological structure, the mereological relation *part-of* (together with its inverse relation, *has-part*) requires a principled account under the following aspects:

1. There must be a clear commitment to the *algebraic foundation* of mereological relations in terms of transitivity, reflexivity and symmetry.
2. Considering the *semantics* of mereological relations, we have to make explicit whether parts and wholes should be understood in a functional or in a topological sense.
3. There should be a clear commitment to either an *open* or *closed world semantics*.
4. The wide-spread use of the relations *part-of* and *has-part* between *concepts* seems intuitive at a first sight, but it contrasts with the classical approach to mereology (Simons 1987) which focuses on *individual entities*, not classes of entities or concepts<sup>6</sup>.

None of the conceptual models of human anatomy or biological structure make sufficient claims regarding these four controversial issues. FMA and GO have a clear commitment to item 1, at least regarding the transitivity property of the general *part-of* relation. The same applies to GALEN with regard to *part-of* subrelations (Rector *et al.* 1994). This view is mainly consistent with classical (i.e., axiomatic) mereology (Simons 1987; Casati & Varzi 1999) which treats generic parthood as reflexive, antisymmetric and transitive.

<sup>6</sup>In this paper we use the term “concept” as a synonym of “class”.

Common conceptualizations in the biological domain, however, suggest that the assumption for *part-of* to be reflexive must be abandoned<sup>7</sup>. We therefore interpret *part-of* in the sense of *proper-part-of* in classical mereology.

A rudimentary commitment to item 2 is reflected by the introduction of several *part-of* subrelations in FMA and GALEN. A closed-world semantics (item 3) underlies the representation language GRAIL (Rector *et al.* 1997) used in GALEN. To the best of our knowledge, no other biomedical domain model makes any statement on this issue.

Finally, the use of mereological associations between pairs of concepts (item 4), such as *part-of/CellNucleus, Cell*, is so ambiguous that conflicting interpretations are likely to evolve: The Gene Ontology (Gene Ontology Consortium 2001) interprets *part-of* in a very loose way (“*part-of* means [something] can be a part of, not [something] is always a part of”) which frequently leads to unwarranted conclusions (Smith, Williams & Schulze-Kremer 2003). In contrast, the Foundational Model of Anatomy (FMA) (Rosse *et al.* 1998) conceptualizes *part-of* in a very strict manner: *A part-of B* means that any instance of *B* has an instance of *A* as part, and any instance of *A* is part of an instance of *B* (Smith & Rosse 2004). As far as other models of organisms are concerned, there is no clear commitment at all to the proper semantics of *part-of*.

The FMA interpretation, however, imposes a mutual dependency between parts and wholes and, therefore, may be too rigid in many cases. As an example, we may want to express that any instance of a *Cell Nucleus* is part of a *Cell*, but that not any instance of a *Cell* has a *Cell Nucleus*. Or consider the integration of pathological or surgical anatomy: Each instance of *Appendix* is part of an instance of *Digestive System* but not each instance of *Digestive System* has an instance of *Appendix* as part.

Additionally, we have to define which concepts can be associated by *part-of* or *has-part*, and which ones cannot. An instance of *Hemoglobin* may be part of an instance of *Cell*, but an instance of *Brain* can never be a part of an instance of *Cell*. This kind of consideration may sound strange from a human reasoning point of view, but it is crucial for correct automated reasoning. A mereological relation between concepts, therefore, cannot be interpreted unambiguously, unless we clarify the dependency status of the whole with respect to its parts, as well as the dependency status of the parts with respect to their whole. Furthermore, instances of two concepts *A* and *B* may be related by *part-of* or *has-part* even in case there is no dependency at all between the corresponding concepts.

In conclusion, for any pair of concepts, *A* and *B*, we make the following distinctions regarding parthood:

1. (*One-sided*) *Part-Whole Dependency*: Any instance of *A* is part of an instance of *B*, but there are instances of *B* which do not have instances of *A* as a part.
2. (*One-sided*) *Whole-Part Dependency*: Any instance of *B* has some instance of *A* as part, but there are some in-

<sup>7</sup>Otherwise, any instance of “stomach” would be an instance of “stomach part”, with the consequence that the concept “*partial resection of stomach*” would subsume “*total resection of stomach*”.

stances of  $A$  which are not part of an instance of  $B$ .

3. *Mutual Mereological Dependency*: Any instance of  $A$  is part of an instance of  $B$ , and any instance of  $B$  has some instance of  $A$  as part.
4. *Mereological Independency*: There is at least one instance of  $A$  which is part of some instance of  $B$ , and there is at least one instance of  $B$  which has an instance of  $A$  as part.
5. *Mereological Disjointness*: There are no instances of  $A$  which are parts of instances of  $B$ , and, conversely, there are no instances of  $B$  which have instances of  $A$  as part.

Interpreted in a strict sense, the five categories are mutually disjoint, i.e., category 3 is not a special case of either 1 or 2, nor does category 4 include the first three ones (Table 1 provides concrete examples from different domains).

Dependency Type	$A$ (part)	$B$ (whole)
Part-Whole Dependency (one-sided)	Cell Nucleus	Cell
	Chlorophyll	Organism
	Prostate Tumor	Prostate
	Touchpad	Laptop
Whole-Part Dependency (one-sided)	Sulfur	Methionin
	Wing	Chicken
	Heart	Drosophila
	Wing	Aircraft
Mutual Mereological Dependency	Cell Membrane	Cell
	Heart Ventricle	Heart
	Vertebra	Vertebrate
	Table Surface	Table
Mereological Independency	Uterus	Mammal
	Sulfur	Amino Acid
	Tooth	Human
	Headlight	Bicycle
Mereological Disjointness	Brain	Cell
	Wing	Mouse
	Sulfur	Alanin
	Wing	Human

Table 1: Examples of Ontological Dependencies Between Parts and Wholes

Mereological independency can be regarded as the default situation assuming an open world (any assertion is possible unless explicitly obviated), whilst mereological disjointness acts as the default under a closed-world assumption (unless ‘positive’ information is available). The distinctions we just outlined are by no means specific for mereological relations only. Rather, all non-taxonomic relations between concepts lead to such considerations when analyzed on the ontological level.

In order to express these concept-to-concept relations we here introduce axiomatically the concept-to-concept relations *Part-Of*<sup>8</sup> for the case of part-whole dependency, and *Has-Part* for the whole-part dependency.

<sup>8</sup>We use upper case relation names for relations between concepts.

We formalize these two relations by introducing, in a similar way as proposed by Smith & Rosse (2004), the irreflexive, non-transitive and antisymmetric relation *instance-of* (abbreviated as  $\iota$ ), which represents *concept class membership*. On this basis, we define *Is-A* as taxonomic subsumption, a reflexive, transitive, and antisymmetric relation between concepts  $A$  and  $B$ , as follows:

$$Is-A(A, B) =_{def} \forall x : (\iota(x, A) \rightarrow \iota(x, B)) \quad (1)$$

We now define the concept-to-concept relations *PartOf* and *HasPart* on the basis of the mereological base relations *has-part* and *partof*:

$$Part-Of(A, B) =_{def} \quad (2)$$

$$\forall x : \iota(x, A) \rightarrow \exists y : (\iota(y, B) \wedge part-of(x, y))$$

$$Has-Part(A, B) =_{def} \quad (3)$$

$$\forall x : \iota(x, A) \rightarrow \exists y : (\iota(y, B) \wedge has-part(x, y))$$

In contrast to the strictly disjoint categories in Table 1 *Part-Of* and *Has-Part* also cover the case of mutual mereological dependency (Mmd):

$$Mmd(A, B) =_{def} Loc(A, B) \wedge Inc(B, A) \quad (4)$$

Finally, we define mereological disjointness (Md) as follows:

$$Md(A, B) =_{def} \forall x, y : \iota(x, A) \wedge \iota(y, B) \rightarrow \neg \exists z : (part-of(z, x)) \wedge part-of(z, y) \quad (5)$$

## Mereology vs. Spatial Inclusion

Not only in biology is it difficult to convene upon non-controversial criteria for a clear distinction between part-hood and spatial inclusion. Let us take the example of an industrial plant. Buildings have multiple spatial subdivisions, there are a variety of physical objects located in these buildings, such as machines, transport systems. Warehouses store products, tanks are filled with chemicals, and persons and vehicles are going in and out. Even if we ignore the time aspect and merely analyze the static ontological relationships between these objects, there will be no controversy that the buildings are part of the industrial plant, but neither the persons working in it, nor the air circulating in its rooms, nor the fuel in the tanks are considered as reasonable parts of it. But what about the machines, tools and instruments, stationary or mobile ones, used for manufacturing, vehicles going in and out (e.g., trucks)? What about the products stored in the warehouses, and what about the hollow spaces in the rooms?

Biology faces similar and even more intricate problems, ontologically speaking. To a much higher degree than observed with artifacts, biological entities are characterized by a continuous exchange of matter with their environment. Water, glucose, amino acids flow in and proteins, hormones, and detritus flow out. Certain kinds of cells (so-called killer cells) ingest other organisms, digest them and then re-assemble their components, cf. Fig. 1. Other cells resemble powerful chemical plants with high output of mucus, hormones, milk, saliva etc, cf. Fig. 2.

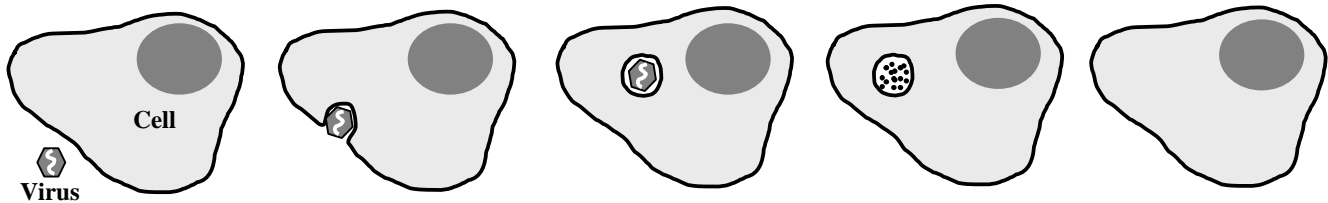


Figure 1: Phagocytosis of a Virus by a Cell

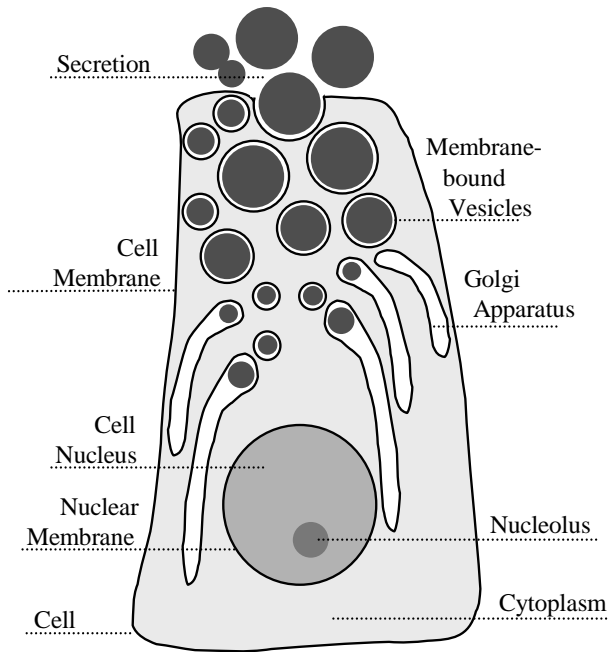


Figure 2: Prototypical Cell from a Salivary Gland

In contrast to artifacts, there is no clear distinction between a system's constituents and its substrates. As far as an engine is concerned, the material of which it consists is quite distinct from the fuel it consumes and the exhaust-gas it produces. Considering an organism, a carbon atom (as part of, e.g., a carbohydrate) ingested by food may play very different roles: It may merely participate in the body's "power supply", after which it is eliminated by the lungs as part of a  $\text{CO}_2$  molecule, or it may be integrated into the body's constitutional structure, e.g. in a muscle. In the latter case it may, in a further phase of the organism's lifecycle, change its role again, since nearly any constitutional matter of an organism can be metabolized ("burned") in order to assure the power supply of the organism, such as in case of hunger or chronic illness.

Organisms interact with other organisms as well. If a virus is in a cell, it is certainly not part of this cell. At which stage of the process of digestion (phagocytosis) do virus or bacteria components become parts of the ingesting cell (cf. Fig. 1)? Considering the cells which produce secretions (Fig. 2), are these substances part of those cells or are they merely located within them? Many tissues such as

the intestinal mucosa or the endometrium undergo permanent renovation, i.e. discharge of cells. Are these cells still part of the original tissue or not?

Besides static components such as fibers, membranes, epithelia, and blood vessels, functional tissues contain all kinds of mobile cells from the blood and the immune system, as well as proteins, hormones and nutrients floating around. How shall we describe these facts in terms of *part-of*? Are the red blood cells in the liver capillaries part of the liver? Are the immunocompetent cells which have migrated from the blood into the connective tissue part of the connective tissue?

Biological tissues and organs are further characterized by a multitude of hollow spaces, holes, cavities, grooves, lumina, vessels and ducts. Is the lumen of the intestine part of the intestine? Is a hollow space such as the cranial cavity part of its host, viz. the skull? We have defended the latter stipulation in (Schulz & Hahn 2001), because this conceptualization is common in the life sciences community (but not uncontroversial (Donnelly 2004)). But this is certainly not tenable from the viewpoint of formal topology, where a hollow space in an object is part of the exterior space but not of its host (Casati & Varzi 1999).

While there is a strong tradition in medicine to conceptualize *ideal* anatomy (the generic human body, as in the Foundational Model of Anatomy (Rosse *et al.* 1998)), additional problems arise when one tries to describe *pathological* anatomy in terms of mereology. Consider the epistemological bias involving parthood in the following example: Is a liver tumor *part-of* a liver? The counterargument proceeds as follows: The kinds of cells that constitute the tumor are different from those normally occurring in the healthy liver. The argument in favor proceeds along the following line: The tumor cells originated from liver cells – in the case of a *primary* tumor. There may be consensus, however, that a *metastasis* which develops from malignant tissue from outside the liver (e.g., from the colon) is *not* part of the liver. Another example of the difficulties of reaching an agreement on parthood is the *fecundation process*: It begins with the contact between a spermatozoon and an oocyte, leads to their fusion, which is followed by the formation and the merge of male and female pronuclei. Then the chromosomes are re-arranged, a zygote is formed and, finally, its cleavage is initiated. In which phase of this process which components of the original oocyte cease to be part of the female organism where the oocyte has developed?

As a result of this discussion we state that the semantic assumptions underlying parthood in the biomedical domain

are difficult to be objectified. If we subscribe, however, to a completely topological account of mereology, it is likely to get a much better agreement, as topological inclusion is based upon strictly geometric criteria.

In order to emphasize this distinction we introduce *loc* (*has-location*) and its converse *inc* (*includes*) as transitive and antisymmetric relations in the same way as we already did for *part-of* and *has-part*. The relation *loc* corresponds to the relation *WL* introduced by Casati & Varzi (1999), except for reflexivity which we exclude for the same reasons as we did for *part-of*.

Spatial inclusion has a strict point-set theoretic semantics, and we assume the extension of spatially relevant solid objects to include the objects' hollow spaces, i.e. those spaces within the convex hull of the object which can be considered a "fillable discontinuity of the object's surface" (Casati & Varzi 1999). Otherwise, all space regions located, e.g., within the blood vessels of an organ, or within the bronchi of the lung would be located in the exterior space – a stipulation which fundamentally violates common conceptualizations in the biomedical domain.

It is, however, not plausible to consider the complete convex hull of a biological object as spatially coinciding with this object. Under this assumption a whole body would nearly co-incide with the circulatory system (since its capillaries permeate all body parts), which is not acceptable at all.

Another example is depicted in Fig. 3. The convex hull of a (simplified) head includes a (dotted) region *C* between the tip of the nose and the chin. It would be implausible to consider this region as a part of the Head, whereas the "real" holes, the nostrils *A* and the mouth *B* are generally considered to be located inside the head. Theoretically, all of them are candidates for being included into the head, as there is no clear geometrical criterion for excluding *C*. Therefore, pragmatic and functional considerations must be involved for determining what is "inside" and "outside". In most cases, artificial spaces such as *C* are not even named because they are ontologically irrelevant.

By introducing the topological relations *loc* and *inc* we face the same difficulties when applying these relations to concept pairs. Analogously to the classification of mereological concept-to-concept associations (cf. Table 1) we here distinguish between includee-includer dependency, includer-includee dependency, mutual inclusion dependency, inclusional independency, and inclusional disjointness. In the same way we introduce the concept-to-concept relations *Loc* for includee-includer dependency and *Inc* for includer-includee dependency:

$$Loc(A, B) =_{def} \quad (6)$$

$$\forall x : \iota(x, A) \rightarrow \exists y : (\iota(y, B) \wedge loc(x, y))$$

$$Inc(A, B) =_{def} \quad (7)$$

$$\forall x : \iota(x, A) \rightarrow \exists y : (\iota(y, B) \wedge inc(x, y))$$

The consequences of conflating the notions of parthood and location have to be carefully traded against the gain in unanimity we stated above:

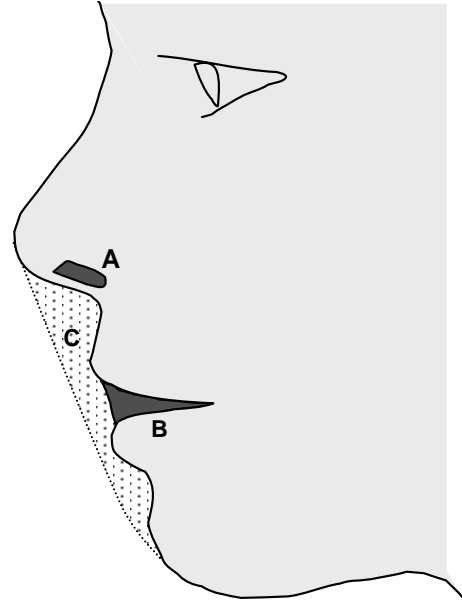


Figure 3: A Face, Its Convex Hull (*C*, dotted), and "Real" Holes (*A* and *B*)

- Even with the above stipulations, controversies may arise about which parts of an object's convex hull are considered to be "inside" or "outside". As an example, compare an open hand with a closed fist. When is an object located "on" the hand, when is it "in" the hand?
- One can conceive objects which are part-of something without being located in it. Examples are removed tissue samples, amputated body parts, cut hairs, cast skins of a reptile, leaves fallen from a tree etc.
- Confusion may arise when we describe an organism which contains another organism (e.g. an embryo, a parasite, or a prey).

An in-depth discussion of possible solutions goes beyond the scope of this paper. One may re-establish a non-controversial notion of "real" parts, based on non-discretionary assumptions. One criterion could be the previously introduced notion of "mutual location":

$$Real-Part-Of(A, B) =_{def} Loc(A, B) \wedge Inc(B, A) \quad (8)$$

under the additional assumption that all instances of *A* and *B* are solid objects. Thus one specifically excludes those objects which are located only temporarily in an organism (e.g. a prey organism in the stomach of a predator).

Another criterion for "real" parthood could be the ontogenetic derivation of a body part. Under this assumption any conglomerate of cells would be part of a given organism if they evolved from the same root.

A further investigation of these criteria – necessarily involving temporal aspects – would be desirable but is beyond the scope of this paper.

## Propagation of Attributes through Mereotopological Hierarchies

So far, we have discussed the semantics of mereological relations between concepts and analyzed the consequences of conflating parthood and spatial inclusion. We will now proceed with the analysis of another phenomenon, *viz.* the propagation of properties between parts and wholes (Rogers & Rector 2000; Rector 2002; Horrocks & Sattler 2003). This reasoning pattern, which can be paraphrased as “*x, which is related to y, is related to z as well, because y is a part of z*”, is equivalent to the following:

$$\forall x, y, z : rel(x, y) \wedge part-of(y, z) \rightarrow rel(x, z) \quad (9)$$

There are numerous examples of this reasoning pattern in biomedicine. For instance, the process of *Insulin Production* is usually considered a *Pancreas* function, because *Pancreatic Beta Cell* are considered *Part-Of* the *Pancreas*. In the same way, *Muscular Movement* would be classified as a *Muscle Function*, since it is a function of the *Actin-Myosin Complex* which is a *Part-Of Muscle Cells*, the latter being *Part-Of Muscle*. Still, there are counterexamples: *Amputation of Toe* cannot be subsumed by *Amputation of Foot* although every *Toe* is *Part-Of* a *Foot* (it might be classified, however, as *Amputation At 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. *DNA Replication* is not subsumed by *Cell Replication* although *DNA* is located in a *Cell*. 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 representation (Rector 2002).

Although computational models (such as the description logics GRAIL (Rector *et al.* 1997) and *RTQ* (Horrocks & Sattler 2003)) have been developed which support role inclusion axioms of this type, there is ample evidence that these inclusion axioms do not always hold (Hahn, Schulz & Romacker 1999). For example, *Gastroenteritis* stands for an inflammatory process of a *Digestive tract* as a whole, and therefore, it does not subsume *Appendicitis*, which refers to a local inflammation. On the other hand, *Nephritis* is usually considered as an inflammation of a *Kidney* or any of its parts and, therefore, subsumes *Glomerulonephritis* which is the inflammation of the *Glomerula*, which are *Part-Of Kidney*.

Rather than introducing additional relations whose propagation properties are clearly defined (such as *inflammation-of<sub>1</sub>*, vs. *inflammation-of<sub>2</sub>*) we have preferred, in our previous work (Hahn, Schulz & Romacker 1999; Schulz & Hahn 2002b) to address separately the concept class of “a whole”, on the one hand, and the concept class of “a whole or any of its constituent parts”. We argued that the distinction between these two concepts of entities is epistemologically valid in the biomedical domain at least, thus justifying the introduction of additional concepts into the ontology.

In the following, we introduce a logic-based solution to this problem:

## A Formal Model of Parthood as Spatial Inclusion

In this formalization, we completely abandon the mereological relations *has-part* and *part-of* (together with their counterparts *Part-Of* and *Has-Part*) and rather treat parthood as spatial inclusion, using *loc* (*has-location*) and its inverse relation *inc* (*includes*). For a concept *A* we introduce two reifier nodes, one for the relation *inc* and another for the relation *loc* (cf. Fig. 4):

$$\forall x : \iota(x, A_{loc}) \rightarrow \exists y : (\iota(y, A) \wedge loc(x, y)) \quad (10)$$

$$\forall x : \iota(x, A_{inc}) \rightarrow \exists y : (\iota(y, A) \wedge inc(x, y)) \quad (11)$$

Preparing the example depicted in Fig. 4 we add another two concepts, *B*, and *C*, together with the corresponding reifier nodes:

$$\forall x : \iota(x, B_{loc}) \rightarrow \exists y : (\iota(y, B) \wedge loc(x, y)) \quad (12)$$

$$\forall x : \iota(x, B_{inc}) \rightarrow \exists y : (\iota(y, B) \wedge inc(x, y)) \quad (13)$$

$$\forall x : \iota(x, C_{loc}) \rightarrow \exists y : (\iota(y, C) \wedge loc(x, y)) \quad (14)$$

$$\forall x : \iota(x, C_{inc}) \rightarrow \exists y : (\iota(y, C) \wedge inc(x, y)) \quad (15)$$

With regard to *loc* and *inc*, we distinguish between the same dependency patterns as the ones we initially introduced for the parthood relations (*A* and *B* denote concepts):

- We can now describe *Includee - Includer Dependency* by a simple taxonomic link, thus eliminating the need for the *Loc* predicate defined in formula (6):

$$Is-A(A, B_{loc}) \quad (16)$$

- *Includer - Includee Dependency* is described analogously:

$$Is-A(B, A_{inc}) \quad (17)$$

- We now have a formal basis for the description of *Mutual Inclusion Dependency*:

$$Is-A(A, B_{loc}) \wedge Is-A(B, A_{inc}) \quad (18)$$

- *Inclusion Independency* is the default assumption if an open world semantics is assumed, unless existential conditions hold (see above) or the range of allowed values is restricted.
- *Inclusion Disjointness* is defined analogously to mereological disjointness (cf. Formula 5)

For instance, we express mutual inclusion dependencies between *A*, *B*, and *C* by the following axioms (cf. Figure 4 for an example using the concepts *Nuclear Membrane* for *A*, *Cell Nucleus* for *B*, and *Interphase Eukaryotic Cell* for *C*). *A* and *B* are linked by

$$Is-A(A, B_{loc}) \quad (19)$$

$$Is-A(A_{loc}, B_{loc}) \quad (20)$$

$$Is-A(B, A_{inc}) \quad (21)$$

$$Is-A(B_{inc}, A_{inc}) \quad (22)$$

*B* and *C* are linked by

$$Is-A(B, C_{loc}) \quad (23)$$

$$Is-A(B_{loc}, C_{loc}) \quad (24)$$

$$Is-A(C, B_{inc}) \quad (25)$$

$$Is-A(C_{inc}, B_{inc}) \quad (26)$$

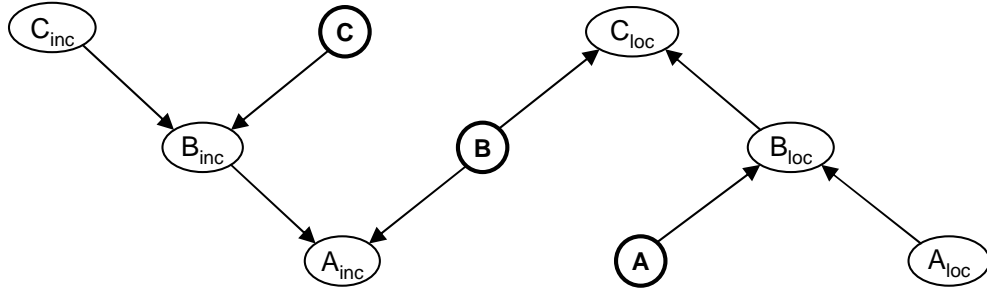


Figure 4: Includer / Includee Encoding Scheme. Arrows depict Taxonomic Links. For Example,  $A = \text{Nuclear Membrane}$ ,  $B = \text{Cell Nucleus}$ ,  $C = (\text{Interphase}) \text{ Eukaryotic Cell}$

In this manner the following properties (see formulae above) are propagated via taxonomic subsumption:  $A$  inherits the properties (12) and (14),  $B$  inherits (11) and (14), and  $C$  inherits (11) and (13). Note that the situation given by these examples represents the case of mutual inclusion dependency. Of course, we are free to express unilateral dependencies (In our example, the link between  $C$  and  $B_{inc}$  must be removed if  $C$  denotes the general concept *Cell*, since cells without nuclei exist.).

The resulting complex graph is still acyclic with respect to the  $Is-A$  relations which express *inc* and *loc* hierarchies, given that there are no cycles in either of these subgraphs.

By the subsumption of  $X_{loc}$  nodes, inclusional disjointness can be expressed on a high level, and is propagated through all subsumed nodes. As an example, the axiom

$$\forall x, y : \iota(x, Trunk) \wedge \iota(y, Head) \rightarrow \neg \exists z : loc(z, x) \wedge loc(z, y) \quad (27)$$

precludes that any object can be simultaneously located in both a *Trunk* and a *Head*. As a consequence, this also holds for any subdivisions of *Trunk* and *Head*, e.g., *Liver* and *Eye*, because  $Is-A(Liver, Trunk_{loc})$  and  $Is-A(Eye, Head_{loc})$ .

Summing up, we express the different forms of includer / includee relationships by means of taxonomic subsumption. Hereby it is not even necessary to use the transitivity property of these relations in order to get the required inferences. This may matter in certain representation formalisms (e.g. the standard description logics  $\mathcal{ALC}$ , which does not support transitive roles).

We will now illustrate how the proposed formalism is capable of accounting for various forms of mereotopological inferences.

“A *nephritis* is an inflammation located at a kidney as a whole or at any of those objects which are necessarily located in the kidney.”

$$\forall x : \iota(x, Nephritis) \rightarrow (\iota(x, Inflammation) \wedge (\iota(x, Kidney_{loc}))) \quad (28)$$

Since  $Kidney_{loc}$  subsumes, e.g.,  $Glomerulum_{loc}$  (according to the encoding principles for compositional hierarchies), *Glomerulonephritis* is correctly subsumed by *Nephritis*, given

$$\forall x : \iota(x, Glomerulonephritis) \rightarrow (\iota(x, Inflammation) \wedge (\iota(x, Glomerulum_{loc}))) \quad (29)$$

and

$$Is-A(Glomerulum_{loc}, Kidney_{loc}) \quad (30)$$

The relation *loc* is used both for the spatial inclusion of *Glomerulum* into *Kidney* and for the relation between the process *Glomerulonephritis* and *Glomerulum*.

“Mitosis is a biological division process which targets and is located in the nucleus of eukaryotic cell.”

$$\begin{aligned} \forall x : \iota(x, Mitosis) \rightarrow & (\iota(x, BiologicalDivision) \wedge \\ & \iota(x, CellNucleus_{loc}) \wedge \\ & \exists y : \iota(y, CellNucleus) \wedge \\ & targets(x, y)) \end{aligned} \quad (31)$$

The relation *targets* addresses a spatial structure, not in “neutral” terms of location where a process takes place, but rather as an object which is modified in the course of this process. With *BiologicalDivision* being subsumed by *Process*, we may infer that Mitosis is a *CellProcess* (equivalent to  $Process \wedge Cell_{loc}$ ), given  $Cell_{loc}$  subsumes  $CellNucleus_{loc}$ . Rightly, we fail to infer that it is, e.g., a liver process, because we cannot infer  $Liver_{loc}$ , although, by way of the Includer-Includee dependency, we may state that *Liver* is an *EukaryoticCell\_{inc}*. In other words, any instance of *Liver* implies the existence of an instance of *EukaryoticCell*, but not every instance of *EukaryoticCell* implies the existence of an instance of *Liver*.

“Insulin production is located in the pancreatic beta cells. Beta cells are included in Langerhans Islets, which are included in the pancreas.”

$$Is-A(InsulinProduction, BetaCells_{loc}) \quad (32)$$

From this and

$$Is-A(BetaCells_{loc}, LangerhansIslets_{loc}) \quad (33)$$

$$Is-A(LangerhansIslets_{loc}, Pancreas_{loc}) \quad (34)$$

we infer that *InsulinProduction* is located at *Langerhans-Islets*, as well as at the *Pancreas*.

“*Amputation of a foot is an amputation which targets a foot and is located at a foot.*”

$$\begin{aligned} \forall x : \iota(x, \text{AmputationOfFoot}) \rightarrow & \quad (35) \\ (\iota(x, \text{Amputation}) \wedge \iota(x, \text{Foot}_{loc}) \wedge \\ \exists y : \iota(y, \text{Foot}) \wedge \text{targets}(x, y)) \end{aligned}$$

“*Amputation at a foot is an amputation which is located at a foot.*”

$$\begin{aligned} \forall x : \iota(x, \text{AmputationAtFoot}) \rightarrow & \quad (36) \\ (\iota(x, \text{Amputation}) \wedge \iota(x, \text{Foot}_{loc})) \end{aligned}$$

Given *Is-A*(*Toe<sub>loc</sub>*, *Foot<sub>loc</sub>*), as a consequence, *AmputationAtToe* is an *AmputationAtFoot*, but not an *AmputationOfFoot*, because the target role does not propagate. Note the subtle but important semantic distinction between “at” (location) and “of” (target).

The already discussed distinction in propagation patterns between *Nephritis* and *Gastroenteritis* can be explained by the same terms. Whilst *Nephritis* is an inflammation “at”, *Gastroenteritis* is not only an inflammation “at” but also an inflammation “of”. This way, specialized roles such as *inflammation-of*, *fracture-of*, or *amputation-of* can easily be reduced to a small set of universal thematic roles, such as *targets* and *loc*.

It is also possible to use the relation *inc*, or the corresponding reificator node *X<sub>inc</sub>* for inferences on biological function. We assume *biological function* to be an inherent ability of biological objects. It means that the specific function is present in full for every instance of this concept. The presence of a function does however not mean that the function gets exercised at any moment.

The following (simplified) example demonstrates how function propagates from parts to wholes. Again, we reduce the inference to taxonomic subsumption, i.e. the transitive closure of *inc* is expressed as taxonomic links between *inc* nodes. We reformulate the pancreas / beta cell example under the aspect of cell function:

“*Pancreatic beta cells produce insulin. An intact pancreas includes beta cells. There are no beta cells in a diabetic pancreas. We infer that an intact pancreas possesses the function of producing insulin. We cannot infer this for a diabetic pancreas.*”

$$\begin{aligned} \forall x : \iota(x, \text{BetaCells}) \rightarrow & \quad (37) \\ \iota(x, \text{InsulinProduction}_{inc}) \end{aligned}$$

$$\begin{aligned} \forall x : \iota(x, \text{BetaCells}_{inc}) \rightarrow & \quad (38) \\ \iota(x, \text{InsulinProduction}_{inc}) \end{aligned}$$

$$\begin{aligned} \forall x : \iota(x, \text{IntactPancreas}) \rightarrow & \quad (39) \\ \iota(x, \text{BetaCells}_{inc}) \end{aligned}$$

According to Formula (38), *IntactPancreas Is-A BetaCells<sub>inc</sub> Is-A InsulinProduction<sub>inc</sub>*, i.e. each *IntactPancreas* includes the function *InsulinProduction*. This is not the case with *DiabeticPancreas*:

$$\begin{aligned} \forall x : \iota(x, \text{DiabeticPancreas}) \rightarrow & \quad (40) \\ \neg \exists y : (\iota(y, \text{BetaCells}) \wedge \text{inc}(x, y)) \end{aligned}$$

## Conclusion

This paper covers two major issues. The first one is concerned with ontological considerations underlying part-whole relations. For physical domains at least, it seems possible to get rid of the ongoing debate about different forms of part-whole reasoning and its underpinnings, by replacing *part-of* / *has-part* relations with locational ones, viz. *has-location* (*loc*) and its inverse *includes* (*inc*). Substituting mereological relations by strict topological inclusion has the advantage of eliminating the difficult and controversial delimitation between the notions of spatial inclusion and generic parthood. In addition, we center bio-ontologies around a multi-purpose relation with an uncontroversial semantics, viz. *loc* (*has-location*).

The second issue has to do with parsimony of the formalization of these locational relations. The relation *loc* is used not only to express the spatial relationships between physical objects and spaces, but also between processes and physically defined structures. In the same way, *inc* relates not only physical objects, but also physical objects with their inherent functions. Thus we do not need additional formal language devices in order to obtain the inferences (propagation of properties across compositional hierarchies) needed in such a reasoning framework. This way, we combine ontological clarity with formal simplicity.

Reducing various types of inferences in compositional hierarchies to one single taxonomic subsumption computation step parallels – on the reasoning level – our claims for formal simplicity with respect to the definition of suitable relations. From a practical perspective, our reductionist approach allows us to reuse off-the-shelf terminological reasoning engines (such as LOOM (MacGregor 1994), FaCT (Horrocks 1998), and RACER (Haarslev & Möller 2001)). The capability of robust inference engines to deal with massive amounts of knowledge is a prerequisite for any serious application concern in the biomedical field, as we have demonstrated by assembling huge knowledge bases using description logics (Schulz & Hahn 2000; Beck & Schulz 2003).

Though we stipulate to replace somewhat “messy” part-whole relations by clear-cut locational ones, many challenging research questions and technical problems remain to be addressed. One problem concerns the prevalence of definitory cycles brought about by mutual concept dependencies. Another one refers to the fusion of parthood with location. Here, a thorough ontological inquiry of the consequences of this abstraction is still due. Last but not least, a large-scale empirical ontology evaluation effort must be started in order to assess whether the proposed formalization is beneficial for bio-medical reasoning tasks as part of intelligent information system functionality, e.g., within the context of information extraction or text mining from biomedical literature (Hahn, Romacker & Schulz 2002).

## Acknowledgements

We wish to thank Lucas Champollion and Kornél Markó for critical reading of the manuscript and fruitful discussions.



## References

- Aitken, J. S., Webber, B. L., and Bard, J. B. L. 2004. Part-of relations in anatomy ontologies: A proposal for RDFS and OWL formalisations. In *PSB 2004 – Proceedings of the Pacific Symposium on Biocomputing 2004*. Hawaii, USA, January 6-10, 2002. Singapore: World Scientific Publishing.
- 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.
- 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. Philadelphia, PA: Hanley & Belfus.
- Casati, R., and Varzi, A. C. 1999. *Parts and Places. The Structures of Spatial Representation*. Cambridge, MA: MIT Press/Bradford.
- Donnelly, M. 2004. On parts and holes: The spatial structure of the human body. In *MEDINFO 2004 – Proceedings of the 11th World Congress on Medical Informatics*.
- Gene Ontology Consortium. 2001. Creating the Gene Ontology resource: Design and implementation. *Genome Research* 11(8):1425–1433.
- Haarslev, V., and Möller, R. 2001. Description of the RACER system and its applications. In *Proceedings of the 2001 Description Logics Workshop*, 132–141.
- Hahn, U., Romacker, M., and Schulz, S. 2002. MEDSYNDIKATE: A natural language system for the extraction of medical information from finding reports. *International Journal of Medical Informatics* 67(1/3):63–74.
- Hahn, U., Schulz, S., and Romacker, M. 1999. Partonomic reasoning as taxonomic reasoning in medicine. In *AAAI'99/IAAI'99 – Proceedings of the 16th National Conference on Artificial Intelligence & 11th Innovative Applications of Artificial Intelligence Conference*, 271–276. Menlo Park, CA, Cambridge, MA: AAAI Press & MIT Press.
- 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*, volume 1, 343–348. San Francisco, CA: Morgan Kaufmann.
- Horrocks, I., Rector, A., and Goble, C. 1996. A description logic based schema for the classification of medical data. In *KRDB'96 – Proceedings of the 3rd Workshop Knowledge Representation Meets Databases*, 24–28.
- Horrocks, I. R. 1998. Using an expressive description logic: FACT or fiction? In *Principles of Knowledge Representation and Reasoning: Proceedings of the 6th International Conference – KR'98*, 636–645. San Francisco, CA: Morgan Kaufmann.
- MacGregor, R. M. 1994. A description classifier for the predicate calculus. In *AAAI'94 – Proceedings of the 12th National Conference on Artificial Intelligence*, volume 1, 213–220. Menlo Park, CA: AAAI Press & MIT Press.
- Patel-Schneider, P. F. 1991. What's inheritance got to do with knowledge representation? In *Inheritance Hierarchies in Knowledge Representation and Programming Languages*. Chichester etc.: John Wiley. 1–11.
- Rector, A., Gangemi, A., Elena, G., Glowinski, A. J., and Rossi-Mori, A. 1994. The GALEN model schemata for anatomy: Towards a re-usable application-independent model of medical concepts. In *MIE'94 – Proceedings of the 12th Conference of the European Federation for Medical Informatics*, 229–233. Amsterdam: IOS Press.
- Rector, A. L., Bechhofer, S., Goble, C. A., Horrocks, I., Nowlan, W. A., and Solomon, W. D. 1997. The GRail concept modelling language for medical terminology. *Artificial Intelligence in Medicine* 9:139–171.
- 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*. Published via <http://CEUR-WS.org/Vol-53/>.
- 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. Converging Information, Technology, and Health Care*, 714–718. Philadelphia, PA: Hanley & Belfus.
- Rosse, C., and Mejino, J. L. V. 2004. A reference ontology for bioinformatics: the Foundational Model of Anatomy. *Journal of Biomedical Informatics*. In press.
- 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 American Medical Informatics Association* 5(1):17–40.
- UMLS. 2003. *Unified Medical Language System*. Bethesda, MD: National Library of Medicine.
- Schulz, S., and Hahn, U. 2000. Knowledge engineering by large-scale knowledge reuse: Experience from the medical domain. In *Principles of Knowledge Representation and Reasoning. Proceedings of the 7th International Conference – KR 2000*, 601–610. San Francisco, CA: Morgan Kaufmann.
- Schulz, S., and Hahn, U. 2001. Mereotopological reasoning about parts and (w)holes in bio-ontologies. In *2nd International FOIS Conference*, 210–221. New York, NY: ACM Press.
- Schulz, S., and Hahn, U. 2002a. A knowledge representation view on biomedical structure and function. In *AMIA 2002 – Proceedings of the Annual Symposium of the American Medical Informatics Association*, 687–691. Philadelphia, PA: Hanley & Belfus.
- Schulz, S., and Hahn, U. 2002b. Necessary parts and wholes in bio-ontologies. In *Principles of Knowledge Representation and Reasoning. Proceedings of the 8th International Conference – KR 2002*, 387–394. San Francisco, CA: Morgan Kaufmann.
- Simons, P. 1987. *Parts: A Study in Ontology*. Oxford: Clarendon Press.
- 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*. San Francisco, CA, September 2004.
- 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. Philadelphia, PA: Hanley & Belfus.
2004. *SNOMED Clinical Terms*. Northfield, IL: College of American Pathologists.