Proceedings of the AMIA Symposium 2005, Washington DC, 669–673. PMC1560856

How to Distinguish Parthood from Location in Bio-Ontologies

Stefan Schulz [a,b] Philipp Daumke [a] Barry Smith [c,d] Udo Hahn [e]
[a]Department of Medical Informatics, Freiburg University Hospital, Germany
[b]Health Informatics Laboratory, Paraná Catholic University, Curitiba, Brazil
[c]Department of Philosophy, University at Buffalo, NY, USA
[d]Institute for Formal Ontology and Medical Information Science, Saarland University, Saarbrücken, Germany
[e] Jena University Language and Information Engineering (JULIE) Lab, Germany

Abstract

The pivotal role of the relation **part-of** *in the description of living organisms is widely acknowledged. Organisms are open systems, which means that in contradistinction to mechanical artifacts they are characterized by a continuous flow and exchange of matter. A closer analysis of the spatial relations in biological organisms reveals that the decision as to whether a given particular is* **part-of** *a second particular or whether it is only* **contained-in** *the second particular is often controversial. We here propose a rule-based approach which allows us to decide on the basis of well-defined criteria which of the two relations holds between two anatomical objects, given that one spatially includes the other. We discuss the advantages and limitations of this approach, using concrete examples from human anatomy.*

INTRODUCTION

Where the classical terminologies used in clinical practice are characterized by taxonomic relationships, the mereological (part-whole) organization of biomedical ontologies has attracted increasing interest in recent years. While taxonomies associate more specific with more general classes, commonly expressed by the relation *is-a*, partonomies associate parts to wholes, or classes of parts to classes of wholes, by means of mereological relations such as parthood, overlap and disjointness. Large partonomic structures can be found in several important biomedical ontologies [7, 12, 11, 1, 10]. Most ontology engineering in the life sciences thus far, however, has been realized in a rather intuitive manner and a cognitive, concept-centered approach was given priority, chiefly motivated by a concern for lexical relations between meanings. Certainly such thesaurus-like systems are sufficient for many applications, e.g. for querying of terms and for the semantic annotation of documents, but they are inadequate for more ambitious logic-based reasoning. Only in this decade have more principled studies, based on logic and formal ontology, investigated in detail the biomedical parthood relation [15] and concluded that the thesaurus-style usage of part-of relations is too informal and leads to unreliable results when more sophisticated reasoning is required. If the basic **part-of** relation is to be granted the status of a foundational relation [2] in ontology, a more formal approach is needed. As we shall show, however, assertions of parthood are tied to issues of human perception and belief, and argue that, because of the problems which then arise, the relation **located-in** would better serve as a foundational relation in biomedical ontologies. At the same time we are convinced that **part-of** nonetheless deserves to play a central role in descriptions of living organisms, and so we propose a cascading list of criteria which can be used for asserting **part-of** relations in different types of cases.

We refer throughout to relations which hold between individuals and not between classes. Class-level parthood and spatial relations are very common in biomedical ontologies (e.g. *Part-Of (Thumb, Hand)*), but they can be formalized only via an approach that is based on instance-level relations [14, 5].

A CASE STUDY

Alanine is a non-essential amino acid that is used by the body to build protein. Consider some alanine molecule in a human body. This molecule is, without doubt, *located* in the body. But is it also *part* of the body? Let us consider the following six scenarios as concerns the destiny of an alanine molecule:

1. It is ingested as food ingredient and excreted by feces without digestion.
2. It is ingested as food ingredient, digested and used for albumin synthesis, some albumin is then excreted with the urine.
3. It is ingested, metabolized and used for collagen synthesis, thereby becoming integrated into the structure of a bone.
4. It is synthesized in the liver, built into a hemoglobin molecule, and leaves the body by bleeding.
5. It is synthesized in the liver, built into a globulin molecule, and then catabolized in a cell.
6. It is present in the zygote and the early embryo, and then catabolized in the maternal organism.

Fig. 1 depicts these scenarios by means of space-time diagrams. The grey areas represent the life cycles of the respective organisms, the black areas the life cycles of the corresponding single alanine molecule in each of the different scenarios. Which scenarios allow

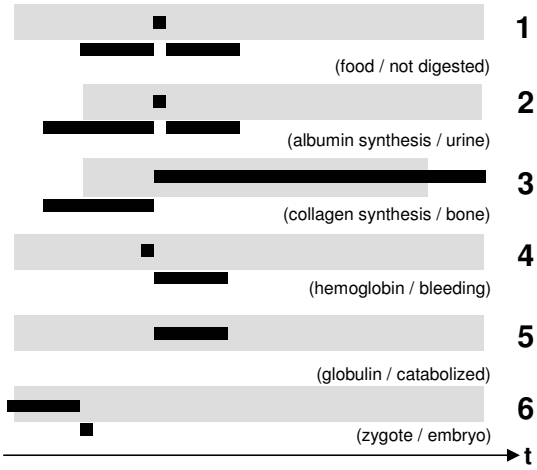

Figure 1: Location of an alanine molecule (black) within the organism (grey) over time

us to state that the alanine molecule is a part of the body? In scenario (1) we may argue that it is not a part but rather that it is only located in (i.e. contained in) a cavity of the body. In (2) one may argue that it is part at least for some time such as to form a part of some protein (albumin) which is itself commonly regarded as a body part. In (3) the molecule would most properly be seen as a constituent of the bone, and therefore again as a part of the body. In (4) and (5), similarly, because the molecule realizes a well-defined function in a protein, we may tend to classify it as a part. And so also in (6), where the molecule is included in relevant biological objects at the very earliest stages of their existence. Note that in none of these cases is the molecule an *essential* part. That is it can in each case be substituted by another object of the same kind, as contrasted for example to the case of the brain.

PARTHOOD AND LOCATION

In order to express instance-level parthood relations we follow the terminology in [15], using **bold face** for relations between instances and *italics* for relations between classes. We use the reflexive, antisymmetric and transitive relation **part-of**, which relates pairs of individual objects, as in:

$$\textbf{part-of}(myThumb, myHand, t),$$

which means that my thumb is a part of my hand at the time instant t. Indexing by time is necessary, since two entities may be related by parthood only in a certain phase of their simultaneous existence [8], as e.g. when

$$\textbf{part-of}(myThumb, myHand, t')$$

is not true for the instant t' because of amputation.
In a similar way we introduce location:

$$\textbf{located-in}(myBrain, mySkull, t)$$

means that at the time t my brain is located in my skull. To understand location formally, we associate with each physical object a spatial region and define a function r which assigns to each anatomical entity c and time t the corresponding spatial region $r(c,t)$. This spatial region is as a matter of definition exactly occupied by c at time t. We can then define the relation of location for anatomical entities as follows [6, 4]:

$$\textbf{located-in}(c,d,t) =_{def} \textbf{part-of}(r(c,t), r(d,t), t)$$

Trivially, by this definition, all parts of biological structures are **located-in** the corresponding wholes but not everything which is located somewhere in a biological structure is also **part-of** that structure.
The cases in question are cases of containment, which we define as follows:

$$\textbf{contained-in}(c,d,t) =_{def}$$
$$\textbf{located-in}(c,d,t) \wedge \neg\textbf{part-of}(c,d,t)$$

Containment is a relation which obtains between a material object (or portion of material substance such as blood or urine), and some immaterial cavity or body space. If the body space itself is part of a material object, then objects located in this space are also contained in that object, by our definition. Because biological objects are involved in a constant exchange of matter with their environment many location relationships are short-lived. Moreover, the continuous nature of the phenomena of matter exchange [3], as illustrated in the alanine example above, suggests that there may be relations intermediate between parthood and containment, of the type illustrated in the alanine example above.

CRITERIA FOR PARTHOOD

Under what condition, then, is an entity **located-in** another entity also **part-of** this entity? Is an embryo **part-of**, or merely **located-in**, a uterus? Is a bolus of food **part-of**, or merely **located-in**, a digestive tract? Is an oxygen molecule **part-of**, or merely **located-in**, a lung? We here offer four kinds of criteria which may be helpful in providing answers to such questions. These criteria were elicited after an in-depth analysis of 100 examples of medical and biological location done by two of the authors.
1. *Genetic identity*: We introduce the relation of genetic identity between two biological objects (**same-genetic-origin**(c,d)), and assert that one object is **part-of** another only if they stand in this relation. An embryo, on this criterion, is not a part of its mother's uterus. This criterion, which is favored for example by the FMA [12], faces problems in application to atoms as well as to those portions of bodily samples such as glucose, water, urea which are not gene products. A strict application of this rule would further mean that

mitochondria are not parts of cells (since they have their own DNA). This rule is also not sufficient to assert parthood. It is of no help, for example, in deciding whether a lymphocyte in a piece of tissue is part of this tissue or is merely contained therein, or if an insulin molecule attached to a receptor in a cell membrane is or is not a part of this cell membrane. We therefore state, taken into account that it is only applicable for c and d which have a genetic origin:

$$\textbf{located-in}(c,d,t) \wedge \neg\textbf{same-genetic-origin}(c,d) \rightarrow \textbf{contained-in}(c,d,t)$$

2. *Sortality*: If an object c is part of an object d, then c and d must be of appropriate sorts to make this possible, i.e. they must instantiate compatible classes. Thus if d is an instance of *Organism*, then it is ruled out that c should be an instance of an *artifact* (e.g., a heart pacemaker, a bullet, a dental filling), or that d should be a second whole organism (a symbiont, parasite, prey, embryo or fetus). Similarly, if d is a non-material object, then c cannot be a material object: my brain is not **part-of** my cranial cavity.

3. *Life Cycle*: Unless this is already ruled out by sortal constraints, c is part of d at t in the case that c is spatially included in d at t and there is no time at which c is not spatially included in d. For example, my brain stem is for the given reason part of my brain. (See also case 5 in Fig. 1.) More strictly, if c is located in d, then c is also part of d whenever d's existence without cease with the removal of c. For this reason my brain is part of my (living) body, and the surface of my body is part of my body). This phenomenon of ontological dependence is already covered by the above criterion. We should not, however, mix up this kind of dependence on the level of individuals with *generic dependence*. Thus every cell depends on some water molecules in its cytoplasm. This means that the cell cannot exist without some instances of the class *water molecule* though it does not of course depend on some one *particular* water molecule.

The main problem here is that such assertions of parthood require that the whole lifecycle has been recorded. Taking our example 6, this means that if the zygote dies and the alanine molecule is still contained in its cytoplasm then it would have been a **part-of** the cytoplasm, otherwise it would have been only **located-in** the cytoplasm.

In order to allow assertions during the existence of spatially related objects we add the precondition $\textbf{earlier}(t',t)$ and define the relation $\textbf{hitherto-located-in}(c,d,t)$ as follows:

$$\textbf{hitherto-located-in}(c,d,t) =_{def} \textbf{located-in}(c,d,t) \wedge \neg\exists t' : \textbf{earlier}(t',t) \wedge \neg\textbf{located-in}(c,d,t')$$

Thus, the parthood criterion is:

$$\textbf{hitherto-located-in}(c,d,t) \rightarrow \textbf{part-of}(c,d,t)$$

The transitivity of **part-of** allows us to expand this criterion to additional cases such as 3 in Fig. 1. Here, the alanine molecule is located in the bone during the whole period in which the bone exists. Since the bone is part of the body, the alanine molecule is part of the body as well.

4. *Functionality/Integrity*: The last criterion which helps us to specialize **located-in** to **part-of** concerns functionality and integrity. Again, we take two particular objects c and d, with c being **located-in** d. Let c have a functionality f which is essential to d in the sense that d becomes dysfunctional or even dies when f cannot be realized. For example, the pumping function of the heart is *essential* to the functioning of the cardiovascular system, just as the function of the collection of hepatocytes is specific and essential to the functioning of the liver. In contrast, the functioning of one individual macrophage which happens to be **located-in** some organ at some moment t is not essential to the functioning of that organ. For the same reason is the function of one given glucose molecule in a cell not essential to the functioning of that cell.

The functionality criterion faces problems in application to those parts of anatomical structures which are spatially included in others but are not essential to the proper functioning of the latter, or whose functional relevance has disappeared during evolution. Examples are: terminal hairs of the skin, nasal sinuses, or the kidney and other organs supplied to the body in pairs. A solution could be to expand the concept of *functioning* to that of *integrity* (or in other words, correspondence to the canonical constitution of the organism). If c **located-in** d then it would be **part-of** d whenever the removal of c would affect the integrity of d. The problem with such a definition is that it relies upon a clear criterion of what integrity is: it requires a reference to what is understood by a well-formed biological structure, e.g., a canonical description such as the Foundational Model of Anatomy [12].

A precise formalization would largely exceed the scope of this paper. We restrict ourselves to outline the functionality / integrity criterion as follows:

$$\textbf{function-integrity-relevant}(c,d,t) \wedge \textbf{located-in}(c,d,t) \rightarrow \textbf{part-of}(c,d,t)$$

A CLASSIFICATION ALGORITHM

Figure 2 depicts a semi-formal algorithm for the classification of a location relation in terms of parthood or containment, by taking account of the above criteria.

Let us take some examples and consider how they are treated by this algorithm:

```
If located-in (c, d, t)
   If Artifact(c) then
     contained-in(c, d, t)
   Else
     If function-integrity-relevant (c, d , t) then
       part-of (c, d, t)
     Else
       If not same-genetic-origin (c, d, t) or
                 (instance-of (c, Material) and
                  instance-of (d, Immaterial)) then
         contained-in (c, d, t)
       Else
         If hitherto-located-in (c, d, t) or
          (hitherto-located-in (c, m, t) and
           part-of (m, d, t)) then
             part-of (c, d, t)
         Else
             contained-in (c, d, t)
         End If
       End If
     End If
   End If
End If
```

Figure 2: Algorithm for specializing **located-in** to **contained-in** or **part-of**

Amalgam filling in a tooth: The filling is an artifact, hence parthood is discarded in the first decision step.

A transplanted lung in an organism: The lung is now functionally related to the organism, so parthood is accepted in the second decision step (in spite of the difference in genetic origin).

A fingernail in a finger: The fingernail has never been outside the finger, hence it is part of the finger.

A portion of urine in a bladder: The two are not functionally related (urine does not have any function which is also a function of the bladder), and nor does it have its origin in the bladder. Therefore the urine portion is contained in the bladder but it is not a part thereof.

A metastasis of a breast cancer in the brain: The metastatic tumor is not functionally related to the brain nor does it have its origin in the brain, since the cells from which the metastasis originated migrated from somewhere else into the brain. Therefore the origination criterion does not hold, and we conclude that the metastasis is contained in the brain but is not part of it.

A glioblastoma in the brain: Again, the tumor is not functionally related to the brain. But it has its origin therein, since the cells from which the metastasis originated themselves originated from the brain. Therefore the origination criterion can here be applied. The glioblastoma is a part of a (pathologically altered) brain.

An alanine molecule in the lumen of an intestine: The molecule does not originate in the intestine, nor does its function coincide with the function of the intestine. Therefore it is contained in the intestine but is not a part thereof.

An alanine molecule in a bone: The alanine molecule is functionally essential for the collagen fiber it is included in. It is therefore part of this collagen fiber. The collagen fiber is synthesized within the bone, hence it has its origin therein and thus the parthood relation can be assumed between the fiber and the bone also. Due to the transitivity of the parthood relation, the alanine molecule is also part of the bone.

DISCUSSION

We will limit our discussion to two aspects, *viz.* the analysis of borderline cases, and the relevance of the above to ontology engineering.

The algorithm given in Fig. 2 suggests that we have solved the problem of deciding between parthood and containment once and for all. This, of course, is not the case. However, it is not the fault of our rules, but rather a consequence of certain underlying ontological assumptions which are not in every case satisfied in a clear-cut way. An example is the exact sortal delimitation between artifacts and biological matter, given the existence of engineered tissue or genetically modified cells. To take the tumor example, one could argue as follows: a single malignant cell may not yet be properly considered to be a metastasis, since one might hold that the metastasis comes into being only at the moment of the first cleavage of the tumor cell, something that may occur already at the final location. In this case, it becomes questionable whether the brain metastasis is **contained-in** or is **part-of** the brain. This is, however, a problem of identity (is the metastasis the same entity as its originating cell or is it a different one?) rather than of mereology. Another problem of identity arises when considering defined volumes of matter, e.g., the blood in my heart, or the air in your lung. One may regard the volume of air in your lung a continuant which preserves its identity across time (despite the rapid exchange of its molecules). In this case the function of the air would be as that of preventing the lung from collapsing and thus assuring gas exchange.

So far, our view has been limited to individual (token) objects [15]. Ontologies, however, generally deal with types or classes of individuals. Whenever we want to make assertions about classes, such as *fingernail part-of finger*, *brain metastasis located-in brain*, then we must take care that for each and every indi-

vidual in the former class there exists some individual in the latter class related by a corresponding instance level relations. The confusion between those relations which hold between individual objects (such as **part-of**) and their cognate relations which hold between classes of objects (*part-of*) has a long-standing tradition and has only recently been clearly addressed in the (bio)medical informatics community [13, 16]. We believe that it has still not been addressed in other branches of informatics. The algorithm depicted in Fig. 2 can therefore not be applied at the level of classes of objects. Let us take the albumin example. For each case the algorithm permits a clear statement of whether **located-in** specializes to **part-of** or **contained-in**. If we take, instead, the *class* of all alanine molecules which are **located-in** human beings and the class of all human beings, then we cannot use our criteria precisely because there are some alanine molecules which are merely contained (e.g., those in the gatrointestinal tract) and others which are parts (e.g. those within bones).

CONCLUSION

As a contribution to future bio-ontologies we relativized the role of the relation **part-of** and juxtaposed it to the equally important **contained-in** relation (see also [2]). Both relations are specializations of the relation **located-in** which relates all objects whose spatial regions exhibit topological inclusion. We proposed an algorithmic approach for specializing location to either parthood or containment using four criteria: genetic relation, type constraints, origination, and functionality. We conclude that the algorithm is useful to support the assignment of relations in describing biological organisms. An assessment of the cognitive adequacy of this approach based on domain experts is in preparation. The limitations of our approach are mostly due to vague boundaries between kinds, controversial conceptualizations of the lifes of biological objects of different types and an imprecise understanding of biological functionality, for which a well-founded ontological account is still required [9].

References

[1] SNOMED *Clinical Terms*. Northfield, IL: College of American Pathologists, 2004.

[2] Thomas Bittner, Maureen Donnelly, and Barry Smith. Individuals, universals, collections: On the foundational relations of ontology. In *FOIS 2004 – Formal Ontology in Information Systems. Proceedings of the 3rd International Conference*, pages 37–48. 2004.

[3] Anthony G. Cohn. Formalising bio-spatial knowledge. In *FOIS 2001 – Formal Ontology in Information Systems. Collected Papers from the 2nd International FOIS Conference*, pages 198–209, 2001.

[4] M. Donnelly. Relative places. In *FOIS 2004 – Formal Ontology in Information Systems. Proceedings of the 3rd International Conference*, pages 249–260. 2004.

[5] M. Donnelly. A formal theory for spatial representation and reasoning in biomedical ontologies. *Artificial Intelligence in Medicine*, 2005. (accepted for publication).

[6] Maureen Donnelly. A formal theory of reasoning about parthood, connection, and location. *Artificial Intelligence*, 160:145–172, 2004.

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

[8] P. Grenon and B. Smith. SNAP and SPAN. towards dynamic spatial ontology. *Spatial Cognition and Computation*, 4:69–103, 2004.

[9] I. Johansson, B. Smith, K. Munn, N. Tsikolia, K. Elsner, D. Ernst, and D. Siebert. Functional anatomy: A taxonomic proposal. *Acta Biotheoretica*, 2005.

[10] OBO. *Open Biological Ontologies (OBO)*. `[http://obo.sourceforge.net. ]`, 2005. Last accessed June 26th, 2005.

[11] Alan L. Rector, Aldo Gangemi, Elena Galeazzi, Andrzej J. Glowinski, and Angelo Rossi Mori. The GALEN model schemata for anatomy: Towards a reusable application-independent model of medical concepts. In *MIE'94 – Medical Informatics Europe 1994*, pages 229–233, 1994.

[12] Cornelius Rosse and José Leonardo V. Mejino. A reference ontology for bioinformatics: the Foundational Model of Anatomy. *Journal of Biomedical Informatics*, 36:478–500, 2003.

[13] S. Schulz. Bidirectional mereological reasoning in anatomical knowledge bases. In *AMIA 2001 – Proceedings of the Annual Symposium of the American Medical Informatics Association*, pages 607–611, 2001.

[14] Stefan Schulz and Udo Hahn. Representing natural kinds by spatial inclusion and containment. In *ECAI 2004 – Proceedings of the 16th European Conference on Artificial Intelligence*, pages 403–407, 2004.

[15] B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A. Rector, and C. Rosse. Relations in biomedical ontologies. *Genome Biology*, 2005.

[16] B. Smith and C. Rosse. The role of foundational relations in the alignment of biomedical ontologies. In *MEDINFO 2004 – Proceedings of the 11th World Congress on Medical Informatics. Vol. 1*, pages 444–448, 2004.