A Knowledge Representation View on Biomedical Structure and Function

Stefan Schulz^a Udo Hahn b

aFreiburg University Hospital, Department of Medical Informatics (http: //www. imbi.uni-freiburg.de/medinf) b_{Freiburg} University, **CF** Text Knowledge Engineering Lab (http://www.coling.uni-freiburg.de)

Abstract

In biomedical ontologies, structural and functional considerations are of outstanding importance, and concepts which belong to these two categories are highly interdependent. At the representational level both axes must be clearly kept separate in order to support disciplined ontology engineering. Furthermore, the biaxial organization of physical structure (both by a taxonomic and partonomic order) entails intricate patterns of inference. We here propose a layered encoding of taxonomic, partonomic and functional aspects of biomedical concepts using description logics.

INTRODUCTION

Many conceptualization efforts in the biomedical domain coincide in an upper-level distinction between tangible objects of the physical world, which are characterized by a spatial extension, on the one hand, and defined (changes of) states of the physical world, on the other hand.¹ Although these two subdomains are tightly related, their reasoning patterns are by no means straightforward. For example, the secretion of insulin is usually considered as a function of the pancreas, because the pancreatic beta cells are part of it. In the same way, muscular movement would be classified as a function of the muscle, since it is a function of the actin-myosin complex which is a component of muscle cells, the latter being part of the muscle. Such inferences $-$ function propagates from physical parts to w holes $-$ seem plausible. However, there are obvious counterexamples: Mitosis is a cell function, but certainly not a pancreas or liver function although these organs have cells as parts.

In this paper, we will discuss and classify typical reasoning patterns from which we derive our requirements for adequate domain modeling. Our focus will be on Biological Function, as a specialization of Event, and its interaction with Biological Structure. Both are of paramount importance for the fields of anatomy, physiology, pathology as well as for cell and molecular biology. We then propose ^a suitable ontology engineering approach, based on our previous work on biomedical knowledge representation [13, 11, 12].

STRUCTURE AND FUNCTION

In knowledge representation, concepts are characterized by their attributes, also called slots in frame-based representations or *roles* in description logic systems (e.g., has-location, part-of,...), while role filler restrictions constrain their sortal ranges. Liver, e.g., is a reasonable filler of the role has-location of Hepatitis. With C being a concept, r one of its roles filled by the role filler F , we make the following distinctions:

- Forbidden roles. F is disallowed to be a filler of the role r at the concept C. As an example, for $C =$ Dermatitis and $r = has-location$, a filler such as $F =$ Liver must be rejected.
- Optional roles. Those roles of a concept C which can be filled but need not. As an example, the role has-part.Nucleus is an optional one when we define the concept Cell, since cells with and without nuclei exist (e.g. red blood cells).
- \bullet Mandatory roles. Those roles of a concept C which are required to be filled as, e.g., in the case of haslocation. Liver in the above definition of Hepatitis. This allows to infer that for each concrete hepatitis there must be a concrete liver it affects.

In the following, we will focus on the distinction between optional and mandatory roles in the definition of Structure and Function concepts. We introduce the relations part-of and its converse has-part (abbreviated as p and i (i for "includes as part")), which hold between Structure and Structure. Accordingly, we introduce has-function and function-of (abbreviated as hf and fo) which relate Structure and Function. Given these distinctions, when we analyze our domain the following observations can be made:

• In Structure concept definitions, roles filled by Function concepts mostly have the 'optional' status, since most functions in living systems can cease without any implications for the identity of the objects concerned. A Liver Cell, e.g., has normally

¹The UMLS [14] distinguishes between Entities and Events, the GALEN ontology [6] between GeneralizedStructure and GeneralizedProcess, the Gene Ontology [4] between Cellular Component and Biological Process, and in the TAMBIS Ontology [2] this distinction is reflected by GeneralizedStructurelGeneralizedSubstance vs. GeneralizedProcesslGeneralizedFunction.

the function of Lipid Biosynthesis, but this is not a necessary condition, since even dead Liver Cells are still liver cells.²

• In Function concept definitions most roles are mandatory. The instantiation of Insulin Biosynthesis, e.g., requires the role function-of to be filled by one or more Structure concepts, e.g., Beta Cells.

As an intermediate representation we now specialize the relations hf and fo to distinguish between mandatory and optional functions such as given in Table 1:

Table 1: Specialized relations between Function and Structure. The upper two relations are subrelations of functionof (fo) , the lower ones are subrelations of has-function (hf) . Domain characterizes the concept to be defined, Range the conceptual constraint on the role filler.

The distinction between optional and mandatory roles is equally important when we limit our view to Structure concepts and the mereological relations part-of and has-part. Such partonomies play a prominent role in biomedical ontologies (e.g., the Gene Ontology [4] and the Digital Anatomist Foundational Model [9]), and various researchers have already investigated their semantic foundations [1, 13, 8]. We subscribe to ^a nonconstrained understanding of parthood in which the relations p and ⁱ are considered as reflexive, antisymmetric and transitive [3].

In a strict sense, mereological relations can only hold between individuals, not between concepts. It is obvious that for each pair of individuals (I_1,I_2) , related by p , the inverse relation i (has-part) always turns out to be true between (I_2,I_1) . The transfer of this symmetry to the level of concept definitions (as in the UMLS metathesaurus where there is exactly one part-of link for each has-part link) can only be justified under the assumption of all roles being optional. This constraint is, however, too weak for many purposes. So we distinguish between optional and mandatory parts and wholes. Hence, symmetry between part-of and haspart is no longer justified, taking common conceptualizations of the domain into account:

(i) Every Cell has Protein as part, but not every Protein is part of a Cell.

(ii) Every Cell Nucleus is part of a Cell, but not every Cell has a Cell Nucleus.

Just as with the relations fo and hf , we resort to specialized partitive relation predicates as introduced in Table 2 and may assert:

(i') Cell hmp Protein, but Protein how Cell,

(ii') Cell hop Cell Nucleus, but Cell Nucleus hmw Cell.

Relation Name	Abbr.	Domain	Range
has-mandatory-part	hm p	Structure	Structure
has-optional-part	hop	Structure	Structure
has-mandatory-whole	hmw	Structure	Structure
has-optional-whole	how	Structure	Structure

Table 2: Relations between Structure and Structure. hmp and hop are subrelations of has-part (i), hmw and how of part-of (p) .

We now analyze interactions between Structure and Function for plausible inferences. All possible combinations are listed in Table 3.

binations are itsied in Table 5.					
٥	$\frac{WhmpP}{N}$ P hmw W	$\frac{WhmpP}{PhowW}$	WhopP\ PhmwW	WhopP \land PhowW	
F mpb W					
$Fmbb$ P	F mpb W	F opb W	F mpb W	F opb W	
F opb W					
Fopb P	F opb W	F opb W	F opb W	Fopb W	
W hm f -F					
P hm	WhmfF	$\mathbf W$ hof F	W ho . F	ΓF W ho	
W ho f					

rative 5: composition alore or plaasible rolation ename, in volving relations between *Structure* Concepts ($W =$ wholes, Table 3: Composition table of plausible relation chains, in- $P =$ parts) and *Function* Concepts (F)

We now illustrate the reasoning pattems which derive from Table 3:

• Insulin synthesis is mandatorily performed by (mpb) the pancreatic Beta Cells. Beta Cells have Islets of Langerhans as ^a mandatory whole (hmw). We infer that *Insulin Synthesis* is also mandatorily performed by (mpb) Islets of Langerhans:

• The structure concept *Beta Cells* has the optional function (hof) Insulin Synthesis (it is not mandatory as it may be inactive in certain states). Islets of Langerhans have Beta Cells as an optional part (in case of certain types of diabetes these islets are devoid of beta cells). We infer that Islets of Langerhans, too, have Insulin Synthesis as an optional function:

• The inferences given in Table 3 can however not be taken for granted such as this atypical example illustrates:

²Mandatory roles filled by Function concepts occur occasionally when a new concept is defined and the mandatory function role is taken as a definitional criterion. As an example, the concept ReplicatingDNA is defined as a DNA as long as the function Replication is performed.

This is a questionable inference, at least. There is also a subtle difference between a function performed by and a function performed at. For the latter we observe fewer exceptions.

As a result, we stipulate that a model capable of dealing with the standard inference patterns given in Table 3 should also be flexible enough to comply with exceptions to these patterns. Such an approach should overcome the limits of the Grail concept representation language [5], which although it provides means for dealing with the role propagation across part-whole hierarchies, has turned out to be too rigid to account for reasoning anomalies[7] such as discussed above.

A DESCRIPTION LOGICS APPROACH

The problems stated in the previous section are addressed by the PI-model (from "part" / "include") which extends our previous work on the *SEP* [13] and Extended SEP [11] models. Our approach, which has already proved valid for the integration of partonomic and taxonomic reasoning, is based upon a set-theoretic semantics underlying the description logics language $ALC[10]³$.

The PI-model is centered around the relations p and ⁱ with the semantics introduced above. We emulate transitive mereological reasoning by taxonomic reasoning by introducing additional concepts which are introduced as 'reificators' of the relations p and i . For each structure concept $(S$ in this example) we introduce, for both the p and the i relation, two artificial concepts as common subsumers for all those concepts which must have, by definition, the role p filled by S , or the role ⁱ filled by S, respectively.

$$
S_P \sqsubseteq \exists p.S \tag{1}
$$

$$
S_I \sqsubseteq \exists i.S \tag{2}
$$

So, S_P subsumes all concepts that have S as a mandatory whole, and S_I subsumes the ones having S as a mandatory part. As an example, the concept Handp subsumes both Finger and Thumb.

Cascading subsumption of concepts (in parallel with the corresponding 'part' node) by 'part' nodes of their mandatory wholes emulates transitivity of the p relation (see Fig. 1). The same applies to the emulation

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

of i transitivity: In this case, concepts – together with their corresponding 'include' nodes – are subsumed by 'include' nodes of their mandatory parts.

Fig. ¹ depicts a scenario in which mandatory parts correspond to mandatory wholes. Each instance of Cell Membrane (M) has its role p filled by an instance of $Cell (C)$, because M is subsumed by C_P . Each Cell (C) has the role ⁱ filled by an instance of Cell Membrane, because C is subsumed by M_l (the same applies to Cell Membrane (M) and Cell Membrane Protein (P)). As P is subsumed by M_P and M_P by C_P , we infer that not only M , but also C is a mandatory whole for P . Accordingly, C is subsumed not only by M_I , but also by P_I , therefore both M and P are mandatory parts for C (in other words, each instance of Cell must have an instance of Cell Membrane and an instance of Cell Membrane Protein related by the relation has-part). In the lower plane of the figure an example for a subsumee of Cell Membrane (together with i and p node) is given in order to illustrate that the proposed encoding scheme is clearly distinct from a taxonomic order in an "orthogonal" way.

For a formal reconstruction of these considerations, let us assume C , M and P to denote the concepts to be modeled, and C_P , M_P , and P_P to denote "part" nodes, related to C, M, and P via the role p. Accordingly C_l , M_I , and P_I denote "include" nodes related to C, M, and P via the role i :

$$
M \sqsubseteq C_P \sqsubseteq Structure \qquad (3)
$$
\n
$$
M_P \sqsubseteq C_P \sqsubseteq Structure \qquad (4)
$$
\n
$$
P \sqsubseteq M_P \sqsubseteq C_P \sqsubseteq Structure \qquad (5)
$$
\n
$$
P_P \sqsubseteq M_P \sqsubseteq C_P \sqsubseteq Structure \qquad (6)
$$
\n
$$
M \sqsubseteq P_I \sqsubseteq Structure \qquad (7)
$$
\n
$$
M_I \sqsubseteq P_I \sqsubseteq Structure \qquad (8)
$$
\n
$$
C \sqsubseteq M_I \sqsubseteq P_I \sqsubseteq Structure \qquad (9)
$$
\n
$$
C_I \sqsubseteq M_I \sqsubseteq P_I \sqsubseteq Structure \qquad (10)
$$

 $3ALC$ allows for the construction of hierarchies of concepts and relations, where \sqsubseteq denotes subsumption and \doteq definitional equivalence. Existential (\exists) and universal (\forall) quantification, negation (\neg) , conjunction (\neg) and disjunction (U) are supported. Role fillers are linked to the relation name by a dot, e.g., $\exists r.C$.

Figure 2: PI-Architecture: Asymmetric situation. Emulation of transitivity of both p and its converse i . The is-a links with label (1) emulate the part-of (p) hierarchies, the is-a links with label (2) emulate the has-part hierarchies (i). The dotted arrows have to be removed when the domain allows for pathological modifications concept Pancreas (which may be devoid of Beta Cells in certain types of diabetes).

$$
C_P \sqsubseteq Structure \sqcap \exists p.C \qquad (11)
$$

$$
M_P \sqsubseteq Structure \sqcap \exists p.M \qquad (12)
$$

$$
P_P \sqsubseteq Structure \sqcap \exists p.P
$$
 (13)

$$
C \sqsubseteq Structure \sqcap \exists i.C
$$
 (14)

$$
C_I \sqsubseteq Structure \sqcap \exists i.C
$$
 (14)

$$
M \sqsubset Structure \sqcap \exists i.M
$$
 (15)

$$
M_1 \subseteq \text{structure} \quad \sqcup \quad \sqcup \ldots \sqcup \sqcup \sqcup
$$

$$
P_I \sqsubseteq Structure \quad \sqcap \quad \exists i.P \qquad \qquad (16)
$$

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

The flexibility of this approach is shown in Fig. 2. Here the (dotted) is-a links between I and B_I , and I_I and B_I , respectively, constrain the concept Islets of Langerhans in terms of disallowing pathologically modified instances (which have no Beta Cells) requiring Beta Cells as mandatory part. Another pair of links between Glucose Receptor Protein and Beta Cell is omitted intentionally, since the latter is not mandatory (glucose receptor proteins occur in many kinds of cells).

In the following we will discuss whether the proposed part-whole model is capable of accounting for the most common inferences given in Table 3. Note that the model itself does not use any of the specialized relations from Table 3.

For Structure concept definitions we raise the following claims:

 \bullet A *Structure* concept W has a mandatory part P. P has the optional function F (e.g., $W = Lung$, $P =$ Alveolus, $F = Gas$ Exchange). The rule to be applied here, is P hof F \circ W hmp P \rightarrow W hof F. Since our language specification does not support the notion of optional roles proper (it allows only existential and universal quantification of roles), we have to introduce specialized roles, here for the function F the role $hf-F \sqsubseteq hf$ which has its range restricted to F. This role is attached not only to P , but also to the "include" node P_I . Thus the role is inherited to all subsumees of P_I , i.e. to P and all other concepts which have P as a mandatory part.

$$
P_I \quad \sqsubseteq \quad \exists i.P \tag{17}
$$

$$
W_I \quad \sqsubseteq \quad \exists i. W \tag{18}
$$

$$
W \subseteq P_1 \tag{19}
$$

$$
W_I \quad \subseteq \quad P_I \tag{20}
$$

$$
P \subseteq \forall hf\text{-}F.F \qquad (21)
$$

$$
P_I \quad \sqsubseteq \quad \forall h f \text{-} F \text{-} F \tag{22}
$$

Wherever this behavior does not comply with shared inferential intuitions, expression (22) can be omitted. Otherwise, the function role is propagated through the 'includes' hierarchy, so we obtain the subsumptions:

$$
W \subseteq \forall hf\text{-}F.F \qquad (23)
$$

and so on for W_I and its subsumees.

• The following inference is not supported by our encoding: P hof F \circ W hop P \rightarrow W hof F (e.g., W $=$ Pancreas, $P = Beta$ Cell, $F = Insulin Biosynthe$ sis), because of the already mentioned problem to express the notion of optional roles proper by our knowledge representation language.

For Function concept definitions we stipulate the following:

 \bullet A Function concept E is defined, being mandatorily functional for the Structure concept P . P has a mandatory whole W . We want to enforce inferences such as

E mpb P o P hmw $W \rightarrow E$ mpb W. In case another Function concept F exists which is mandatorily functional for $W \sqcup W_P$, then F subsumes E, since P is subsumed by W_P :

$$
P_P \subseteq \exists p.P \qquad (24)
$$

$$
W_P \quad \sqsubseteq \quad \exists p. W \tag{25}
$$

$$
P \sqsubseteq W_P \tag{26}
$$

$$
P_P \quad \sqsubseteq \quad W_P \tag{27}
$$

Next, the functions are defined:

$$
E \doteq Function \sqcap \exists fo.P \qquad (28)
$$

$$
F \doteq Function \sqcap \exists fo.(W \sqcup W_P) \qquad (29)
$$

So we infer:

$$
E \subseteq F \qquad (30)
$$

because P is subsumed by W_P (e.g., $W =$ Kidney, $P = Glomerulum, E = Glomerular Filtration, F =$

Renal Function). This behavior can be obviated by restricting the range to W only (excluding W_P):

$$
F \doteq Function \sqcap \exists fo.W \qquad (31)
$$

 \bullet A Function concept F is defined, being mandatorily performed by the *Structure* concept P . P has a mandatory whole W . We want to allow inferences such as F mpb P \circ W hmp P \rightarrow F mpb W (e.g., again $W = Lung$, $P = Alveolus$, $F = Gas Exchange$.

$$
P_I \quad \sqsubseteq \quad \exists i.P \tag{32}
$$

$$
W_I \quad \sqsubseteq \quad \exists i. W \tag{33}
$$

$$
W \subseteq P_I \tag{34}
$$

$$
W_I \quad \sqsubseteq \quad P_I \tag{35}
$$

In its definition F has the role fo filled by the disjunction of P and its I-node P_1 :

$$
D \doteq Function \sqcap \exists fo(P \sqcup P_1) \qquad (36)
$$

Since P_I subsumes W , the latter is contained in the range of the fo role in the definition of F . Here too, an "exception handling" is possible by excluding P_I from the range of the function fo in expression (35) (e.g., with $W = Organism$, $P = Eukaryotic Cell$, $F =$ Mitosis).

 \bullet When defining functions related to *optional* wholes, the expected inferences would require additional nodes, as discussed above.

Summing up, the PI encoding provides not only an elegant way to simulate part-of and has-part transitivity by taxonomic subsumption, but also more flexibility in defining both Function concepts and Function roles in Structure concept definitions. By adding two "proto-nodes" for every Structure concept, several control mechanisms for the propagation of roles within mereological hierarchies exist which allow for a more precise adaptation of the ontology to common conceptualizations in the biomedical domain.

CONCLUSIONS

The notions of biomedical structure and function not only require to sort out appropriate general ontological categories but also to take care 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 as reificators for the relations part-of and has-part. These artifacts not only enable the propagation ofroles across partonomic hierarchies, but also the accommodation of exceptions from standard reasoning patterns. There are other approaches to the handling of such exceptions, e.g., by specific subrelations of mereological relations [8], but in any case the gain in flexibility has to be traded against the proliferation of (artificial) concepts and relation hierarchies. Due to the parsimony of the underlying language scalability is an asset, as we have demonstrated in a previous study in which a very large terminological knowledge base was constructed [12] using a similar approach.

References

- [1] A. Artale, E. Franconi, N. Guarino, and L. Pazzi. Part-whole relations in object-centered systems: An overview. Data & Knowledge Engineering, 20(3):347- 383, 1996.
- [2] P.G. Baker, C. A. Goble, S. Bechhofer, N.W. Paton, R. Stevens, and A. Brass. An ontology for bioinformatics applications. Bioinformatics, $15(6):510-510$, 1999.
- [3] R. Casati and A. Varzi. Parts and Places. The Structures of Spatial Representation. MIT Press, 1999.
- [4] Gene Ontology Consortium. Creating the Gene Ontology resource: Design and implementation. Genome Research, 11(8):1425-1433, 2001.
- [5] A. Rector, S. Bechhofer, C. Goble, I. Horrocks, A. Nowlan, and D. Solomon. The GRAIL concept modelling language for medical terminology. Artificial Intelligence in Medicine, 9:139-171, 1997.
- [6] A.L. Rector, J.E. Rogers, and P. Pole. The GALEN high level ontology. In Proceedings of Medical Informatics in Europe - MIE 96, pages $174-178$, 1996.
- [7] Alan L. Rector. Analysis of propagation along transitive roles: Formalisation of the galen expe rience
with medical ontologies. In *DL02 - 2002 International* Workshop on Description Logics, 2002. Published via http://CEUR-WS.org/Vol -531.
- [8] J. Rogers and A. Rector. GALEN's model of parts and wholes: Experience and comparisons. In Proceedings of the AMIA 2000, pages 714-718, 2000.
- [9] C. Rosse, J.L. V. Mejino, B. R. Modayur, R. Jakobovits, K.P. Hinshaw, and J.F. Brinkley. Motivation and organizational principles for anatomical knowledge representation: The Digital Anatomist symbolic knowledge base. JAMIA, 5(1):17-40, 1998.
- [10] M. Schmidt-SchauB and G. Smolka. Attributive con-cept descriptions with complements. Artificial Intelligence, 48(1):1-26, 1991.
- [11] S. Schulz. Bidirectional mereological reasoning in anatomical knowledge bases. In AMIA 2001, pages 607-611,2001.
- [12] S. Schulz and U. Hahn. Medical knowledge reengineering - converting major portions of the UMLS into a terminological knowledge base. International Journal of Medical Informatics, 64:207-221, 2001.
- [13] S. Schulz, M. Romacker, and U. Hahn. Part-whole reasoning in medical ontologies revisited: Introducing SEP triplets into classification-based description logics. In AMIA 1998, pages 830-834, 1998.
- [14] UMLS. Unified Medical Language System. Bethesda, MD: National Library of Medicine, 2002.