User Centred Networked Health Care
A. Moen et al. (Eds.)
IOS Press, 2011

doi:10.3233/978-1-60750-806-9-739

Architectural Approach for Providing Relations in Biomedical Terminologies and Ontologies

Mathias BROCHHAUSEN[a,b,1], Bernd BLOBEL[c]
[a] IFOMIS, Universität des Saarlandes, Saarbrücken, Germany
[b] Department of Philosophy, University at Buffalo, Buffalo, USA
[c] eHealth Competence Center, University Hospital Regensburg, Regensburg, Germany

Abstract. The representation of multiple relations is one of the main criteria of ontologies. In formalizing both ontologies and terminologies in biomedicine relations are used to code axioms for the classes of the ontology. However, a huge number of relations represented in medical ontologies and terminologies are derived from language and formal definition is omitted. We present a strategy based on an architectural approach to facility formal analysis of relations for use in ontology systems in biomedicine and in general.

Keywords. Biomedical Ontologies, Medical Terminologies, System Theory, Biomedical Relations, Architecture

1. Introduction

Relations are central features in ontologies and all terminologies that aim to provide more than just a single hierarchy or completely flat representation of reality. In the past the representation of relations has often been done language-based or inductive from observed instance-instance relations. In order to fully understand representation of relations in ontologies it is important to be aware of the difference between instances (particulars) and types. Basically, there are three kinds of relations: instance-instance relations, type-type relations and instance-type relations.

[1] is a key contribution to unify use of relations in the entire biomedical domain. defining type-type relations based on undefined empirical relations between instances. Even though this was a first step towards a more controlled usage of relations in the biomedical arena, the methodology should be improved to gather more formal differences between the relations. The formal definitions provided by the Open Biological and Biomedical Ontologies (OBO) Relation Ontology use expressions that go beyond binary relations and thus, cannot be coded in Web Ontology Language (OWL), in which relations are exclusively binary.

In this paper we aim not at ontologies re-using the OBO Relation Ontology, but we discuss issues in widely used medical terminologies that arise from a lack of theoretical well-funded strategies regarding the representation of granularity, thereby focusing on

[1] Corresponding author. Dr. Mathias Brochhausen, IFOMIS, Saarland University, P.O. Box 15 11 50, 66041 Saarbrücken, Germany; Phone: +49 681 30264770; Email: mathias.brochhausen@ifomis.uni-saarland.de

the architectural aspect of any concrete or abstract system and its representation [5]. Over the last decade a number of problems have been detected and discussed in SNOMED CT, National Cancer Institute Thesaurus (NCIT) and other terminologies and ontologies in the eHealth arena [2, 3]. We hold that the problems arise from a deficiency of the approaches taken by these controlled vocabularies and ontologies: a lack of well-founded strategies to connect different levels of granularity and medical disciplines. We suggest to use the systems approach in order to accomplish a complete analysis of spheres involved [4]. The Generic Component Model (GCM) provides an architecture framework for the complete process of representation and systematization of a given domain, including the domain's decomposition/composition [5]. We will use GCM to clarify the relations between the entities in the domain and the informational representation of them.

2. Materials and Methods

2.1. Trans-Granular Relations in Biomedical Ontologies

The problem is that in the terminologies mentioned above "concepts" are related with each other based on observations from medical practice, more or less regardless of their position within a system and its subsystems. Sometimes non-matching entities are represented as being linked by not matching relations, as in the example from NCI Thesaurus:

(1) (Acinic Cell Breast Carcinoma) Disease_May_Have_Finding (Pain) [6].

The NCIT does not give a textual definition of "Acicnic Cell Breast Carcinoma", but it seems obvious that this term ought to refer to the physical neoplasm that is part of the patient's body. "Pain" defined by the NCIT as: "The sensation of discomfort, distress, or agony, resulting from the stimulation of specialized nerve endings" [6]. Thus, pain is the sensation that is experienced by the organism and is not the same as the stimuli of nerve endings. Correctly, pain is a process or a state of the entire organism. Notably, a physical entity such as a neoplasm is a disease according to the NCIT. However, we stress that a more powerful and coherent interpretation of disease is given by Scheuermann et al. [7]. According to this paper, disease is a disposition to undergo pathological processes. The first problem is the idea to link an organismal structure to a phenomenon of the organism as a whole and its finding within a diagnostic process. This is done using a relation that needs a disease as domain.

In order to give a formally more adequate representation we may view the organism, its parts and the processes that take place within or adjacent to the organism as a system and its subsystems. We propose that different levels of granularity (e.g. molecular level, cell level, tissue level, organ level, organismal level) should be viewed as subsystems of one big system. Thus, we are able to distinguish relations within one subsystem from relations that bridge between different subsystems.

2.2. The GCM

The GCM is an architecture framework that enables the representation of any real or virtual system including both the system architecture from its business perspective and the system's development process for the ICT solution supporting or enabling that business. The approach allows for modelling systems by reducing their complexity and separating the phases of their design, specification, implementation and deployment by representing and interrelating different views, namely Enterprise View, Information View, Computational View, Engineering View (Fig. 1) [5]. For our purpose we can focus on:

- **Enterprise View** - captures the real world business process, in our case all relevant biological processes (physiological and pathological), biomedical process and medical processes.
- **Information View** - captures the informational expression of the Enterprise View, in our case the representation in a terminology or in an ontology.

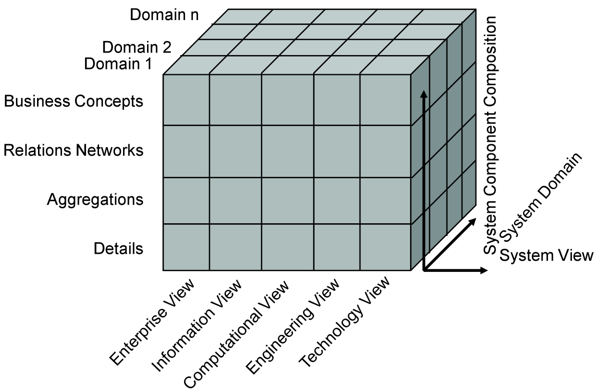

Figure 1: The General Component Model (GCM)

3. Results

Our aim is to give a system theoretic consistent, architecture centric reformulation of (1). Our starting point is that the different levels of granularity, which we have to take into account regarding biological structures and processes, can be viewed as a sequence of interrelated systems and sub-systems. Figure 2 illustrates the fact that within an organism organismal components are system components, however we can view each of these components as systems themselves. Thus the system of interest can be defined at different level of granularity from body through organ, tissue, and cell down to the level of the molecular structure of cells, always considering the system and different granularity levels of subsystems.

Within each level of granularity we have a multitude of relations between the components of the system; for instance within an organism we have relations between its organismal components (Fig. 2). Our basic approach is to keep two types of relation distinct: (A) relations at the same granularity level and (B) relations between different granularity levels. We expect that the number of Type B relations can restricted to quite some extend due to the limited scope of selected interesting processes.

We have to reformulate (1) as follows:

(2) (Acinic Cell Breast Carcinoma) causes (Stimulated Nocireceptors)
(3) (Stimulated Nocireceptors) lead_to (Pain Perception in Organism 1)
(4) (Pain Perception in Organism 1) is_reported_by (Organism 2)

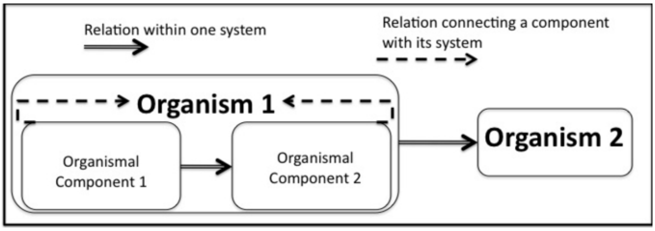

Figure 2: Systems and components with the two different types of relations

For our approach, it is important to note that "lead_to" is quite different from "causes" as "lead_to" is relating a system component with its system, thus bridging between two levels of granularity. From a formal point of view, this is important since besides the linguistic differentiations between the two verbs used to name the relation this provides us with a formal criterion for difference of the two relations. Note that one could well use "lead_to" in both cases. However, the aim of providing an ontological representa-tion is to provide language-neutral and machine understandable semantic criteria. We hold that a systemic analysis of a given domain and its relations yields to criteria which can be formalized to differentiate relations which are scarcely differentiated by natural language use, even in a more technical or scientific setting like medicine.

However, it is important to note that from the GCM perspective there is something even more important happening. The physiological processes of the patient are all captured on the Enterprise View, whereas the abduction from information about the physical state of the patient to information about the symptoms that we can expect is part of the Information View.

So, within the GMC framework, the following reformulation of (1) applies:

(2') (Enterprise View) Patient has_Part Acinic Cell Breast Carcinoma

(3') (Information View) Report of Acinic Cell Breast Carcinoma is_positively_correlated with Report of Pain

(4') (Enterprise View) Expectation: Patient experiences Pain

Notably, (1) is an example of a sentence expressing probabilistic assertion rather than an assertion about a type being true for all individuals of that type. Rector [8] points out that knowledge of this type for formal reasons cannot be displayed in ontologies. These kind of assertions need to be part of a "background knowledge resources". Based on this distinction, Schulz et al. put forward an argument that knowledge representation is not a task of formal ontologies [9]. We hold that this is an overstatement. Yet, we agree that ontologies cannot be the only resource in a knowledge management system, since it represents only knowledge about entire types and their properties at any time. We have demonstrated that the GCM helps to raise awareness of the different resources within a knowledge management system.

4. Conclusions and Discussion

From the above we learned two things: a system theoretic, architecture centric analysis of relations in reality offers interesting opportunities to find formalizable differences between relations and will help to fix the semantics of relations for machine-machine communication. The problem with representing relations in applied ontology is that the entities represented in most clinical terminologies and ontologies are located in the mesocosm [10]. This fact puts the developers in danger of mixing different types of granularities as they appear in the transition from microcosm (molecular processes) to medical reality (therapeutic processes). Exact rules of relating the entities in the domain representation are missing, but using the GCM and its system-theoretical background can be a first step towards systematizing the representation of relations.

GCM will help representing knowledge management systems as a whole, including the different components, e.g. ontology and background knowledge resources, and the different operations carried out by them.

We started our analysis of relations by pointing out the difference between instance-instance relations and type-type relations. The system-theoretical approach raises the question how we plan to keep this distinction up in a framework where we view entities as systems to better grasp the properties they bear for ontological representation. In viewing real world phenomena from the system-theoretical, architecture centric perspective we can do both, viewing cells in general or viewing one particular cell. The latter will not lead to any kind of ontological knowledge regarding cells in general, but nevertheless can be important in a medical knowledge management system (e.g. an HIS). Only the analysis of cells in general can provide us with the properties that need to be represented in an ontology.

We would like to add that the newest version of the Web Ontology Language, OWL 2, provides the opportunity to create property chains [11]. This methodology would support the definition of (1) by supplying the chain (2) - (4) as its definition. Nevertheless, in order to distinguish the relations used in the definitory chain, the system-theoretical criteria described above ought to be used.

References

[1] Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, et al. Relations in Biomedical Ontologies, *Genome Biology*, 6, 5 (2005), R46. PMC1175958

[2] Bodenreider O, Smith B, Kumar A, Burgun A. Investigating subsumption in SNOMED CT: An Exploration Into Large Description Logic-Based Biomedical Terminologies, *Artif Intell Med* 39/3 (2007), 183/95.

[3] Ceuster W, Smith B, Goldberg L. A Terminological and Ontological Analysis of the NCI Thesaurus, *Methods Inf Med* 44 (2005), 498-507.

[4] Lopez DM, Blobel B. A Development Framework for Semantically Interoperable Health Information Systems, *Int J Med Inf* 78, 2 (2009), 83-103.

[5] Blobel B. Architectural Approach to eHealth for Enabling Paradigm Changes in Health, *Methods Inf Med* 49/2 (2010), 123-34.

[6] http://ncit.nci.nih.gov. Last accessed April 28, 2011

[7] Scheuermann RH, Ceusters W, Smith B. Toward an Ontological Treatment of Disease and Diagnosis, Proceedings of the 2009 AMIA Summit on Translational Bioinformatics, 2009, 116-120.

[8] Rector A. Barriers, Approaches and Research Priorities for Integrating Biomedical Ontologies, 2008. Available from: www.semantichealth.org/DELIVERABLES/Semanti- cHEALTH_D6_1.pdf. Last accessed: April 28 2011.

[9] Schulz S, Stenzhorn H, Boeker M, Smith B. Strengths and limitations of formal ontologies in the biomedical domain, *Elect. J. Commun. Inf. Innov. Health*. Rio de Janeiro, v.3, n.1, 31-45, Mar., 2009.

[10] Smith B. Ontologie des Mesokosmos: Soziale Objekte und Umwelten, *Zeitschrift für philosophische Forschung* 52 (1998), 521-40.

[11] Goldbreich C, Wallace EK. OWL 2 Web Ontology Language New Features and Rationale. Oct 27 2009. http://www.w3.org/TR/owl2-new-features. Last accessed April 28, 2011