From: *Kooperative Versorgung, Vernetzte Forschung, Ubiquitäre Information* (Proceedings of GMDS Innsbruck, 26-30 September 2004), Niebüll: Videel OHG, 199–201.

Normalizing Medical Ontologies using Basic Formal Ontology

Thomas Bittner[1] and Barry Smith [1,2]

[1]*Institute for Formal Ontology and Medical Information Science, Saarbrücken, Germany*
[2]*Department of Philosophy, SUNY Buffalo, USA*

Introduction

Description Logics [1] are nowadays widely accepted as formalisms for implementing rigorous domain ontologies, and have been used in biomedicine in projects such as GONG [2] and SNOMED-CT [3]. A key feature of such ontologies is that the associated reasoning facilities allow us to discover inconsistencies and other problems in an automatic fashion. This is important since ontologies of complex domains such as medicine are large and complex and have been built by many people over long periods of time.

The DL reasoning facilities allow ontologies to be developed in modular fashion, where changes in one module that affect other modules are propogated through the system automatically by the reasoner in a way which helps to maintain consistency and stability in the ontology as a whole.

Modular ontologies and normalization

For the feature of modularity to be utilized, however, requires that domain ontologies be represented in a *normalized form* [4]. This means that: (i) distinct modules must be represented as disjoint classification trees (for example by separating an anatomy and a disease module); (ii) binary relations between classes in distinct modules must be established (for example *hasLocation*, *participatesIn*, *hasRole*, *isContainedIn*, *causes*, *hasQuality*, *playsRole*, etc.), with axioms stating for example that pneumonia is an inflammation which *hasLocation* lung. This allows the classifier to compute the subsumption hierarchy which results when the modules are combined. Often the resulting hierarchy is not a tree.

Normalization and top-level ontologies

To normalize a domain ontology, then, we need to find (i) an appropriate set of trees that form its skeleton and that represent ontologically significant categorial distinctions, and (ii) an appropriate set of binary relations. We will show below that it is the job of a top-level ontology to provide the basic categories within which the different tree structures reside, and also to provide a list of the binary relations together with axioms that specify their semantics.

There are currently multiple top-level ontologies under development, e.g., DOLCE [5] or Basic Formal Ontology (BFO) [6]. In the context of this paper we will consider the latter. Based on BFO we discuss (a) the basic categories that give raise to the different tree structures in normalized domain ontologies and (b) the formal relations, which establish the relationships between classes within and across category trees.

Top-level categories

The most basic categorial distinction between entities at the top-level relates to different modes of persistence through time. Two categories of persistent entities can be distinguished: endurants and perdurants. Endurants are wholly present (i.e., all their proper parts are present) at any time at which they exist. For example, you (an endurant) are wholly present in the moment you are reading this. No part of you is missing.

Perdurants, on the other hand, are extended in time in virtue of possessing different temporal parts at different times. In opposition to endurants they are only partially present at any time at which they exist – they evolve over time. For example, at this moment only a (tiny) part of your *life* (a perdurant) is present. Larger parts of your life – such as your childhood – are not present at this moment.

In the context of ontological normalization it is therefore critical to distinguish between perdurants and endurants at the very top of the classification hierarchy. In BFO perdurants and endurants form disjoint category trees.

Endurants are divided into two major categories: independent endurants such as organisms and organs, and dependent endurants such as body spaces, qualities, roles, states, functions, etc. BFO distinguishes the following kinds of independent endurants: substances, fiat parts of substances, aggregates of substances, boundaries of substances, and cavities.

Substances are maximally connected entities, i.e., they have connected bona fide boundaries, i.e., boundaries which correspond to discontinuities in the underlying reality. Neither your nose nor your arm are substances. Both are fiat parts of you. In biomedical ontologies fiat parts are usually referred to as subdivisions [7, 8]. Aggregates of substances are not substances either. Examples of aggregates are: your family, your digestive tract, etc. Examples of cavities are the lumen of the lung or bladder, and the thorax cavity.

Dependent endurants are entities that cannot exist without some other entity or entities upon which they depend. They include: qualities, functions (the function of the heart is to pump blood), states (Mary is in the state of being pregnant), roles (Roger has the role of being a medical doctor), etc. A disease is a dependent endurant as is a quality such as temperature.

Processes, the second top-category in BFO, are perduring entities. BFO distinguishes processes, their fiat parts, aggregates, and boundaries. Your life is a process, as is the course of a disease. Your youth is a fiat part of your life. The lives of your family members is an aggregate of processes.

Formal relations

A top-level ontology such as BFO provides also systematic means to identify the relations between the classes which are organized in the separate tree-structures. These include not only those listed above, all of which are recognized in BFO,but also structural relations such as *isDependentOn*, *isRealizedIn*, etc. The resultant methodology then complements the BFO-based methodology for integrating ontologies in systems such as LinkSuite [9]. Within the framework of a top-level ontology also the axioms, which specify the semantics of these relations, can be provided along the lines set out in [10].

Literature

[1] The Description Logic Handbook, edited by F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider, Cambridge University Press, 2002

[2] C.J. Wroe, R. Stevens, C.A. Goble, M. Ashburner, A Methodology to Migrate the Gene Ontology to a Description Logic Environment Using DAML+OIL Pacific Symposium on Biocomputing 8:624-635(2003).

[3] College of American Pathologists. Snomed Clinical Terms® Technical Reference Guide, July 2003 release.

[4] Rector, A. Modularisation of domain ontologies implemented in description logics and related formalisms including OWL, In Proceedings of the international conference on Knowledge capture, 2003, 121-128.

[5] Gangemi, A., Guarino, N., Masolo, C., Oltramari, A. and Schneider, L., Sweetening Ontologies with DOLCE, Proceedings of EKAW, 2002

[6] Grenon, Pierre, Barry Smith, Louis Goldberg: Biodynamic Ontology: Applying BFO in the Biomedical Domain. in: Pisanelli, Domenico M. (ed.): Ontologies in Medicine: Proceedings of the Workshop on Medical Ontologies, Rome, October 2003. IOS Press, Amsterdam.

[7] Mejino, J. L. V. and Agoncillo, A. V. and Rickard, K. L. and Rosse, C. (2003) Representing Complexity in Part-Whole Relationships within the Foundational Model of Anatomy. In Proceedings, American Medical Informatics Association Fall Symposium, pages 450-454.

[8] Rogers J, Rector A. GALEN's Model of Parts and Wholes: Experience and Comparisons. AMIA Annual Symposium 2000; Los Angeles, California; 2000.

[9] Werner Ceusters, Barry Smith and James Matthew Fielding, "LinkSuite?: Software Tools for Formally Robust Ontology-Based Data and Information Integration", in Proceedings of DILS 2004 (Data Integration in the Life Sciences), (Lecture Notes in Bioinformatics, 2994), Berlin: Springer, 2004, 124-139.

[9] Bittner, T., Axioms for parthood and containment relations in bio-ontologies, In Proceedings of KR-Med, 2004