

Interoperability, non-monotonicity and core ontologies

Bio-Ontology Research Group Leipzig
<http://bioonto.de>

Heinrich Herre*(heinrich.herre@imise.uni-leipzig.de), Robert Hoehndorf*[†]
(hoehndorf@eva.mpg.de), Janet Kelso[†](kelso@eva.mpg.de), Frank Loebe*
(frank.loebe@imise.uni-leipzig.de), Roberto Poli[‡](roberto.poli@soc.unitn.it)

1 Introduction

Recent trends have shown a growing body of biomedical knowledge being classified in ontologies. In supporting researchers who need flexible access to explore and analyse these heterogeneous data sources, it has become clear that there is a need to improve the inter-operability of biomedical ontologies. What is less obvious is the best means by which to achieve and promote this inter-operability while supporting the development of new ontologies that can be easily integrated into existing ontology frameworks.

Integrating biomedical ontologies has been recommended as a way to support the inter-operability required for research applications. The integration of ontologies was defined by J. F. Sowa [9] as the creation of a new ontology C from two ontologies A and B , such that C facilitates the inter-operability between information systems based on ontologies A and B . It should be noted that the inter-operability between ontologies A and B is weaker than the integration of A and B , in that inter-operability requires information systems based on A to translate statements from B while retaining parts of the semantics from B and vice versa.

We propose that there is a need for non-monotonic formalisms, as well as for an extended understanding of core ontologies, to facilitate the integration of biomedical ontologies.

*Onto-Med Research Group, Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig

[†]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology

[‡]University of Trento

2 Non-monotonic Aspects

We have focused on two sub-groups of biological ontologies. The first group describes a *canonical* or idealized view of a domain, for example, an ontology of canonical anatomy such as the Foundational Model of Anatomy (FMA) [5]. We call this group *canonical ontologies*. The FMA describes a prototypical, idealized human anatomy. Many ontologies concerned with structure, such as cell structure or anatomy, are canonical in this sense. A canonical anatomy ontology such as the FMA contains rules such as:

Every instance of a human body has as part an appendix. (1)

This rule does not necessarily apply to every real human body; an individual human body may *lack* an appendix as part. The rule, however, describes an idealized or *default* human.

This observation becomes more important when considered with the other group of ontologies, namely *pathology ontologies*. A pathology ontology describes *deviations* from an idealized case, including exceptions, abnormalities or malfunctionings. For example, the Mouse Pathology Ontology [6] includes the term “growth arrest” as a specific type of “general developmental defect”, and the Mammalian Phenotype Ontology [8] contains the term “absent tail” as a specific type of “abnormal tail morphology”. Pathology or disease ontologies cover a different kind of knowledge than that captured in canonical ontologies. For example, an individual may be both an instance of a human body as described in the FMA (which implies an appendix as part) and an instance of the category “human body with absent appendix”. Making the negation in “absent appendix” explicit yields an inconsistency. According to (1) all human bodies have an appendix as part, but there are instances of “human body” that do not have an appendix as part.

The integration of canonical and pathology ontologies is difficult, because inconsistencies are inevitable when both are interpreted according to classical logical formalisms. Little attention has been paid to resolving this issue. In order to prevent inconsistencies while preserving the intuition behind statements such as “a human has an appendix as part”, such statements in the canonical ontology must be weakened. We propose using *non-monotonic* logic which will allow the statements in a canonical ontology to be treated as true by default. Adding knowledge from a pathology ontology or using a statement involving the **lacks** relation (and therefore negation), may then invalidate the conclusions previously drawn, but without necessarily resulting in inconsistency.

Non-monotonicity is provided in a number of formalisms, for example, auto-epistemic logic, circumscription, default logic and context-based reasoning. For some of these, ready-to-use implementations exist. We use a prototype implementation of non-monotonic rules using the DLVHEX system [1]. DLVHEX provides the ability to add datalog programs to a description logic knowledge base using the answer set semantics. Our implementation is coupled with the

biological core ontology GFO-Bio¹.

3 Core Ontologies and Upper Domain Ontologies

There have been a number of attempts to integrate biomedical ontologies within a common ontological framework such as the BioTop Ontology [7], GFO-Bio, the Simple Bio Upper Ontology² and the ONIONS methodology [2]. Current approaches, such as those listed above, have been primarily intended to provide a domain top-level or a link between domain-specific categories and categories in some top-level ontology. We refer to these ontologies as *upper domain ontologies*. Integrated upper domain ontologies include categories from different (sub-)domain ontologies (GO, cell type, mammalian phenotype, FMA, etc.). These integrated ontologies may be designed following different design principles, exhibiting different views or are based on different granularities. Losing these aspects of domain ontologies in an integrated upper domain ontology possibly restricts the inter-operability of information systems based on the domain ontologies. In accordance with [10], we argue that it is necessary to make the nature of a domain precise and to provide a clear methodology for including categories in the ontology. We demonstrate our approach by explaining the basic ideas of the core ontology GFO-Bio, which is being developed by the Bio-Ontology Research Group Leipzig. GFO-Bio is centered around the concept of *autopoiesis*. The aspects of each domain ontology regarding autopoietic systems are made explicit as relations among the categories of the domain ontology and the category of autopoietic systems. This analysis is based on an ontological theory of levels of reality [4].

We believe that there is a difference between core ontologies, as intended by us, and upper domain ontologies. In contrast to upper domain ontologies like BioTop or SBUO, the core ontology GFO-Bio exhibits the following features:

- It provides a method for making the nature of the biological domain explicit by using the theory of autopoiesis. The category of autopoietic system is used as a category with a special status which we call a *principal category* for biology.
- GFO-Bio allows for an alignment of domain ontologies according to their view of the principal category of autopoiesis. This is achieved by specifying the relations – originating from the top-level ontology GFO [3] – that categories of domain ontologies have to the principal categories. In general, we consider this functionality a key feature of core ontologies, that benefits both the comparison and translation of domain ontologies.
- Beyond biology, ontologies of different domains can be integrated into a

¹<http://onto.eva.mpg.de/gfo-bio.html>

²<http://www.cs.man.ac.uk/~rector/ontologies/simple-top-bio/>

similar style, i.e., by relating their categories to the principal categories in GFO-Bio.

- GFO-Bio is an ontology of categories and individuals. In particular, it groups categories and their interrelations together in domains, using higher-order categories and relations.
- It provides a means for integrating ontologies of different biological domains based on different formalisms (first order logic, description logic, datalog, etc.).

4 Conclusion

We have outlined the problem of integrating canonical and pathology ontologies, which requires an extended logical setting for expressing ontologies. For a principled integration approach, we advocate using core ontologies, which add aspects of domain structuring and meta-categories to upper-domain ontologies.

References

- [1] T. Eiter, G. Ianni, R. Schindlauer, and H. Tompits. Nonmonotonic description logic programs: Implementation and experiments. In *Logic for Programming, Artificial Intelligence, and Reasoning*, volume 3452 of *Lecture Notes in Computer Science*, pages 511–527. Springer Berlin / Heidelberg, 2005.
- [2] A. Gangemi, D. M. Pisanelli, and G. Steve. An overview of the ONIONS project: Applying ontologies to the integration of medical terminologies. *Data & Knowledge Engineering*, 31(2):183–220, 1999.
- [3] H. Herre, B. Heller, P. Burek, R. Hoehndorf, F. Loebe, and H. Michalek. General Formal Ontology (GFO) – A foundational ontology integrating objects and processes [Version 1.0]. Onto-Med Report 8, Research Group Ontologies in Medicine, Institute of Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, 2006.
- [4] R. Poli. The basic problem of the theory of levels of reality. *Axiomathes*, 12(3-4):261–283, 2001.
- [5] C. Rosse and J. L. V. Mejino. A reference ontology for bioinformatics: The foundational model of anatomy. *Journal of Biomedical Informatics*, 36(6):478–500, 2003.
- [6] P. N. Schofield et al. Pathbase: a new reference resource and database for laboratory mouse pathology. *Radiation Protection Dosimetry*, 112(4):525–528, 2004.

- [7] S. Schulz, E. Beisswanger, U. Hahn, J. Wermter, A. Kumar, and H. Stenzhorn. From Genia to BioTop: Towards a top-level ontology for biology. In B. Bennett and C. Fellbaum, editors, *Formal Ontology in Information Systems: Proceedings of the Fourth International Conference (FOIS 2006)*, Baltimore, Maryland, USA, Nov 9-11, pages 103–114, Amsterdam, 2006. IOS Press.
- [8] C. L. Smith, C.-A. W. Goldsmith, and J. T. Eppig. The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information. *Genome Biology*, 6(1):R7, 2004.
- [9] J. F. Sowa. *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Brooks/Cole, Pacific Grove, 2000.
- [10] A. Valente and J. Breuker. Towards principled core ontologies. In B. R. Gaines and M. A. Musen, editors, *Proceedings of the 10th Knowledge Acquisition Workshop (KAW'96)*, Banff, Alberta, Canada, Nov 9-14, pages 301–320, 1996.