BioTop and ChemTop – Top-Domain Ontologies for Biology and Chemistry

Holger Stenzhorn
Stefan Schulz
University Medical Center Freiburg
Institute for Medical Biometry and Medical Informatics
Stefan-Meier-Straße 26
79104 Freiburg, Germany
holger.stenzhorn@uniklinik-freiburg.de
stschulz@uni-freiburg.de

Elena Beißwanger
Udo Hahn
University Language and Information Engineering (JULIE) Lab
Fürstengraben 30
07743 Jena, Germany
elena.beisswanger@uni-jena.de
udo.hahn@uni-jena.de

László van den Hoek
Erik van Mulligen
Erasmus MC - University Medical Center Rotterdam
Department of Medical Informatics PO Box 1738
3000 DR Rotterdam, The Netherlands
lhoek@science.uva.nl
e.vanmulligen@erasmusmc.nl

1. INTRODUCTION

The recent flood of data and factual knowledge in the area of biomedicine requires some principled approaches to their proper analysis and management. A major cornerstone in this effort constitutes the precise and complete description of the fundamental entities within this domain.

But although this is generally accepted, current ontology developments often do not adhere to some of the elemental ontology design principles: For example, even prominent domain terms lack precise (logical) definitions and descriptions in many cases. Such issues impede the semantic standardization needed for adapting ontologies in data/knowledge management tasks. Rather, they lead to inconsistencies within and between ontologies and so to a situation of fragmentation ontologies are intended to solve in the first place.

In light of this we introduce BioTop and ChemTop [4], two top-domain ontologies define and describe the foundational entities needed to characterize phenomena in the life sciences area. The basic intention of those ontologies is twofold: Firstly, to serve as top-level basis for creating new focused domain ontologies or as aid for ameliorating existing ones. Secondly, to be used as "semantic glue" in aligning and mapping biomedical ontologies from disparate provenience.

2. METHOD

To clarify the role of BioTop and ChemTop as top-domain ontologies, we make the following distinction:

Top Ontologies contain only a very restricted set of high-level, generic classes such as "Continuant", "Function" or "Object", not related to any particular domain.

Top-domain Ontologies hold the general core classes of a given domain to interface both top and domain ontologies, like "Organism", "Tissue" or "Cell" for biology.

Domain Ontologies include only domain-specific classes to comprehensively describe a certain domain, e.g. "Antisense RNA Transcription" or "DNA Replication" from the Gene Ontology (GO) [1] for gene-related research.

To create our ontologies we investigated a sample from the domain ontologies of the Open Biomedical Ontologies (OBO) Foundry [5] — a community initiative to create interoperable biomedical ontologies adhering to good ontology design principles. There we looked at the commonalities of the Gene Ontology (GO), the Cell Ontology (CL), the Chemical Entities of Biological Interest (ChEBI) and others and used this as one basis to devise an ontology able to align several biomedical ontologies together. As second basis and "scaffold" for our work we employed GENIA [4], a lightweight ontology applied for corpus annotation in biological text mining contexts, which we wanted to comprehensively and formally redesign and expand.

The mentioned alignment between the employed OBO Foundry ontologies enables users to get an integrated view on them and renders possible a more systematic and concerted application. In this vein, one goal is to continuously consider more and more OBO Foundry (but also external biomedical) ontologies in our work. One current "external" mapping of BioTop is to the UMLS Semantic Network [3] whereby a bridge and possible integration from this resource to the OBO Foundry ontologies is created.

In our ontology design we complied with the OBO Foundry principles. Some of the principles coined for our case include:

- Both BioTop and ChemTop have a clearly defined and delineated content matter that does not overlap with other ontologies found in the OBO Foundry.
- BioTop and ChemTop are open-source and accessible to everybody. To adopt them in other projects a developer must only acknowledge its original source and agree not to alter and distribute the modified ontology.
- The implementation of BioTop and ChemTop is based on OWL-DL, a formally defined language and official standard for creating Semantic Web ontologies published by the World Wide Web Consortium (W3C).
- The ontologies contain clear textual definitions to avoid the ambiguity many terms possess in the life sciences.

This makes the ontologies understandable not only to computer tools but also to human (experts).

- BioTop and ChemTop are based on the top ontology Basic Formal Ontology (BFO) [2], successfully applied by many OBO Foundry ontologies and thus showing its practical applicability to biomedicine. We further used the relations from the Relation Ontology, also a standard ontology proposed by the OBO Foundry.

3. IMPLEMENTATION

As said above, BioTop and ChemTop are implemented in OWL-DL. Besides fulfilling the OBO Foundry principle for using a formally-defined representation language, this entails the availability of a wealth of documentation and supporting tools. For example, it was crucial to continuously classify our ontologies during development with a terminological reasoner in order to check for potential bugs in the rather complex encoding of axioms. In our implementation we also refrained from asserting multihierarchies but let the reasoner infer those from the axioms specified for the classes. Most importantly, OWL-DL makes it possible to align BioTop and ChemTop with other existing biomedical ontologies implemented in OWL-DL.

4. MODULARIZATION

Pursuing the principle to describe as many classes as possible in terms of necessary and sufficient conditions and realizing the need to cover the relation between biology and chemistry more thoroughly as we considered further OBO ontologies, BioTop needed to be significantly expanded towards biochemistry after its initial version.

But this expansion was problematic in that BioTop lost more and more its focus on biology. To countersteer this, we decided to reenginer BioTop towards a systematic modularization to highlight again its original focus. To this end, a significant amount of classes contained in BioTop was migrated into a newly created ontology, called ChemTop. In the modularization process, the following principles were obeyed:

- The boundary of all modules should coincide with a particular subdomain, e.g. biology vs. chemistry.
- All modules must follow the same (OBO Foundry) ontology design principles.
- The size of each module should be such that it can be handled easily by humans and tool, e.g. reasoners.
- Modules covering neighbor subdomains may exhibit a limited (and documented) degree of overlap.
- Bridging files link the various modules and apply the usual OWL import feature in straightforward fashion.

The following modules emerged as result of our modularization efforts. Fig. 1 schematizes how the core BioTop/ChemTop modules are related to each other but also how the top ontology BFO as well as the domain ontologies are linked.

`biotop` **and** `chemtop` self-standing BioTop and ChemTop ontologies (without specific chemistry resp. biological classes or reference to BFO and RO)

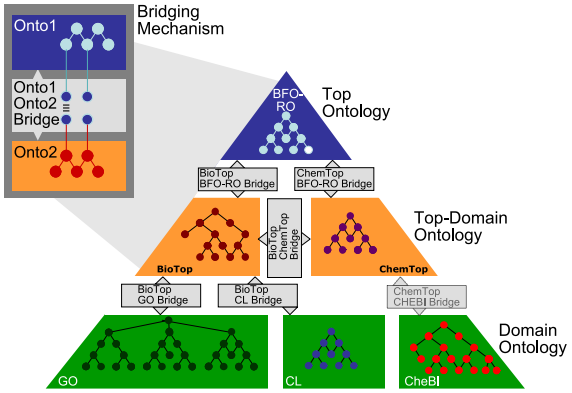

Figure 1: BioTop and ChemTop linked to top ontology BFO and domain ontologies through bridges

`biotop-chemtop` bridge that connects both BioTop to ChemTop (without reference to BFO and RO)

`biotop-bfo-ro` **and** `chemtop-bfo-ro` bridges that connect BFO and RO with BioTop and ChemTop respectively

`biotop-chemtop-bfo-ro` bridge that includes `biotop-bfo-ro` and `biotop-chemtop` and connects BioTop with BFO and RO and also ChemTop.

5. AVAILABILITY

All BioTop and ChemTop related material can be found at `http://purl.org/biotop` and `http://purl.org/chemtop`.

6. REFERENCES

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

[2] Pierre Grenon, Barry Smith, and Louis Goldberg. Biodynamic ontology: Applying BFO in the biomedical domain. In Domenico Pisanelli, editor, *Ontologies in Medicine: Proceedings of the Workshop on Medical Ontologies*, pages 20–38. IOS Press, 2004.

[3] Donald Lindberg, Betsy Humphreys, and Alexa McCray. The Unified Medical Language System. *Methods of Information in Medicine*, 32(4):281–291, 1993.

[4] Stefan Schulz, Elena Beisswanger, Udo Hahn, Joachim Wermter, Anand Kumar, and Holger Stenzhorn. From GENIA to BIOTOP: Towards a top-level ontology for biology. In Brandon Bennett and Christiane Fellbaum, editors, *Proceedings of the 4th International Conference on Formal Ontology in Information Systems*, pages 103–114. IOS Press, 2006.

[5] Barry Smith, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis J Goldberg, Karen Eilbeck, Amelia Ireland, Christopher Mungall, Neocles Leontis, Philippe Rocca-Serra, Alan Ruttenberg, Susanna-Assunta Sansone, Richard Scheuermann, Nigam Shah, Patricia Whetzel, and Suzanna Lewis. The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25(11):1251–1255, 2007.