Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-441

The BioTop Family of Upper Level Ontological Resources for Biomedicine

Stefan SCHULZ[a,1], Martin BOEKER[b] and Catalina MARTINEZ-COSTA[a]
[a,b] *IMI – Medical University of Graz, Austria*
[a] *IMBI – University of Freiburg, Germany*

Abstract. BioTop is a domain upper level ontology for the life sciences, based on OWL DL, introduced ten years ago. This paper provides an update of the current state of this resource, with a special focus on BioTop's top level, BioTopLite, which currently contains 55 classes, 37 object properties and 247 description logics axioms. A bridging file allows harmonising BioTopLite with the classes of Basic Formal Ontology BFO2. The updated OWL resources are available at http://purl.org/biotop. They build the core of several upper level ontological artefacts including bridging ontologies to other upper level resources.

Keywords. Upper-level Ontology, Semantic Interoperability, Description Logics

1. Introduction

Ontology engineering requires knowledge (i) in the domain to be represented, (ii) in the representation language, (iii) in the discipline of Applied Ontology [1], and (iv) in building software following design specifications. Upper-level ontologies (ULOs) guide this process by supporting the creator of domain ontologies with a consistent framework, which can guarantee for interoperability if used by different domain ontologies.

Although ULOs are often seen as domain-independent, popular ULOs like DOLCE [2] and BFO [3] target cognitive sciences and natural science, respectively. In the past, several ULOs have been created explicitly for biology and medicine, e.g. the GALEN upper level [4], the UMLS SN [5], the OBO RO [6], GFO-BIO [7], and SIO [8].

Whether ULOs positively influence the resulting artefacts or whether they contribute to excessive complexity that does not pay off from an application point of view, has been controversially discussed for two decades. However, many application ontologies exhibit a rather naïve and haphazard makeup that does not commit to any pre-existing ULO. "Quick and dirty" solutions may solve immediate problems, but they fail when it comes to maintenance, extension, modularisation and re-use. This is our rationale for ULOs. Their use requires some additional effort, involving a process towards "ontological thinking". The constraints ULOs impose on the modeller may provoke adverse reactions, but they keep them on the track and prevent them from committing modelling errors. This is the current mission of the biological upper level ontology BioTop and its overarching component BioTopLite. The objective of this paper is to provide an update of BioTop, ten years after its first publication and use.

[1] Corresponding author: Stefan Schulz, Medical University of Graz, Austria; E-mail: stefan.schulz@medunigraz.at

2. Evolution

BioTop was first launched in 2006 in OWL DL [9]. Its design has been inspired by the analysis of the GENIA ontology for cell signalling [10]. In response to a series of fundamental design problems with this resource, BioTop was proposed as an alternative, but it soon went beyond the scope of GENIA, in order to cover a broad range of categories relevant for application in all areas of life sciences. BioTop was not intended to compete with established ULOs, but rather to integrate with them. Therefore, its developers created bridging ontologies to DOLCE, BFO, and RO, and left the uppermost level deliberately flat, thus lending itself as a top-level layer for domain ontologies, without coercing developers into a certain, pre-existing ULO.

BioTop has always set a strong focus on constraining axioms, which at the time was only available for DOLCE but not for BFO and RO. This changed its scope to the integration of more classes considered fundamental for the representation of biological entities. The attempt to provide full definitions led to a further expansion into the realm of biochemistry. Experiences from the @neurIST project [11] and BioTop's use as a top-level ontology in DebugIT [12] revealed performance problems, which were mitigated by factoring out most of the chemistry classes into a separate ontology named ChemTop. This module was however not further developed due to the re-emerging ChEBI ontology [13]. Another alignment project integrated BioTop with the UMLS semantic network (SN)[14]. The resulting ontology showed, again, considerable performance problems, so that its intended use for validating UMLS sources had to be postponed. However, the task of covering the whole content of SN provided a good external criterion for content scoping [15]. BioTop's growth and ensuing performance issues motivated the creation of a “lite” version, with a minimum of content but still sufficient to link to life sciences ontologies and, nonetheless, to provide a sound framework and guidance for developers. This version was then released as BioTopLite. It was used in several experimental ontologies in which future evolutions of SNOMED CT were tested [16][17]. BioTopLite was furthermore used in experimental ontologies within the EU SemanticHealthNet project [18], in the TNM ontology [19] and in the CELDA ontology [20]. The dynamic evolution of BioTopLite, until its current version BioTopLite2, abbreviated as BTL2, led to some disruptions regarding the original BioTop ontology. Recently, however, BioTop was stripped by its upper level classes and most relations, importing BTL2, instead. Currently BTL2 has 55 classes, 37 object properties and 247 logical axioms, whereas and BioTop (without the imported BTL2 component) has 358 classes, 46 object properties and 580 logical axioms.

3. BioTop: Current Status

Figure 1 shows the current BTL2 class and relation trees. In the following, we briefly describe some of the important features:

Intuitive labels: BTL2 sets a focus on intuitive labels (*Material object*, *Quality*, *Information object*, *is part of, has participant*), which give an intuition of the meaning of the respective class or relation and is therefore better suited as "entry points" for developers compared to the class and relation names in BFO and DOLCE.

Simplified relation (object property) hierarchy: The earlier distinction between process parts and object parts, as well as between parts and proper parts has turned out to confuse rather than to help the users. We found out that one relation pair *has part / is*

part of is sufficient. Constraining axioms, necessary to control the domains and ranges of these relations were included at the class level and as general class inclusion axioms.

The class *Condition*: There is an inherent ambiguity (or "logical polysemy", according to [21] of medical terms: "Ulcer" may denote both an ulceration process as well as its result, e.g. a gastric ulcer. "Allergy" can be interpreted as an allergic disposition or as allergic manifestation.

Figure 1. BTL2 class hierarchy (left), relation hierarchy (right)

Such category distinctions (as, e.g. proposed by OGMS [22]) are not necessary in many clinical reasoning patterns, which motivated us to add the class *Condition* as a disjunction of *Material entity*, *Process*, *Disposition*, and *Function*.

The class *Situation*: Medical discourse often refers to time segments of a (biological) life, defined by the presence of a condition (in the above sense), as empirical investigations with SNOMED CT [23] have shown [24]. E.g., "gastric ulcer" would therefore refer to the life segment, called "clinical situation" or "clinical life phase", in which a gastric ulcer process unfolds, or in which a gastric ulcer structure is present. This motivated the addition of the class *Situation*, which can therefore be used as a robust bridge to what is called disorders, diseases, findings, symptoms etc. in clinical ontologies.

The classes *Universe* and *Subatomic particle*: notorious errors in OWL ontologies derive from the confusion between existential and value restrictions ("some" and "only" in Manchester syntax). In combination with transitive relations such as '**has part**', this often leads to incorrect expressions like:

'*Cell culture*' subclassOf '**has part**' only *Cell* (1)

By adding the upper level axiom:

'*Material object*' subclassOf '**has part**' some *Subatomic particle* (2)

with the latter being disjoint from other material entities a logical error occurs, which forces the modeller to revise expression (1).

Temporally qualified continuants: In 3-dimensionalist ontologies a known issue is the representation of relations between continuants, i.e. objects that exist during time, undergo temporal change and have no temporal parts like processes or time intervals. This would require three-valued relations, such as **has part** (*a, b, t*) with time as third argument. OWL-DL does not support this. The lack of temporal qualification provokes ambiguities and may lead to wrong entailments. BioTopLite 2 mitigates lack of ternary, time-dependent relations in OWL-DL by introducing time-dependent entities, subscribing to the principle of temporally qualified continuants, still under development [25]. The class '*Particular at some time*' is of no real ontological relevance but it has proven useful as a means to enforce that instances of time-dependent classes be placed in a temporal context. Class-level axioms are such that the reasoner infers that the relata must be of the type '*Particular at some time*', cf. examples in [26]. Since 2016, the following, updated ontologies are available via http://biotopontology.github.io/ which is mapped to the URL http://purl.org/biotop
The following updated ontologies are available

- BioTopLite2: http://purl.org/biotop/btl2.owl
- BioTop (importing btl2): : http://purl.org/biotop/biotop.owl
- BioTopLite2 - BFO - bridge: http://purl.org/biotop/btl2-bfo.owl

An update of the bridge files to RO, DOLCE, and UMLS SN is under way.

4. Conclusion

Ten years' experience with BioTop has shown the need for adaptation of an upper level ontology to the user's context thus providing additional, domain-specific content. A recent step towards modularization was the re-harmonisation of the original BioTop

ontology with the "lite" version BTL2, which had more dynamically evolved, primarily by disjunctive classes, simplified relations and the definition of all entities as '*Particulars at some time*'. This approach allows differentiating, by simple means, between relations that hold at some times and those that (generically) hold at all times. A gradual update of existing bridging files to other ontologies is under way, targeting the release of a family of upper-level ontological resources for biomedicine to be released in 2017.

References

[1] Smith B. Applied Ontology: A new discipline is born. Philosophy Today 12 (29):5-6 (1998)
[2] Borgo S, Masolo C. Ontological foundations of DOLCE. In Staab S, Studer R (eds.), Handbook on Ontologies (Second Edition), Springer Verlag, 2009: 361-382.
[3] Grenon P, Smith B. (2004) "SNAP and SPAN: Towards dynamic spatial ontology", Spatial Cognition and Computation, 4: 1, 69-103.
[4] Rector A, Rogers J. Patterns, properties and minimizing commitment: Reconstruction of the GALEN Upper Ontology in OWL. EKAW*04 Workshop on Core Ontologies. http://ceur-ws.org/Vol-118/
[5] McCray AT. An upper-level ontology for the biomedical domain. Comp F Genomics. 2003; 4 (1): 80-84
[6] Smith B wt al.Relations in bio-medical ontologies. Genome Biology 2005; 6 (5): R46.
[7] Hoehndorf R, Loebe F, Poli R, Herre H, Kelso. GFO-Bio: A biological core ontology. Applied Ontology, 2008, 3 (4), 219-227.
[8] Dumontier M (2013). The Semanticscience Integrated Ontology (SIO) http://code.google.com/p/semanticscience/wiki/SIO .
[9] Beisswanger E, Schulz S, Stenzhorn H, Hahn U: BioTop: An upper domain ontology for the life sciences. Applied Ontology, 2008; 3 (4): 205-212.
[10] Rak, R., Kurgan, L., & Reformat, M. (2007). xGENIA: A comprehensive OWL ontology based on the GENIA corpus. *Bioinformation*, *1*(9), 360.
[11] Boeker M , Stenzhorn H, Kumpf K, Bijlenga P, Schulz S, Hanser S. The @neurIST ontology of intracranial aneurysms: Providing terminological services for an integrated IT infrastructure. AMIA Annu Symp Proc. 2007;2007:56–60.
[12] Schober D et al. The DebugIT core ontology: semantic integration of antibiotics resistance patterns. Stud Health Technol Inform. 2010;160(Pt 2):1060-1064.
[13] Hastings J et al. The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Res. 2013 Jan;41
[14] Schulz S et al. Alignment of the UMLS semantic network with BioTop: methodology and assessment. Bioinformatics. 2009 Jun 15; 25 (12): i69-76.
[15] Schulz S et al. Scalable representations of diseases in biomedical ontologies. *Journal of Biomedical Semantics.* 2011, *2*(2), 1.
[16] Schulz S, Martínez-Costa C. Harmonizing SNOMED CT with BioTopLite: An exercise in principled ontology alignment. *Studies in health technology and informatics 2015*, *216*, 832.
[17] Cheetham E et al. Formal representation of disorder associations in SNOMED CT. Proc. of International Conference on Biomedical Ontology (ICBO). 2015
[18] SemanticHealthNet Network of Excellence (2013). http://www.semantichealthnet.eu/
[19] Boeker M et al. TNM-O an ontology for the Tumor-Node-Metastasis classification of malignant tumors: a study on rectal cancer.
[20] Seltmann S et al. CELDA – an ontology for the comprehensive representation of cells in complex systems. BMC Bioinformatics. 2013 Jul 17;14:228
[21] Pustejovsky J. The Generative Lexicon, MIT Press, 1995., xi+ 298pp. *Studies in English literature*, *1999*, 109-116.
[22] Ceusters W, Smith B. Foundations for a realist ontology of mental disease. J Biomed Semantics. 2010 Dec 9;1(1):10
[23] SNOMED CT. International Health Terminology Standards Development Organisation (IHTSDO). http://www.ihtsdo.org/snomed-ct
[24] Schulz S, Rector A, Rodrigues JM, Spackman K. Competing interpretations of disorder codes in SNOMED CT and ICD. AMIA Annu Symp Proc. 2012; 2012: 819-827.
[25] Grewe N et al. Permanent generic relatedness and silent change. Competition workshop at FOIS 2016.
[26] Schulz S, Boeker, M. BioTopLite: An Upper Level Ontology for the Life Sciences. Evolution, Design and Application. In *GI-Jahrestagung* 2013 (pp. 1889-1899).