ICBO: International Conference on Biomedical Ontology
July 28-30, 2011 · Buffalo, NY, USA

Higgs Bosons, Mars Missions, and Unicorn Delusions: How to Deal with Terms of Dubious Reference in Scientific Ontologies

Stefan Schulz[1,2], Mathias Brochhausen[3], Robert Hoehndorf[4]

[1]Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria
[2]Institute for Medical Biometry and Medical Informatics, Freiburg University Medical Center, Germany
[3]Institute for Formal Ontology and Medical Information Science, Saarland University, Saarbrücken, Germany
[4]Department of Genetics, University of Cambridge, UK

Abstract. Realist ontologies claim to represent what exists. Scientific discourse, however, often contains terms of dubious reference when describing current or past hypotheses, plans, or ideas. We present a framework in which a realist ontology is embedded in a description logics theory, the latter being indifferent regarding the existence of class members. It therefore may include units for various kinds of such terms. Using a taxonomy of terminological units we are able to distinguish between different kinds of classes in description logics theories based on whether the classes are believed to have instances or not. We also demonstrate how discourse using terms of dubious reference can be represented without departing from the principle of realist ontologies. An example OWL file can be downloaded from: http://purl.org/steschu/misc/ICBO2011.

1 Introduction

Biomedical ontologies are increasingly advocated for the representation of scientific discourse [1], available both as human language content in free-text narratives and the structured content in scientific databases. For this purpose considerable effort has been put into a steadily growing repository of ontologies in biology and medicine, made available through the NCBO BioPortal [2] and the OBO Foundry [3,4].

Upper level ontologies like the Basic Formal Ontology (BFO) [5] or DOLCE [6] ontology, or domain specific top levels like BioTop [7], GFO-Bio [8] and the OBO Relation ontology [9] have been proposed as ordering and constraining principles guiding the development of ontologies in the field of biomedicine.

The OBO Foundry [4] is a collection of biomedical ontologies, following explicit guidelines and adopting a systematic approach to ontology building which is grounded in theoretically well-founded upper-level ontologies and quality criteria. The OBO Foundry recommends BFO as upper level ontology, which adopts a position of *ontological realism*. In contrast to the proposal that biomedical terminologies should be concept-oriented [10], ontological realism claims that classes in an ontology extend *universals*[1] that exist in reality [12]. Universals, often equated with types, are considered to be entities that exist in their instances. Thus, there can be no uninstantiated universal.

A representational framework increasingly used for biomedical ontologies is Description Logics (DL) [13]. In particular the DL dialects which are standardized by the Semantic Web community as the Web Ontology Language (OWL) [14] have no means to differentiate between universals and non-universals, or between classes that have members in the world and classes that haven't. For example, we can create the classes *Animal* and *Plant* intended to correspond to the universals <u>*Animal*</u>, and <u>*Plant*</u>, respectively. DL theories tolerate nonsense classes like *Planimal* (the conjunction of *Animal* and *Plant*), which have no instances at all. We can state that *Animal* and *Plant* have no instances in common by

[1] As a less restrictive alternative to universals, the more general term *repeatables* has also been recommended [11]. For the sake of simplicity we here use the term "universal" in a broader sense.

declaring them as disjoint. Here, the DL axioms '*Animal* and *Plant* SubClassOf *Nothing*' and '*Planimal* SubClassOf *Nothing*' are equivalent.

DL theories are *ontologically neutral:* there is no impediment to creating classes that are not the extensions of universals, or to creating classes that have no instances. Whereas the latter is contrary to the basic assumption of realist ontologies, the former construct, so-called defined classes [15], e.g. '*Person born in Belgium*' have been accepted in principle for inclusion in OBO Foundry ontologies [16].

Scientific discourse and discourse in clinical practice often contains terms for which the existence of instances in this world is hypothetical, unknown, purely fictional, or scientifically implausible. End users and application builders may want to use some kind of semantic artefact to link these terms to, for example in the semantic annotation of natural language text.

Ontologists who defend a realist view are pressed to find ways to properly represent assertions using terms for which it is believed or known that no instances exist, without abandoning their principles. In this paper, we investigate how this goal can be achieved. We will use the name *terms of dubious reference* (TDRs) for language expressions which do not denote anything in reality.

We will outline a formal approach of how TDRs are semantically accounted for in description logics based theories in a way which as little as possible contradicts the principles of realist ontologies. To this end we will sketch a typology of TDRs which are commonly used in scientific and clinical discourse.

2 Typology of Terms of Dubious Reference (TDRs)

We define as TDRs those expressions in a natural or logic language, for which the reference to something that exists or has existed is ruled out or hypothetical. We distinguish (see Table 1 for examples):

1. Speculative entities, whose existence cannot be derived from accepted scientific theories, but which are referred to in the discourse in several branches of medicine, e.g. in family practice or complementary medicine.
2. Associations of signs or symptoms which are referred to as a unity according to culture-specific constructs about health and disease, although not backed by any scientific theory.
3. Hypothetical entities, whose existence is postulated by accepted scientific theories, although there is no evidence for their existence. Suitable experiments are expected to collect this evidence.
4. Artefacts or processes that may come into existence only after the execution of a plan in the future. This also includes non-existent chemical compounds, whose structures do not offend the rules of valence but which have never been prepared or found in nature.
5. Fictional concepts, whose existence is not even speculatively assumed but which may be used as metaphors or imaginary patterns.
6. Historic concepts with consensus about their obsolescence.

We exclude from this typology shortcut interpretations suggested by naively interpreted syntactic compositions (missing thumbs, prevented pregnancies, …), as well as terms including negative statements (non-insulin-dependent diabetes mellitus), because in all those cases there can be referents in reality once we avoid interpreting these expressions at face value.

3 TDRs in Biomedical Ontologies

In the following we will propose a formal solution as to how TDRs can be accommodated in realist ontologies based on classes in description logics. We distinguish between two approaches. In the first approach, the information that a class is not the extension of a universal is not made explicit. In the second approach, it is conveyed by an additional ontology layer.

Throughout this paper we clearly distinguish between *realist ontologies* and *DL models*. Realist ontologies may be – at least partly – represented as DL models, whereas DL models may include units which are irrelevant for realist ontologies.

Category of Entity / Concept of Reference	Term	Definition	Occurrences in MEDLINE
Speculative, not scientifically plausible	Qi	The vital life force in the body, supposedly able to be regulated by acupuncture	299
Hypothetical	Higg's Boson	Hypothetical elementary particle predicted to exist by the standard model of particle physics.	19
Culture-specific constructs	Koro	Syndrome of someone's belief that his/her external genitals disappear	139
Subjects of unrealized plans	Manned mars mission	Process during which a human disembarks on Mars	8
Fictional concepts	Unicorn	Horse with a horn	65
Obsolete concepts of historic interest	Phlogiston	a fire-like element released during combustion	18

Table 1. Typology for NRUs and their occurrences in the MEDLINE abstracts

3.1 Ontology Without Explicit Reference to Universals

The main tenet is the following: While some classes in a DL theory may correspond to universals, there are other classes that do not. In the subsequent formalisms we will use Underline for universals, *Italics* with leading capital for classes extending universals and *lower case italics* for classes not extending universals. **Bold face** is used for individuals and relations involving individuals, for all other relations *lower case italics* are used. For DL expressions we use the OWL Manchester syntax [17].

The extension function *ext* relates a universal to the class of instances of this universal, e.g.

J EquivalentTo *ext* (J) (1)

K EquivalentTo *ext* (K) (2)

Using DL syntax we can easily generate new classes that are not the extensions of universals, assuming that the set of universals is not closed under DL operators:

b EquivalentTo *J* and **rel** some *K* (3)

However, the DL syntax does not include special symbols for universals or the extension relation *ext*. Formula (1) and (2) are therefore not made explicit, and we therefore refrain from an explicit reference to universals. The following definition, based on the classes *MarsMission* and *Human* as well as the relation **hasParticipant**, does not make any claim on the existence of instances of *mannedMarsMission*:

mannedMarsMission EquivalentTo *MarsMission* and **hasParticipant** some *Human* (4)

This class has no instances in our world yet, but may have some instances in the year 2030 (or never). As soon as an entity **m** satisfies the defining conditions of the class *mannedMars Mission,* it can be classified as *MannedMars Mission* and a new universal MannedMars Mission may emerge with the first instance of this class. Then, the class has instances that are in the extension of a universal:

MannedMarsMission EquivalentTo *ext* (MannedMarsMission) (5)

The class *MannedMarsMission* can be used in new definitions, e.g.

MannedMarsMissionPlan EquivalentTo Plan and **realizedBy** only *mannedMarsMission* (6)

Here we define what a manned mars mission plan is, regardless of whether such a plan has ever been drafted, let alone realized. If desired, we could replace every defined term with its definiens, so that only classes which extend universals are explicitly named in the DL theory:

MannedMarsMissionPlan EquivalentTo *Plan* and **realizedBy** only (*MarsMission* and **hasParticipant** some *Human*) (7)

The latter definition would be in line with the principles of realist ontologies, because it does not name any non-referring unit.

These examples show that the application of DL operators to a set of classes that are the extensions of universals allows the straight-forward generation of classes that do not represent universals. These classes can then be used for inferences. For example, from the definition of *MannedMarsMissionPlan*, together with the definition:

MarsMissionPlan EquivalentTo *Plan* and **realizedBy** only *MarsMission* (8)

we can infer that

MannedMarsMissionPlan SubClassOf *MarsMissionPlan* (9)

While mars missions are scientifically feasible, unicorns are scientifically implausible. Nevertheless, a psychiatrist may need to use the term "unicorn" when describing the topic of delusional disorder of patients who feel themselves persecuted by unicorns.

In contrast to a realist ontology, a DL theory may contain a defined class *unicorn*:

unicorn EquivalentTo *Horse* and **hasPart** some *Horn* (10)

without stating that it has instances. Such a statement asserts that, if there are ever horses with horns, than they would be classified as unicorns. This demonstrates that classes in a DL theory may specify the meaning of terms within a vocabulary without any commitment to whether the terms in the vocabulary refer to universals or not.

When describing a unicorn delusion we can refer to the class *unicorn* by a universal restriction without asserting any existence of instances of *unicorn*, e.g.

UnicornDelusion EquivalentTo *Delusion* and **isAbout** only *unicorn* (11)

We can again replace all defined terms with their definiens to name only those classes in the definition that we believe to represent universals:

UnicornDelusion EquivalentTo *Delusion* and **isAbout** only (*Horse* and **hasPart** some *Horn*) (12)

This would then be compatible with the tenet of realist ontologies. We could even go further and define, within the realist framework, the class *PinkUnicornDelusion*:

PinkUnicornDelusion EquivalentTo *Delusion* and **isAbout** only (*Horse* and **hasPart** some *Horn* and **bearerOf** some *PinkColor*) (13)

A description logics classifier then computes

PinkUnicornDelusion SubClassOf *UnicornDelusion* (14)

because every instance of *PinkUnicorn* would be an instance of *Unicorn*, provided there were such instances.

It may not always be possible to avoid TDRs in ontologies. Therefore we may discuss whether a realist ontology implemented in DL should be allowed to have extensions for classes for which we do not yet know whether they have members. For example, whether the scientifically postulated Higgs bosons exist is one of the most exciting research questions in modern physics.

ScalarBoson SubClassOf *Boson* (15)

higgsBoson SubClassOf *ScalarBoson* and *NeutralParticle* (16)

The latter axiom uses the intersection between the (non-empty) classes *ScalarBoson* and *NeutralParticle*. We can define the class

HiggsBosonResearchProject EquivalentTo *ResearchProject* and **isAbout** only *higgsBoson* (17)

which is then classified a subclass of, e.g. *BosonResearchProject*, defined as

BosonResearchProject EquivalentTo *ResearchProject* and **isAbout** only *Boson* (18)

These examples show how DL models can produce correct inferences using classes that are the extensions of TDRs. They also demonstrate how DL expressions with the same power can be built without such classes, thus corresponding to the principles of realist ontologies. An important modeling feature is the universal quantifier used with the relations **isAbout** and **realizedBy**. In opposition to the information artifact ontology (IAO) we do not claim that the ranges of these relations always have instances.

3.2 DL Theory Extended by a Taxonomy of Terminological Units

The above approach treats TDRs as classes in a DL theory which can be distinguished from classes extending universals within the formal theory itself. Although this may be sufficient for many use cases, it may be important in certain cases to explicitly state whether a class is believed to represent a universal or not. Therefore, we will introduce a simple theory of terminological units which allows explicit references to TDRs. Such a theory requires an extension to the upper level and introduces a root class *TerminologicalUnit* that has terms as instances. These terms can then be explicitly asserted to have referents (in this world) or not, which makes it possible to distinguish between terms that represent universals and those that do not.

Our proposed extension uses a **has Instance** relation that is rarely explicitly introduced in OWL theories. The main reason for this is that the notion of instantiation is commonly considered as part of DL semantics (in the form of class membership as determined by an interpretation function). However we here make a distinction between instantiation (in this world) and class membership. Whereas class membership relates a class to an individual, instantiation relates representational units (which are instances of the *TerminologicalUnit* class) to individuals.

Modifying the above formatting conventions, we use Underline for terms in a broader sense. Terminological units can refer to diverse flavors of universals, but also simple names or concepts (as entities of thought). The common characteristics of terminological units is that they have classes as their extensions. Classes are not required to have members. In contrast to sets, classes are defined intensionally. Therefore, empty classes are not necessarily identical. We introduce the following ground axiom: All members of the class *J*, which is the extension of a terminological unit J, are instances of J.

$$\forall \mathbf{x}{:}\mathbf{hasMember}(J,\mathbf{x}) \Leftrightarrow \mathbf{hasInstance}(\underline{J},\mathbf{x}) \quad (19)$$

Whereas the class instantiation is a built-in OWL feature, the explicit instantiation relation between a terminological unit and its instances needs to be manually asserted. Due to OWL restrictions this requires that terminological units are treated as individuals. We express this in DL as follows (with **instanceOf** being the inverse relation of **hasInstance**):

J EquivalentTo **instanceOf** value J (20)

This is to be read: every member of the class *J* is an instance of some terminological unit J.

We therefore propose an extended DL theory with a bipartition into (i) *Particular* and (ii) *TerminologicalUnit*, with a taxonomy of terminological units under the latter. The terminological units that correspond to the classes below *Particular* are represented as individual members of the class *Terminological Unit*.

Let us illustrate this extended theory using the example *Animal*. AnimalTerm is an instance of the class *Universal*, which is a subsumed by *DenotingTerminologicalUnit* and *TerminologicalUnit*.

AnimalTerm rdfs:type *Universal* (21)

Animal SubClassOf **instanceOf** value AnimalTerm (22)

We can express classes that have no instances as follows:

EmptyClass EquivalentTo*Terminological Unit* and not **hasInstance** some *Thing* (23)

Note that this is not equivalent to the class *Nothing*, which is a class that could not have any instances (via rdfs:type).

EmptyClass is therefore a class with an empty extension with respect to the explicitly introduced **hasInstance** relation. With this statement we can use an OWL reasoner such as HermiT [18] to ensure that this class is not instantiated (with respect to **hasInstance**), while it is at the same time possible to create subclasses and reason over them.

4 Implementation

The examples in section 3 were implemented in an OWL file with DL SROI expressivity (see http://purl.org /steschu/misc/ICBO2011.

Fig. 1 demonstrates the three terms PinkElephantTerm, PinkUnicornTerm, and UnicornTerm as members of the class *Non DenotingTerminologicalUnit*. The latter is defined as a terminological unit which has no instances. As a consequence of the axiom

(**instanceOf** value UnicornTerm) EquivalentTo *Unicorn* (24)

the class *Unicorn*, by design placed under *Object,* becomes inconsistent (subsumed under *Nothing*).

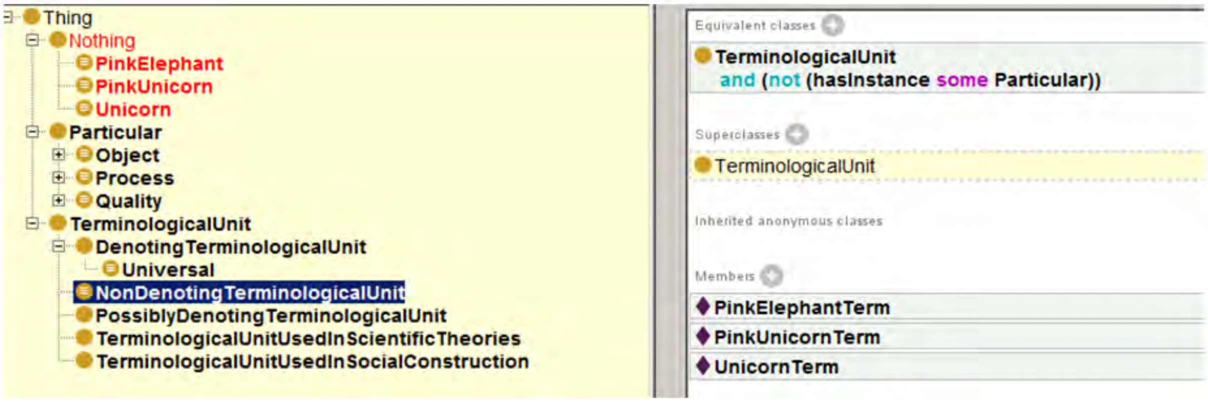

Figure 1. Example Ontology

5 Discussion

Fig. 1 illustrates our example ontology including a hierarchy of terminological units (Fig. 1). We consider the current arrangement of classes under TerminologicalUnit as preliminary and plan to harmonize it, in the future, with other proposals such as in [19]. So far, the following classificatory axes are proposed.

Possibility for existence. The first axis is the possibility for existence: is it logically possible for an instance of a subclass of *EmptyClass* to exist or not? Logical impossibility represents an unsatisfiable class, i.e., a logical contradiction. For example, all processes of raining and not-raining are logically impossible, because the definition of the class contains a contradiction. Arguably, the class of square circle is also inconsistent, given the background knowledge about "square" and "circle" is provided formally, and their mutual exclusiveness can be derived from the definitions and axioms. Unicorns, Higgs bosons or ideal humans, on the other hand, can possibly exist, only the world is such that they (probably) do not exist. In particular, they are not self-contradictory. Even unsatisfiable classes may be classified further. Square circles and rain-and-not-rain-events may both be unsatisfiable, yet they have different intensions [20]. It is, however, out of the scope of this paper to address differences in unsatisfiable classes.

Use in scientific theories. The second axis is its use in scientific theories. An entity can be predicted by a scientific theory which predicted other entities that we believe to exist. For example, Higgs bosons are predicted by a scientific theory that is compatible with a large portion of collected empirical evidence. The problem is then to devise experiments and strategies to collect evidence for the existence of such predicted hypothetical entities.

Use in society, dependency on social construction. The third axis refers to idealizations as reference theories. These idealizations are often based on very similar entities that are observed, e.g., almost canonical humans.

6 Conclusions

Several biomedical ontologies have adopted the strategy to only include classes that either correspond to *universals* or are attributive collections, non-empty subsets of the extension of universals with an axiomatized inclusion criterion. The term "attributive collection" has been introduced recently to distinguish these classes in an ontological theory from the notion of "defined class" in a DL model [21]. In particular when ontologies are used for the representation of discourse, it becomes important to include classes that specify ideas about the terms used in scientific communication, independent of whether these terms refer to universals, attributive collections or not.

On the other hand, it may be relevant in some applications to include provenance information, including whether or not a class is believed to correspond to a universal or not. To provide this information, we started to develop a theory of terminological units and implemented this theory in OWL. Within this theory, terminological units, including both universals as well as classes that may not have instances in our world, are treated as instances of a

TerminologicalUnit class; and an explicit **hasInstance** relation is used to distinguish between universals (as terminological units that have instances) and empty classes (as terminological units that have no instances). The application of this theory will allow the development of ontologies that specify the meaning of terms within a vocabulary without sacrificing philosophical assumptions.

References

1. Michel Dumontier, Robert Hoehndorf (2010) Realism for scientific ontologies. Formal Ontology in Information Systems, Proceedings of the Sixth International Conference, FOIS 2010, 387-399.
2. National Center for Biomedical Ontology. Bioportal. bioportal. bioontology.org/
3. The Open Biological and Biomedical Ontologies. www.obofoundry.org/
4. Smith B et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007 Nov;25(11):1251-1255.
5. Basic Formal Ontology. ifomis.org /bfo.
6. DOLCE: a Descriptive Ontology for Linguistic and Cognitive Engineering, www.loa-cnr.it/DOLCE.html
7. BioTop. A top level ontology for the life sciences. purl.org/biotop/
8. GFO-Bio. A biomedical core ontology. onto.eva.mpg.de/gfo-bio.html
9. OBO Relation Ontology. obofoundry. org/ro/
10. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods of Information in Medicine, 37(4/5):394-403, 1998.
11. Schulz S, Johansson I. Continua in Biological Systems, 2007. The Monist 90 (4)
12. Smith, B. (2006). From concepts to clinical reality: An essay on the benchmarking of biomedical terminologies. Journal of Biomedical Informatics, 39, 288-298.
13. Baader F, Calvanese D, McGuinness DL, Nardi D, Patel-Schneider PF. The Description Logic Handbook Theory, Implementation, and Applications (2nd Edition). Cambridge, 2007.
14. OWL2 Web Ontology Language. W3C. www.w3.org/TR/owl2-overview/
15. Ceusters W, Smith B. A unified framework for biomedical terminologies and ontologies. Stud Health Technol Inform. 2010;160(Pt 2):1050-1054.
16. Ruttenberg A. (2009) Defined classes. obi-ontology.org/page/Defined_classes
17. Horridge M, Patel-Schneider P (2009) OWL 2 Web Ontology Language Manchester Syntax. www.w3.org/TR/2009/NOTE-owl2-manchester-syntax-20091027/
18. Glimm B, Horrocks I, and Motik B. Optimized Description Logic Reasoning via Core Blocking. In Proc. of the 2010 Description Logic Workshop (DL 2010), volume 573 of CEUR, 2010. ceur-ws.org/
19. Ceusters W, Smith B. Foundations for a realist ontology of mental disease. J Biomed Semantics. 2010 Dec 9;1(1):10.
20. Loebe F, Herre H. Formal semantics and ontologies: Towards an ontological account of formal semantics. Formal Ontology in Information Systems, 49-62, 2008.
21. Smith B, Ceusters W. Ontological realism: A methodology for coordinated evolution of scientific ontologies. Applied Ontology 5 (2010) 139-188.