A Realism-Based Approach to the Evolution of Biomedical Ontologies

Werner CEUSTERS [a] , Barry SMITH [a,b]

[a] Center of Excellence in Bioinformatics and Life Sciences, and National Center for Biomedical Ontology, University at Buffalo, NY, USA
[b] Institute for Formal Ontology and Medical Information Science, Saarbrücken, Germany, and Department of Philosophy, University at Buffalo, NY, USA

Abstract
We present a novel methodology for calculating the improvements obtained in successive versions of biomedical ontologies. The theory takes into account changes both in reality itself and in our understanding of this reality. The successful application of the theory rests on the willingness of ontology authors to document changes they make by following a number of simple rules. The theory provides a pathway by which ontology authoring can become a science rather than an art, following principles analogous to those that have fostered the growth of modern evidence-based medicine. Although in this paper we focus on ontologies, the methodology can be generalized to other sorts of terminology-based artifacts, including Electronic Patient Records.

1 Introduction

An ontology is commonly defined as '*a shared and agreed upon conceptualization of a domain*'. An ontology such as the UMLS Semantic Network correspondingly takes the form of a graph, whose nodes refer to *concepts.*[1] The combinations of nodes and edges in such a graph provide both *concept descriptions* and also, in the best case, *concept definitions.* Unfortunately, the documentation of such concept-based ontologies leaves unspecified what *concepts* actually are, or to what, if anything, they might correspond in reality.[2]

Of a different sort are those ontologies that are based on philosophical realism and require the nodes and edges in an ontology graph to correspond not to concepts but rather to entities in reality, for example to lesions or diseases or neoplasms on the side of the patient. Here the nodes in the graph refer to *universals* (such as *person*, *organ*, *liver*, *tumor*) which are shared in common by open-ended families of similar instances and which form the objects of scientific research. The edges in the graph correspond accordingly to relationships between universals, as expressed in assertions such as: *liver is_a organ*, *human liver part_of human being*, and so on. Such ontologies may then be used in association with inventories of those particulars that instantiate the corresponding universals, built out of assertions such as: patient #324 *instance_of person*.

Following a recently proposed terminology,[3] we use the term *portion of reality* (POR) to denote particulars, universals, and the simple and complex combinations thereof. Examples of ontologies conforming to these principles are Basic Formal Ontology[4] (BFO) and DOLCE,[5] and the same principles serve also as the basis for the Relation Ontology (RO) laid down by the Open Biomedical Ontologies consortium as part of its OBO Foundry initiative.[6] Examples of systems that are approaching satisfaction of these criteria are the most recent versions of the Foundational Model of Anatomy[7] (FMA) and the Gene Ontology[8] (GO).

2 Objectives

Interestingly, both conceptualist and realist ontologies share a common defect. When new versions of such ontologies are released, very little information is provided about the reasons for the changes made. As witnessed by two recent surveys,[9,10] efforts in the domain of ontology versioning and evolution have focused thus far on techniques for keeping track of which entries in an ontology appeared, disappeared, became fused or split in successive versions. Because the question is not raised as to *why* such changes are made, crucial distinctions are missed between the different kinds of changes in an ontology, reflecting for example:

(1) changes in the underlying reality (does the appearance or disappearance of an entry in a new version of an ontology relate to the appearance or disappearance of entities or of relationships among entities?);
(2) changes in our scientific understanding;
(3) reassessments of what is relevant for inclusion in an ontology;
(4) encoding mistakes introduced during ontology curation (for example through erroneous introduction of duplicate entries reflecting lack of attention to differences in spelling).

That such differences are overlooked is no surprise in the case of concept-based ontologies, where, because

entities in reality are thought of as playing at best a secondary role, the associated reasoning machinery takes care only of *internal* consistency. An example is the CONCORDIA model for managing divergence in concept-based terminologies,[11] which consists of a well-elaborated change model that is able to capture 27 different sorts of changes such as adding or merging concepts, or adding and deleting terms, but provides no facilities to log motivations for these changes along the lines proposed in what follows.

Ontologies based on realism are, we believe, capable of doing a better job. To demonstrate this, we develop a metric that allows us to measure the improvements obtained in successive versions of an ontology by drawing on reality as benchmark.

3 Material and methods

We base our metric on the distinction between three levels which have a role to play wherever ontologies are used as artifacts for annotation and automated reasoning in the field of biomedicine:

- Level 1: the *reality on the side of the patient*;
- Level 2: the *cognitive representations* of this reality embodied in observations and interpretations on the part of clinicians and others;
- Level 3: the *publicly accessible concretizations* of such cognitive representations in *representational artifacts* of various sorts, of which ontologies and terminologies are examples.

Different ontology authors maintain different positions concerning the correspondence between their representations and reality. Authors of realism-based ontologies maintain that their ontologies are intended to mirror reality; authors of concept-based ontologies maintain that their ontologies are intended to mirror cognitive representations on the part of domain experts. Our metric is based on the realist view, which means that it seeks to use objective reality as benchmark of correctness. This means in turn that it assumes that it is possible for humans to gain access to this reality, for example through the methods of evidence-based medicine. Since human cognition is fallible, both our cognitive representations and the representational artifacts based thereon may contain mistakes. But such mistakes can also be corrected, and it is above all this fact which makes possible a metric along the lines proposed.

4 Representations

In line with the theory of granular partitions,[12] we see complex representations as being composed in modular fashion of sub-representations built out of *representational units* that are assumed to correspond to PORs.

Some characteristics of the units in a representation created for clinical or research purposes are:

1) each such unit is assumed by the creators of the representation to be veridical, i.e. to conform to some relevant POR as conceived on the best current scientific understanding (which may, of course, rest on errors);
2) several units may correspond to the same POR by presenting different though still veridical views or perspectives, for instance at different levels of granularity (one thing may be described both as being brown and as reflecting light of a certain wavelength, or one event as an event of buying and of selling);
3) what is to be represented by the units in a representation depends on the purposes which the representation is designed to serve.

We concentrate in what follows on representational artifacts such as ontologies and terminologies, in which the representational units are terms from some natural or formal language, which are assumed to refer to universals or defined classes.[3]

5 The relevance and veridicality of expressions

Because ontologies, as conceived on realist terms, are artifacts created for some purpose (e.g. to serve as controlled vocabulary, or to provide domain knowledge to a software application), and because they are at the same time intended to mirror reality, and because reasoning with ontologies requires efficiency from a computational point of view, we argue that an optimal ontology should constitute a representation of *all and only those portions of reality that are relevant for its purpose*. Clearly, things may go wrong on the way to achieving this. First, ontology developers may be in error as to what is the case in their target domain, leading to *assertion errors*. Second, they may be in error as to what is objectively relevant to a given purpose, leading to *relevance errors*. Third, they may not successfully encode their underlying cognitive representations, so that particular representational units fail to point to the intended PORs because of *encoding errors*.

An ideal ontology, now, would be marked by none of these three types of errors. Each term in such an ontology would designate (1) a single POR, which is (2) relevant to the purposes of the ontology and such that (3) the authors of the ontology intended to use this term to designate this POR Moreover, (4) there would be no PORs objectively relevant to these purposes that are not referred to in the ontology.

Table 1 shows this ideal case and the possible types of departure therefrom, divided into two groups, labeled ‘P’ and ‘A’, denoting respectively the presence or absence of an expression (in or from an ontology). These cases reflect the different kinds of mismatch between what the ontology author believes

to exist (BE) or to be relevant (BRV) on the one hand, and matters of objective existence (OE) and objective relevance-to-purpose (ORV) on the other. The encoding of a belief can be either correct (R+) or incorrect, either (a) because the encoding does not refer (¬R) or (b) because it does refer, but to a POR other than the one which was intended (R-).

Table 1: Typology of expressions included in and excluded from an ontology in light of relevance and relation to external reality

	Reality		Under-standing		Encoding		G	E
	OE	**ORV**	**BE**	**BRV**	**Int.**	**Ref.**		
P+1	**Y**	Y	**Y**	Y	Y	R+	G1	0
A+1	**N**	–	**N**	–	–	–	G2	0
A+2	**Y**	N	**Y**	N	–	–	G3	0
P-1	**N**	–	**Y**	Y	Y	¬R	–	3
P-2	**N**	–	**Y**	Y	N	¬R	G4	4
P-3	**N**	–	**Y**	Y	N	R-	G5	5
P-4	**Y**	Y	**Y**	Y	N	¬R	G4	1
P-5	**Y**	Y	**Y**	Y	N	R-	G5	2
P-6	**Y**	N	**Y**	Y	Y	R+	G1	1
P-7	**Y**	N	**Y**	Y	N	¬R	G4	2
P-8	**Y**	N	**Y**	Y	N	R-	G5	3
A-1	**Y**	Y	**Y**	N	–	–	G3	1
A-2	**Y**	Y	**N**	–	–	–	G2	1
A-3	**N**	–	**Y**	N	–	–	G3	1
A-4	**Y**	N	**N**	–	–	–	G2	1

Legend: OE: objective existence; ORV: objective relevance; BE: belief in existence; BRV: belief in relevance; Int.: intended encoding; Ref.: manner in which the expression refers; G: typology which results when the factor of external reality is ignored. E: number of errors when measured against the benchmark of reality. P/A: presence/absence of term. (See text for details.)

Looking down the columns labeled OE and BE in Table 1, we can then distinguish four OE/BE value pairs, as follows:

- Y/Y: correct assertion of the existence of a POR;
- Y/N: lack of awareness of a POR, reflecting an assertion error;
- N/N: correct assertion that some putative POR does not exist (for example: 'there is no one-horned mammal');
- N/Y: the false belief that some putative POR exists (another kind of assertion error).

Note that these four pairs provide more information than would result from stating only that the assertions in question are true or false. As concerns the ORV and BRV columns in the table, these do not receive a value (cases marked '–') whenever either OE or BE, respectively, has the value N. An expression is included in an ontology only when BRV has the value Y. Wherever ORV has a different value than BRV, a relevancy error has been committed.

Out of the 15 alternative types of included and excluded expressions here distinguished, only 3 are desirable: P+1, which consists in the presence in an ontology of an expression that correctly refers to a relevant POR; and A+1 and A+2, which consist in the correct exclusion of an expression from an ontology, either because there is no POR to be referred to, or because this POR is not relevant to the ontology's purpose. A-3 and A-4 are borderline cases, in which errors made by ontology authors are without deleterious effect, either because something that is erroneously assumed to exist is deemed irrelevant, or because something that is truly irrelevant is over-looked. There are 9 different kinds of P cases, i.e. of cases which arise where an expression is present in an ontology. Of these, interestingly, only expressions of types P+1 and P-6 refer correctly to a corresponding POR: the former reflects our ideal case referred to above; the latter is marred by the incorrect inclusion of an expression which lacks relevance.

Note that our typology reflects what might initially appear to be an unacceptable idealization. For the type of an (included/excluded) expression depends upon two factors – of objective relevance-to-purpose and relation to objective reality – the assessment of which is something which could be correctly carried out only by someone able to adopt the god's eye perspective. As we shall see in section 6, however, the inclusion of these factors can in fact bring significant practical benefits when applied to actual cases. The key heuristic idea is to consider each new version of an ontology as incorporating, when measured in relation to its predecessor, some ingredient of the god's eye perspective.

Excluding objective reality and relevance from the typology – the practice defended, in effect, by proponents of the concept orientation in ontology development – would make the 15 types collapse into just 5, as indicated by column 'G' in Table 1. Defenders of the concept-based paradigm believe, in effect, that the terms in an ontology refer to concepts, rather than to entities in reality. This view has serious consequences. First, it collapses types P+1 and P-6 onto the single type G1. This makes the justifiable inclusion of a correctly encoded term in an ontology indistinguishable from the unjustifiable inclusion because of irrelevance to purpose.

Similarly, as indicated by the distribution of types G2 and G3 in Table 1, it confounds the two types of justifiable exclusion of a term from an ontology (A+1 and A+2) with various unjustifiable exclusions. Note that type P-1 has no counterpart within the G-typology because when an encoding captures what is intended, then it cannot have the value ¬R for its reference slot.

The last column of Table 1 shows the numbers of

mistakes committed when an expression of each given type is included in or left out of an ontology as measured against its corresponding baseline 'best case'. These baselines are P+1 for P-4, P-5, A-1 and A-2; A+1 for P-1, P-2, P-3 and A-3; and A+2 for all the others. These figures, as will be explained later, can be used to assess quality changes in successive versions of an ontology using reality as benchmark.

6 Ontology evolution

The minimal requirement for releasing an ontology on the realist paradigm is that its authors assume in good faith that all its constituent expressions are of the P+1 type. A stronger requirement would be that the authors advance the ontology as complete, i.e. as containing expressions designating all PORs deemed relevant to its purpose. Successive versions of an ontology should approximate ever more closely to this latter ideal, though in the biomedical domain it is of course unlikely that it will in fact be achieved.[13]

Documenting the changes made in an ontology by means of the typology described in Table 1 provides a way to quantify the improvements in its successive versions. This involves registering whether or not the changes are dictated by changes in (1) the underlying reality, (2) objective relevance of an included expression to the purposes of the ontology, (3) the ontology authors' understanding of each of these, and also by (4) the correction of encoding errors. If, for some purpose, we require only a sequence of 'snapshot ontologies' that mirror the entities existing in a given domain at successive points in time, then the disappearance of a POR requires merely the deletion of the corresponding expression: an expression of type P+1 would then give way, in the new version, either to one of type A+1 if the ontology authors are aware of the change, or to one of type P-1 if they are not. In the latter case the quality of the ontology decreases even though there is no change in the ontology itself. If, in contrast, the purpose requires representing changes which unfold over time, then the disappearance of one POR will require the addition of a new process term to designate the corresponding change. There would then, again in the best case, arise a new term of type P+1, and otherwise a new term of type P-4 or P-5.

In the following, we limit our analysis to the snapshot ontology case. Table 2 shows how an ontology might evolve under a simple scenario in which (1) a change in reality will not immediately lead to a change in the ontology authors' understanding thereof and (2) if an encoding change is introduced, e.g. by making some syntactic correction to an existing term, then this does not result in a term which wrongly refers. The described scenario is of course insufficiently refined for practical purposes; we use it merely to demonstrate a simple application of the kind of ontology benchmarking calculus that we are developing. It is 'simple' because it leads for each type of expression at each time *t* to only one possible type-assignment for the correction of that expression at time *t*+1.

For a number of expression types, certain transitions cannot occur; thus there can be no change in ORV if there is no POR to start with. These cases are indicated by empty cells in Table 2. The information displayed in the non-empty cells includes (1) the expression type that arises after application of the change indicated by the column header at time *t*+1, (2) the type of change in the expression, and, quality improvement realized, expressed as the difference in number of errors for an expression of

Table 2. The effect on the veridicality of terms in an ontology of different sorts of changes

t	***t*+1**				
	ΔOE	**ΔORV**	**ΔSE**	**ΔSRV**	**ΔInt**
P+1	P-1 nc -3	P-6 nc -1	A-2 D -1	A-1 D -1	P-4 C -1
A+1	A-4 nc -1		P-1 A -3		
A+2	P-1 A -3	A-1 nc -1	A-4 nc -1	P-6 A -1	
P-1	P-6 nc +2		**A+1** **D +3**	A-3 D +2	P-2 C -1
P-2	P-7 nc +2		**A+1** **D +4**	A-3 D +3	P-3 C -1
P-3	P-8 nc +2		**A+1** **D +5**	A-3 D +4	P-1 C +2
P-4	P-2 nc -3	P-7 nc -1	A-2 D 0	A-1 D 0	**P+1** **C +1**
P-5	P-3 nc -3	P-8 nc -1	A-2 D +1	A-1 D +1	**P+1** **C +2**
P-6	P-1 nc -2	**P+1** **nc +1**	A-4 D 0	**A+2** **D +1**	P-7 C -1
P-7	P-2 nc -2	P-4 nc +1	A-4 D +1	**A+2** **D +2**	P-6 C +1
P-8	P-3 nc -2	P-5 nc +1	A-4 D +2	**A+2** **D +3**	P-6 C +2
A-1	A-3 nc 0	**A+2** **nc +1**	A-2 nc 0	**P+1** **A +1**	
A-2	**A+1** **nc +1**	A-4 nc 0	**P+1** **A +1**		
A-3	**A+2** **nc +1**		**A+1** **nc +1**	P-1 A -2	
A-4	**A+1** **nc +1**	A-2 nc 0	**A+2** **nc +1**		

Legend: *Columns*: ΔOE: change in objective existence; ΔORV: change in objective relevance; ΔBE: change in belief about existence; ΔBRV: change in belief in relevance; ΔInt: change in encoding. *Cells*: nc: no change; A: addition of an expression; C: change in an expression; D: deletion of an expression. (See text for details.) Changes that lead to a correct result are printed in bold.

the given type at *t*+1 as compared to *t*. Note that a

quality improvement can be obtained even where the result of a change is still incorrect.

7 Towards an ontology benchmarking calculus

We argue that each time a new version of an ontology is released, or, better still, each time an individual expression is changed, added or deleted, the authors should document that change by indicating the sort of improvement they assume to have effected. Their assumption will always be that changes are towards the P+1, A+1, or A+2 cases. The purpose of the calculus is not however to demonstrate how good an individual version of an ontology is, but rather to measure how much it is believed to have been improved as compared to its predecessor. As an example, consider an expression that at stage *t* is assumed to be of type P+1 but is in fact of type P-7. At stage *t*+1 the ontology authors become aware of the unintended encoding and correct it. They then assume once again that this new term is of type P+1. This means that they have to believe at *t*+1 that the type at *t* was P-4 rather than P+1. The assumed gain in quality, according to table 1, would then be +1. At stage *t*+2, however, the authors change their minds, and assume that the expression refers to nothing at all, and they thus delete it from their ontology. They thus assume that at that stage it is of type A+1. This forces them to believe that at *t*+1 its type should have been P-1. Then however they must revise their belief with respect to the typology of the corresponding term at stage *t*, recognizing now that it should properly have been classified as being then of type P-2 (rather than P-4 as believed at stage *t*+1). The believed gain would then be +3 compared to *t*+1, and +4 compared to *t*. The registering of these kinds of transition chains is a process that can be easily automated. It requires only that ontology versioning software be supplemented with check-sheet technology allowing ontology authors, after appropriate training, to register for each term the appropriate values for BE, BRV, Int and Ref.

The resulting information can then be used to assess not only the quality of ontologies but also the skills of ontology authors through the tracking of the history of their revisions. A possible refinement of the method to assess the latter, would be to account also for deliberate omissions in the ontology when an author wishes to remain agnostic on certain issues. This would involve using a separate metric for author assessment such that, for instance, the number of errors committed or improvements realized will not be counted when an expression is added for which the author previously declared himself to be agnostic.

Our calculus also opens up interesting perspectives for documenting scientific discoveries. We believe that what we have called the 'objective relevance' of a representational unit in an ontology is something that is measurable, perhaps indirectly, when the ontology is used in implementations. Suppose that we are able to gauge improvements in the performance of an application after incorporation of a new version of an ontology, a version in which the existence of some POR is assumed. Then this raises the likelihood that the POR in question (or something very like it) truly exists. Equally interesting, from a philosophical perspective, are transition chains which lead from P-1 via P-6 to P+1: here something that was assumed to be the case was in reality not the case at the time the assumption was made, but, for completely independent reasons, became the case at some later time. This is what, in other circumstances, is called successful prediction.

References

[1] McCray AT. An upper-level ontology for the biomedical domain. Comp Funct Genom 2003;4:80-84.

[2] Smith B: Beyond concepts: Ontology as reality representation. Formal Ontology and Information Systems (FOIS) 2004. p. 73-84.

[3] Smith B, Ceusters W. Towards a coherent terminology for principle-based ontologies. Under review.

[4] Grenon P, Smith B, Goldberg L. Biodynamic ontology: applying BFO in the biomedical domain. Ontologies in Medicine, Amsterdam 2004, 20-38.

[5] Gangemi A, et al. Sweetening ontologies with DOLCE. Proc. EKAW (LNCS 2473) 2002, 166-81.

[6] Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, Mungall C, Neuhaus F, Rector AL, Rosse C. Relations in biomedical ontologies. Genome Biol, 2005;6(5):R46.

[7] Rosse, C. and Mejino, JLV. A reference ontology for bioinformatics: The Foundational Model of Anatomy. J Biomed Inform. 2003;36:478-500.

[8] Gene Ontology Consortium. The Gene Ontology (GO) project in 2006. Nucl Acids Res, 2006;34: D322–D326.

[9] Tagger B. A literature review for the problem of biological data versioning. July 2005. http:// www.cs.ucl.ac.uk/staff/btagger/LitReview.pdf.

[10] Haase P, Sure Y. State-of-the-art on ontology evolution. SEKT deliverable D3.1.1.b, 2004.

[11] Oliver DE, Shahar Y. Change management of shared and local versions of health-care terminologies. Methods Inf Med. 2000;39:278-90.

[12] Bittner T, Smith B. A theory of granular partitions. In Foundations of Geographic Information Science, M Duckham, MF Goodchild and MF Worboys (eds.), London: Taylor & Francis, 2003, 117–151

[13] Smith B. From concepts to clinical reality: an essay on the benchmarking of biomedical terminologies, J Biomed Inform. 2006;39(3):288-298.