Ceusters W. Towards A Realism-Based Metric for Quality Assurance in Ontology Matching. In: Bennett B, Fellbaum C. (eds.) Formal Ontology in Information Systems, IOS Press, Amsterdam, 2006;:321-332. Proceedings of FOIS-2006, Baltimore, Maryland, November 9-11, 2006.

# Towards A Realism-Based Metric for Quality Assurance in Ontology Matching

Werner CEUSTERS [1]
*Center of Excellence in Bioinformatics and Life Sciences, University at Buffalo, NY, USA*

**Abstract**: Ontology matching is commonly defined as a matter of dealing with semantic correspondences between terms in ontologies and thus refers to more specific activities such as mapping or aligning, possibly with ontology merging in mind. However, it has been pointed out that there still prevails no common understanding of what such 'semantic correspondences' are supposed to be, and that in consequence "human experts do not agree on how ontologies should be merged, and we do not yet have a good enough metric for comparing ontologies." In what follows we define such a metric, which is designed to allow assessment of the degree to which the integration of two ontologies yields improvements over either of the input ontologies. We start out from the thesis that if two or more ontologies are to be considered for matching, then, however much they may reflect distinct views of reality on the part of their authors, the portions of reality to which they refer must be such as to overlap. Our approach takes account of the fact that both authors and users of ontologies may make mistakes (the former in their interpretation of reality and in the formulation of their views, the latter in misinterpreting the former's intentions).To do justice to such factors, we need to draw a distinction between three levels of: (1) reality; (2) cognitive representations; and (3) publicly accessible concretizations of these representations. We can then define 'semantic correspondence' not, as is usual, in terms of (horizontal) relations of 'association' or 'synonymy' between the terms within the ontologies to be matched, but rather in terms of the (vertical) relation of reference: terms correspond semantically if they refer to the same entities in reality. One conclusion of our argument is that, when ontology matching has been used as the first step towards ontology merging, then the merged ontology can contain inconsistencies only if there are already inconsistencies in at least one of the source ontologies.

**Keywords**: ontology matching, quality assurance, realism

## 1. Introduction

An ontology is commonly defined as '*a shared and agreed upon conceptualization of a domain*'. Often, an ontology so conceived takes the form of a graph, whose nodes are seen as referring to what are called 'concepts.' The combinations of nodes and edges in such a graph provide *concept descriptions*, and sometimes, in the best case, can be used to generate *concept definitions*. Unfortunately, the documentation of such concept-

[1] Corresponding author: Werner Ceusters, Center of Excellence in Bioinformatics and Life Sciences, Ontology Research Group, 701, Ellicott Street, Buffalo NY, 14203, USA. Email: ceusters@buffalo.edu.

Ceusters W. Towards A Realism-Based Metric for Quality Assurance in Ontology Matching. In: Bennett B, Fellbaum C. (eds.) Formal Ontology in Information Systems, IOS Press, Amsterdam, 2006;:321-332. Proceedings of FOIS-2006, Baltimore, Maryland, November 9-11, 2006.

based ontologies leaves unspecified what *concepts* actually are, and to what, if anything, they might correspond in reality [1]. We have argued elsewhere that this lack of specification leaves so much free play for ontology developers that a series of mismatches are created both within and between ontologies, with consequences which can be compared to the consequences of creating an international railway system in which it is left unspecified what gauge should be used by the separate national systems. The many different kinds of mistakes detected in existing terminologies and ontologies have a variety of sources (including, increasingly, the unjustified belief in the quality assurance capacities of Description Logic-based languages [2]). We believe, however, that the uncertainty as to what is meant by 'concept' is still the main reason why such mistakes arise [3, 4, 5].

Increasingly, however, and especially in the domain of biomedicine, ontologies are being built that are not based on 'concepts' but on philosophical realism. Here the nodes and edges in the ontology graph are required to correspond not to concepts but rather to entities in reality, for example to molecules, or tumors, or diseases on the side of patients. More precisely, they are required to refer to *universals* (such as person, organ, liver, tumor). It is universals which form the objects of scientific research. Where we acquire knowledge about 'concepts' by examining ideas or thoughts or meanings, universals are directly accessible through their instances in reality, for example as they appear in the lab or clinic: they are entities that are multiply located in space and time through their particular instances and they are identified by discovering that particular families of instances share in common certain corresponding intrinsic features and dispositions.

On the realist paradigm, the nodes in an ontology graph correspond to universals and the edges in the graph to relations between such universals, as expressed in assertions such as: *liver is_a organ, liver part_of mammal*, and so on – relations which are themselves defined in turn in terms of further relations obtaining among the underlying instances. Thus when we say that universal *A* stands in the *part_of* relation to universal *B*, then what we mean is that every instance of *A* stands to some instance of *B* in that instance-level parthood relation which is defined through the standard axioms of mereology [6].

The use of relations so defined allows us to ensure that ontologies have a direct relation to instances in reality, so that they may be used in association with realism-based *inventories* of such instances, built out of assertions such as: patient #324 *instance_of person*, meningitis #4612 *instance_of disease_of_nervous_system* in the construction of an electronic health record system that is designed to support instance-based reasoning. Instances in reality can hereby provide a benchmark of correctness for the assertions in an ontology [7], and instance information can also be used to help solve some of the problems of ontology merging to be addressed in what follows.

## 2. Terminological Conventions

Following a recently proposed terminology [8], we will use the expression '*portion of reality*' to denote instances, universals, and the simple and complex combinations they form when combined through relations of the mentioned sorts. By 'instance-level portion of reality' we mean: individuals and collections thereof: you, your digestive system, your family, your favorite hospital.

The terminology used by practitioners of ontology matching is not consistent, but

Ceusters W. Towards A Realism-Based Metric for Quality Assurance in Ontology Matching. In: Bennett B, Fellbaum C. (eds.) Formal Ontology in Information Systems, IOS Press, Amsterdam, 2006;:321-332. Proceedings of FOIS-2006, Baltimore, Maryland, November 9-11, 2006.

we believe that it is best conceived as an operation on pairs of ontologies yielding an input to support further operations such as *merging*, *fusion*, or *integration* of ontologies designed to yield new single artifacts out of the ontologies with which we begin. Matching consists in dealing with what are called 'semantic correspondences' between the representational units (the single terms) of the individual ontologies. It involves activities such as '*ontology mapping*', which is mostly concerned with the *representation* of correspondences between ontologies, and '*ontology alignment*', which is concerned with the (semi-)automatic discovery of such correspondences [9].

This task of identification is still normally addressed from the concept-based point of view; thus ontology representational units or terms are 'matched' when, as it is said, they express, refer to, or represent the *same concepts*. Ehrig and Sure for example describe ontology mapping as follows: "*given two ontologies A and B, mapping one ontology with another means that for each concept (node) in ontology A, we try to find a corresponding concept (node), which has the same or similar semantics, in ontology B and vice verse.*" [10] To say that two concepts have similar semantics, on this account, means roughly that they occupy similar places in the associated graphs (called 'concept lattices').

An analogous approach is advanced by Kalfoglou and Schorlemmer, who define an ontology as "*a pair O* = (*S*, *A*), *where S is the (ontological) signature – the vocabulary – and A is a set of (ontological) axioms – specifying the intended interpretation of the vocabulary in some domain of discourse*". The ontological signature itself they see as "*a hierarchy of concept symbols together with a set of relations symbols* [sic] *whose arguments are defined over the concepts of the concept hierarchy*". Ontology mapping is then: "*the task of relating the vocabulary of two ontologies in such a way that the mathematical structure of ontological signatures and their intended interpretations, as specified by the ontological axioms, are respected*" [11].

Bouquet *et al*. define ontology mapping in similar vein as "*a formal expression that states the semantic relation between two entities belonging to different ontologies*", and they continue: "*Simple examples are: concept c1 in ontology O1 is equivalent to concept c2 in ontology O2; concept c1 in ontology O1 is similar to concept c2 in ontology O2; individual i1 in ontology O1 is the same as individual i2 in ontology O2*", and so on [12].

Based on this same point of view, Kotis *et al.* define ontology merging as follows: "*Given two source ontologies O1 and O2* [we can] *find an alignment between them by mapping them to an intermediate ontology, and then, get the minimal union of their (translated) vocabularies and axioms with respect to their alignment.*" [13].

The problems with all of the above are however clear: ontology matching is defined in terms of the correspondence (equivalence, sameness, similarity) of concepts. But how, precisely, do we gain access to these concepts in order to determine whether they stand in a relation of correspondence (presupposing that we have already solved the prior problem of working out what 'concept' means in any given context). One option is via definitions, but then these definitions themselves, as they are supplied by the different ontologies to be matched, will likely employ different terms (or 'concepts'), so that the problem of matching has merely been shifted to another place. The Ehrig and Sure suggestion of establishing correspondence by looking at the positions of given concepts in their surrounding concept lattices is subject to a similar difficulty. For how, unless we have already matched (some) single concepts, can we compare 'places' in distinct lattices (which as experience shows will likely still – if

Ceusters W. Towards A Realism-Based Metric for Quality Assurance in Ontology Matching. In: Bennett B, Fellbaum C. (eds.) Formal Ontology in Information Systems, IOS Press, Amsterdam, 2006;:321-332. Proceedings of FOIS-2006, Baltimore, Maryland, November 9-11, 2006.

they are lattices at all – have very different mathematical forms). This leaves only a series of more or less statistically-based algorithms involving lexical term-matching, the results of whose application have thus far proved uneven, to say the least [14].

When Euzénat *et al.* carried out a survey of ontology alignment methods, they did indeed find that the majority are based on analyzing either the vocabulary used to label concepts or the structure in terms of which the latter are organized [15]. But more interesting was their statement that "*there is no common understanding of what to align, how to provide the results and what is important*", a conclusion which echoes that of Noy and Musen, according to whom "*human experts do not agree on how ontologies should be merged, and we do not yet have a good enough metric for comparing ontologies*." [16]

## 3. Objectives

Our goal is to define such a metric, based on what the expressions in ontologies are intended to refer to in reality. That is, we hold that ontology matching is possible only if we view expressions in terms of that in reality to which they are believed to refer. In the case of realism-based ontologies such as the Foundational Model of Anatomy [17] the resultant methodology can be applied directly. However it can also be extended quite easily to the analysis of concept-based ontologies, since the expressions of the latter can in many cases be viewed from the realist perspective.

Central to our approach is the claim that expressions (terms in natural language or expressions constructed by means of a formal language) from two or more ontologies can be considered from the point of view of matching only if they are built out of representational units which refer to instance-level portions of reality which overlap. The referents of two expressions are said to overlap if either they or the referents of expressions from out of which they are composed are such that the portions of reality referred to by these expressions share parts. Thus the coverage of an ontology of anatomy will likely overlap with that of an ontology of disease since many diseases are associated with specific bodily locations, as is marked by the use of expressions like 'lung cancer' or 'spinal fracture'. Note that the ontologies in these domains do not overlap *because* they contain expressions of the given sorts. Rather, such expressions are included, and associated relations posited, *because of the relationships that obtain in reality* between the corresponding entities (between spinal fractures and spines, between lung cancers and lungs).

An adequate metric for comparing ontologies and the quality of the matching between them must be able to deal with a variety of problems by which ontology matching endeavors thus far have been affected:

(a) different ontology authors may have different though still veridical views on the same reality,

(b) ontology authors may make mistakes, either when interpreting reality or when formulating their interpretations in their chosen ontology language,

(c) an agent – whether human or machine – who is charged with carrying out the matching (and who from this point forward we will be calling *the assessor*) can never be sure to what the expressions in an ontology actually refer (for this, he would need to be able to adopt a God's eye perspective),

(d) if two ontologies are developed at different times, reality itself may have changed in the intervening period.

Ceusters W. Towards A Realism-Based Metric for Quality Assurance in Ontology Matching. In: Bennett B, Fellbaum C. (eds.) Formal Ontology in Information Systems, IOS Press, Amsterdam, 2006;:321-332. Proceedings of FOIS-2006, Baltimore, Maryland, November 9-11, 2006.

Each of these factors represents a dimension of unknowns in the ontology matching exercise, and we provide the resources to solve for each of these unknowns in the discussion which follows below.

One (expensive) way to create a metric for the quality of ontology matching would be to have experts manually prepare for each given matching problem a gold standard 'ideal' solution, to which matching efforts to be evaluated could be compared [18]. Our solution is of a different sort. It relies on the idea that one can measure what has been gained – which means: that we can count the improvements that have been effected – when the results of a given matching are compared to the ontologies as they had existed earlier. With this type of metric, we are able to assess whether the integration of two ontologies is an *improvement* over either of the input ontologies

## 4. Material and methods

We base our method on the same distinction between three levels that we introduced as part of the methodology for the measurement of quality improvements in single ontologies advanced in [19]. These levels are:

- Level 1: *reality*, consisting of both instances and universals, as well as the various relations that obtain between them;
- Level 2: the *cognitive representations* of this reality embodied in observations and interpretations;
- Level 3: the *publicly accessible concretizations* of such cognitive representations in *representational artifacts* of various sorts, of which ontologies are examples.

For adepts of the concept-based approach, ontologies are representational artifacts which are intended to mirror the cognitive representations shared by domain experts. From the realist perspective, ontologies (independently of the paradigm on the basis of which they were constructed) are to be interpreted as if their expressions are intended to refer to entities in reality. This then gives us the possibility of using objective reality as a benchmark for the correctness of ontologies and of ontology matching efforts.

We have talked thus far of the 'expressions' in an ontology. In line with the theory of granular partitions [20], however, we prefer to talk more precisely of ontologies as being composed in modular fashion out of sub-representations which are built ultimately out of minimal (syntactically non-decomposable) *representational units* (including alphanumeric identifiers to uniquely identify universals), each of which is assumed by its author to correspond to some portion of reality (POR).

We are interested primarily in ontologies created for clinical or research purposes. The representational units of such ontologies are marked by the following characteristics:

1) each such unit is assumed by the authors, on the basis of their best current understanding of reality (which may, of course, rest on errors), to be veridical, i.e. to refer to some relevant POR;
2) several units may correspond to the same POR by representing distinct though still veridical views or perspectives on this POR, for instance at different levels of granularity (one thing may be described both as being brown and as reflecting light of a certain wavelength; one event as being both an act of buying and an act of selling).

Ceusters W. Towards A Realism-Based Metric for Quality Assurance in Ontology Matching. In: Bennett B, Fellbaum C. (eds.) Formal Ontology in Information Systems, IOS Press, Amsterdam, 2006;:321-332. Proceedings of FOIS-2006, Baltimore, Maryland, November 9-11, 2006.

In addition we take it that ontologies in general are characterized by the fact that the choice of what is to be represented by their representational units depends on the purposes which a given ontology is designed to serve (for example: to provide domain knowledge to a software application). 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 to the purpose for which the ontology was built.

Clearly, things may go wrong on the way to achieving such an optimal representation. 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 be in error because their ontologies do not successfully encode the underlying cognitive representations, so that particular representational units fail to point to the intended PORs because of errors of syntax, leading to *encoding errors*.

The ideal (optimal) ontology, now, would be marked by containing no errors of the three just-mentioned types. This means that each representational unit in such an ontology would designate (1) a single POR, that 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. In addition, (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, which we have labeled ‘P’ and ‘A’, respectively, to denote 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 presents a second-order view of how expressions and reality are related together. Thus it allows us to assert for example that certain expressions ought to be in an ontology because there are relevant PORs that need to be referred to.

To see how the table works, consider the second and fourth columns in its main body. We can there 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).

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 value different from that of BRV, a relevancy error has been committed.

Out of the 15 alternative types of included and excluded expressions, only 3 are desirable: P+1, A+1, and A+2. P+1 consists in the presence in an ontology of an expression that correctly refers to a relevant POR; A+1 and A+2 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

**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** | | **E** |
|---|---|---|---|---|---|---|---|
| | **OE** | **ORV** | **BE** | **BRV** | **Int.** | **Ref.** | |
| P+1 | Y | Y | Y | Y | Y | R+ | 0 |
| A+1 | N | – | N | – | – | – | 0 |
| A+2 | Y | N | Y | N | – | – | 0 |
| P-1 | N | – | Y | Y | Y | ¬R | 3 |
| P-2 | N | – | Y | Y | N | ¬R | 4 |
| P-3 | N | – | Y | Y | N | R– | 5 |
| P-4 | Y | Y | Y | Y | N | ¬R | 1 |
| P-5 | Y | Y | Y | Y | N | R– | 2 |
| P-6 | Y | N | Y | Y | Y | R+ | 1 |
| P-7 | Y | N | Y | Y | N | ¬R | 2 |
| P-8 | Y | N | Y | Y | N | R– | 3 |
| A-1 | Y | Y | Y | N | – | – | 1 |
| A-2 | Y | Y | N | – | – | – | 1 |
| A-3 | N | – | Y | N | – | – | 1 |
| A-4 | Y | N | N | – | – | – | 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; **E:** number of errors when measured against the benchmark of reality. **P/A:** presence/absence of term. (See text for details.)

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 overlooked. There are 9 different types 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 an inclusion that is incorrect because the included expression lacks relevance.

The last column of Table 1 shows for each type the numbers of mistakes committed with respect to the 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.

## 5. Ontology matching and merging

The minimal requirement for releasing an ontology, on the realist paradigm, is that the 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 the PORs deemed relevant to its purpose. In reality, of course, any single ontology will contain expressions of the various P-*n* types. It will typically also lack expressions for PORs which are relevant to its purposes. This is certainly the case when ontologies are candidates for merging, since this presupposes at least an incompleteness on the side of both ontologies.

Ceusters W. Towards A Realism-Based Metric for Quality Assurance in Ontology Matching. In: Bennett B, Fellbaum C. (eds.) Formal Ontology in Information Systems, IOS Press, Amsterdam, 2006;:321-332. Proceedings of FOIS-2006, Baltimore, Maryland, November 9-11, 2006.

The matching of expressions in ontologies $O^1$ and $O^2$ now comes down in our view to assessing which expressions $EO^1_i$ present in the first ontology stand in a relationship of reference to the same PORs as expressions $EO^2_j$ in the second.

If two distinct expressions are judged to refer to the same POR, then the expressions can be counted as synonymous (a case which is taken to include the relation between expressions drawn from different languages which are translations of each other).

Trivially, something has thereby been gained, and this gain can be quantified. For the purposes of this communication, however, we will not take this trivial case further into account. Our attentions are focused rather on the more challenging sort of case, where for an expression in ontology $O^1$ no co-designating expression is found in ontology $O^2$. Our methodology then requires that the assessor needs to document what, in his mind, is the reason for this mismatch using the typology described in Table 1. The latter enables him to quantify the seriousness of the mismatch in terms of the number of errors associated which each deviation from the baseline (P+1, A+1 or A+2) cases.

A P-8 configuration for example deviates from its baseline (A+2) in three respects: (1) the POR is falsely believed to be relevant, (2) the expression does not encode what its author intends it to encode, and (3) the expression refers to a POR different from that to which it is intended to refer.

The possibilities for co-reference are restricted by the following principles (which can be incorporated into the software support for the assessor's work):

- there is only one reality to which expressions in $O^1$ and $O^2$ may refer;
- the relevance of a POR is to be assessed in light of the purposes for which the resultant mapped or merged ontology is being created, not in terms of the original purposes of the individual ontologies;
- everything that exists or has existed can be referred to (where necessary by using appropriate temporal indices).

## 6. Dealing with differences in coverage

There are various ways an assessor can deal with differences in the ontologies he is called upon to match. A first option would be to assume that there are no mistakes in the source ontologies, and that any difference is to be accounted for in terms of their different *purposes.* (We use 'pO' in what follows to refer to the purpose of an ontology O.) On this option, if a POR is referred to in $O^1$ but not in $O^2$, then this does not mean that the authors of $O^2$ did not believe in the existence of that POR, but rather that they did not consider the POR to be relevant for $pO^2$. Another option would be to assume (perhaps because one has evidence for the thesis) that these purposes are identical, so that any difference is to be explained on the basis of a false assumption as to the relevance of that POR underlying one or other of the source ontologies. The assessor must then apply some strategy to resolve conflicting information in ways which will allow him to identify the false belief. Here information reflecting quality improvements over successive versions of the ontologies under scrutiny (compiled along the lines developed in [19]) can provide important clues.

Table 2 demonstrates the application of our metric to a case in which two ontologies are to be matched in which the purposes of the input ontologies themselves

Ceusters W. Towards A Realism-Based Metric for Quality Assurance in Ontology Matching. In: Bennett B, Fellbaum C. (eds.) Formal Ontology in Information Systems, IOS Press, Amsterdam, 2006;:321-332. Proceedings of FOIS-2006, Baltimore, Maryland, November 9-11, 2006.

and of the sought for merged ontology do not need to be the same. Rows 1 and 2 depict situations where expressions in $O^1$ and $O^2$ have been mapped to the same POR, and rows 3 and 4 situations in which a POR is referred to in only one ontology. Rows 5 and 6 reflect situations where both source ontologies lack reference to a POR that is either relevant (C5) or irrelevant (C6) for the purposes of the matching.

Table 2 reflects a situation in which the assessor, during the matching process, is assumed to have some procedure which enables him to assess PORs as relevant or not relevant for the purposes of the ontology which is to result from matching or merging (since it will typically not be the case that everything relevant to the source ontologies will also be relevant to the merged result). The assessment of relevance will then yield one or other of: correct presence (P+1), justified absence because of lack of relevance (A+2), or unjustified absence because of missed relevance (A-1) (shown in column (9) of Table 2).

The 'objective relevance' of a POR, and thus of a corresponding expression in an ontology, is something that becomes at least indirectly measurable whenever the ontology is used to solve the problems for which it was designed. To this end we need to measure improvements in the performance of applications after they have started to use the merged ontology. If the postulation of some POR as relevant leads to improvements in such performance, then it is likely that that the POR in question (or something very like it) is indeed relevant.

But even where we have no knowledge as to the objective relevance of a given POR, it is still possible to measure the quality of the source ontologies relative to the matching result. Consider for example row 3 in Table 2: if it is believed that for purpose $pO^m$ of a merged ontology $O^m$ a certain POR is relevant, and that for purpose $pO^2$ it is not, then one must also believe that if $O^2$ would be used for $pO^m$ then there is an unjustified absence of an expression, namely one characterized as being of type A-1.

To have a quantitative assessment of the relative quality of the ontologies with respect to $pO^m$, it suffices to use the number of errors that are involved in an expression of a certain type as indicated in the column labeled 'E' in Table 1 and to use them in appropriate formulas for dealing with errors. Under the assumptions used for Table 2, a simple tally of the percentage of error-free expressions relative to the total number of expressions may suffice, since the number of errors for expressions of types P-6 and A-1 equals 1 in both cases. For more complex cases, in which the number of errors per expression type might be higher than 1 (in the worst case, if all expressions would be of type P-3, there would be 5 times more errors than expressions), error percentage formulas that include adjustments for the number of expressions would yield more adequate results. These calculations may seem at first difficult to master; however, they soon prove themselves to have a high degree of intuitiveness, as is seen in the fact that they can easily by carried out by means of simple software embedded in ontology matching tools. They do however require some additional effort on the part of those involved in ontology matching, in that they are required to assess (through checklists or similar technology) the relevance of each expression in the input ontologies in light of the purposes involved in the intended merger.

## 7. Reality as benchmark

When two or more source ontologies are mapped or merged, then it may happen that inconsistencies are discovered. In [21] a distinction is made between inconsistency and

Ceusters W. Towards A Realism-Based Metric for Quality Assurance in Ontology Matching. In: Bennett B, Fellbaum C. (eds.) Formal Ontology in Information Systems, IOS Press, Amsterdam, 2006;:321-332. Proceedings of FOIS-2006, Baltimore, Maryland, November 9-11, 2006.

**Table 2: Possible combinations of the believed relevance of a POR in source ontologies ($O^1$ and $O^2$) and resulting merger ($O^m$)**

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|
| | $O^1$ | | | $O^2$ | | | $O^m$ | |
| | BRV | ET | | BRV | ET | | BRV | ET |
| | $pO^1$ | $pO^1$ | $pO^m$ | $pO^2$ | $pO^2$ | $pO^m$ | $pO^m$ | $pO^m$ |
| 1 | Y | P+1 | P+1 | Y | P+1 | P+1 | Y | P+1 |
| 2 | Y | P+1 | P-6 | Y | P+1 | P-6 | N | A+2 |
| 3 | Y | P+1 | P+1 | N | A+2 | A-1 | Y | P+1 |
| 4 | Y | P+1 | P-6 | N | A+2 | A+2 | N | A+2 |
| 5 | N | A+2 | A-1 | N | A+2 | A-1 | Y | A-1 |
| 6 | N | A+2 | A+2 | N | A+2 | A+2 | N | A+2 |

Legend: **BRV:** believed relevance; **ET:** expression type (see Table 1); columns (2), (5) and (8): relevance of the POR for the purposes of $O^1$, $O^2$ and $O^m$ respectively; columns (3), (6) and (9): expression types according to $pO^1$, $pO^2$ and $pO^m$ respectively; columns (4) and (7): expression types if $O^1$ or $O^2$ would be used without any modification in $pO^m$.

incoherence: "*An ontology will be called inconsistent iff there is no interpretation satisfying all the DL axioms in the ontology; it will be called incoherent iff it does not satisfy certain predefined constraints or invariants related to efficient ontology design*." Thus it is claimed that the ontology with the axioms (1) birds fly, (2) penguins are birds, and (3) penguins don't fly, is incoherent. Only when an assertion is added to the effect that a particular individual is an instance of *penguin*, we are told, does the ontology become inconsistent. Yet, so the authors continue, an ontology of exactly the same structure consisting of the axioms: (1) horses don't have horns, (2) unicorns are horses, and (3) unicorns have horns, is also incoherent, but will become never inconsistent because "*To the authors' knowledge, there are no unicorns*". This difference (perhaps) makes sense from a concept-based view; not however from a realist perspective. For when reality is taken as benchmark, then it becomes clear that *both* ontologies contained mistakes from the very beginning. Thus, if ontology $O^1$, with expressions (1) birds fly and (2) penguins are birds, becomes merged with ontology $O^2$, which contains (3) penguins don't fly, then it is not such that only the merged ontology *became* wrong; $O^1$ was *already* wrong!

Mistakes of this kind do not *arise* because of merging; rather, they are *discovered* thereby. To find out which of the three axioms is (or are) the source of error is a matter not of applying logic, but rather of looking carefully at reality in light of what the axioms assert. It might be that the ontology authors' understanding of reality was erroneous from the start, so that an assertion error was made; or it might be that the intended representational unit was erroneously encoded. Indeed, it might also be that reality has changed between the times that $O^1$ and $O^2$ were published, perhaps because penguins lost the ability to fly. In each such case, the complete range of possible types of mistakes as shown in Table 1 must be taken into account.

Of course, Table 1 alone is not able to inform an assessor which of the expressions to be mapped or merged are wrong: to find out whether penguins fly or whether they are birds is a matter of scientific discovery. It does however inform an assessor of the improvement in quality which will result from a given merger according to the position that is taken as to the reasons for mistakes which this merger corrects. In [19], we argued that while revising ontologies, authors should keep track of the reasons for any changes made by registering whether or not the changes are believed to be dictated by changes in (1) the underlying reality, (2) the objective relevance of an included expression to the purposes of the ontology, (3) the ontology authors' understanding of each of these, and (4) the correction of encoding errors. Of course, authors will always assume that changes are *towards* the P+1, A+1, or A+2 cases. But this does not prevent the assessor from measuring how much the ontology is believed to have been improved as compared to its predecessor (and it does not prevent him, either from evaluating the skills of ontology authors by tracking the history of their earlier revisions).

## 8. Conclusion

We presented a novel methodology for assessing the quality of ontologies that are mapped upon each other, or that result from merging two or more source ontologies. We concluded that differences between the ontologies should be resolved by resorting to (1) the ontology authors' beliefs in what is the case in the underlying reality, (2) their belief in the relevance of an included expression to the purposes of the ontology, and (3) the possible presence of encoding errors. Differences between these beliefs and what is the case in reality are quantifiable and can be used to assess the adequateness of both the original ontologies and the resultant matching or merging. The methodology for quality measurement thereby provides a pathway by which ontology matching and merging can be transformed from art into science.

## References

[1] B. Smith, Beyond concepts: Ontology as reality representation. Varzi A, Vieu L (eds.), Proc FOIS 2004. International Conference on Formal Onto-logy and Information Systems, Turin, p. 73-84.
[2] W. Ceusters, B. Smith, Ontology and Medical Terminology: why Descriptions Logics are not enough. Proceedings of the conference Towards an Electronic Patient Record (TEPR 2003), San Antonio, 10-14 May 2003 (electronic publication 5pp)
[3] W. Ceusters, B. Smith, A. Kumar, C. Dhaen, Mistakes in medical ontologies: where do they come from and how can they be detected? in Pisanelli DM (ed) Ontologies in Medicine. Proceedings of the Workshop on Medical Ontologies, Rome October 2003, IOS Press, Studies in Health Technology and Informatics, 2004;102: 145-63.
[4] W. Ceusters, B. Smith, A. Kumar, C. Dhaen, Ontology-based error detection in SNOMED-CT. Proceedings of MEDINFO 2004;:482-6.
[5] W. Ceusters and B. Smith, A Terminological and Ontological Analysis of the NCI Thesaurus. Methods of Information in Medicine 2005; 44: 498-507
[6] B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A. Rector and C. Rosse, Relations in biomedical ontologies, Genome Biology (2005), 6 (5), R46.
[7] B. Smith, From Concepts to Clinical Reality: An Essay on the Benchmarking of Biomedical Terminologies, J Biomed Inform, 2006 Jun;39(3):288-98.
[8] B. Smith, W. Ceusters, Towards a Coherent Terminology for Principle-Based Ontologies. (under review).
[9] J. de Bruijn, M. Ehrig, C. Feier, F. Martín-Recuerda, F. Scharffe, M. Weiten, Ontology mediation, merging and aligning. In Davies J, Studer R, Warren P (eds), Semantic Web Technologies: Trends and Research in Ontology-based Systems, Wiley, UK, 2006.
[10] M. Ehrig M and Y. Sure, Ontology mapping - an integrated approach. In Proceedings of the First European Semantic Web Symposium, ESWS 2004, volume 3053 of Lecture Notes in Computer Science, pages 76–91, Heraklion, Greece, May 2004. Springer Verlag.
[11] Y. Kalfoglou and M. Schorlemmer, Ontology mapping: the state of the art. Knowl. Eng. Rev., 18(1):1--31, 2003.

[12] P. Bouquet *et al*. KnowledgeWeb deliverable D2.2.1. Specification of a common framework for characterizing alignment. http://www.inrialpes.fr/exmo/cooperation/kweb/heterogeneity/deli/kweb-221.pdf

[13] K; Kotis, G. Vouros, K. Stergiou, Towards Automatic Merging of Domain Ontologies: The HCONE-merge approach. Elsevier's Journal of Web Semantics (JWS), vol. 4:1, pp. 60-79, January 2006.

[14] S. Staab and A. Maedche. Measuring similarity between ontologies. Lecture notes in artificial intelligence, 2473:251--263, 2002.

[15] J. Euzénat *et al*. KnowledgeWeb Deliverable D2.2.3: State of the art on ontology alignment. V1.2, August 2004. http://www.starlab.vub.ac.be/research/projects/knowledgeweb/kweb-223.pdf

[16] N.F. Noy and M.A. Musen, Evaluating Ontology-Mapping Tools: Requirements and Experience. Workshop on Evaluation of Ontology Tools at EKAW'02 (EON2002). 2002, p1-14.

[17] C. Rosse and J.L.V. Mejino, A reference ontology for bioinformatics: the Foundational Model of Anatomy. J Biomed Inform, 36: 478-500, 2003.

[18] M. Ehrig and J. Euzenat, Relaxed Precision and Recall for Ontology Matching, in: Proc. K-Cap 2005 workshop on Integrating ontology, Banff (CA), p. 25-32, 2005.

[19] W. Ceusters and B. Smith, A Realism-Based Approach to the Evolution of Biomedical Ontologies. Forthcoming in Proceedings of AMIA 2006, Washington DC, November 11-15, 2006. http://ontology.buffalo.edu/bfo/Versioning.pdf

[20] T. Bittner and B. Smith, 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

[21] G. Flouris and D. Plexousakis, Handling Ontology Change: Survey and Proposal for a Future Research Direction, Technical Report 362, ICS-FORTH, Heraklion, Crete, Greece, September 2005.