Negative Findings in Electronic Health Records and Biomedical Ontologies: A Realist Approach

Werner CEUSTERS[a] , Peter ELKIN[b] and Barry SMITH [a,c]

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

[b] *Department of Medicine, Mayo Foundation, Rochester, MN, USA*

[c] *Department of Philosophy and National Center for Ontological Research, University at Buffalo, NY, USA*

Corresponding author:

Werner Ceusters, MD

Center of Excellence in Bioinformatics and Life Sciences

701 Ellicott Street

Suite B2-132

Buffalo, NY 14203, USA.

Email: ceusters@buffalo.edu.

Internet: http://org.buffalo.edu/RTU.

Phone: +1 716 881 8971

Abstract. PURPOSE: A substantial fraction of the observations made by clinicians and entered into patient records are expressed by means of negation or by using terms which contain negative qualifiers (as in "absence of pulse" or "surgical procedure not performed"). This seems at first sight to present problems for ontologies, terminologies and data repositories that adhere to a realist view and thus reject any reference to putative non-existing entities. Basic Formal Ontology (BFO) and Referent Tracking (RT) are examples of such paradigms. The purpose of the research here described was to test a proposal to capture negative findings in electronic health record systems based on BFO and RT. **METHODS**: We analysed a series of negative findings encountered in 748 sentences taken from 41 patient charts. We classified the phenomena described in terms of the various top-level categories and relations defined in BFO, taking into account the role of negation in the corresponding descriptions. We also studied terms from SNOMED-CT containing one or other form of negation. We then explored ways to represent the described phenomena by means of the types of representational units available to realist ontologies such as BFO. R**ESULTS**: We introduced a new family of 'lacks' relations into the OBO Relation Ontology. The relation **lacks_part**, for example, defined in terms of the positive relation **part_of**, holds between a particular p and a universal U when p has no instance of U as part. Since p and U both exist, assertions involving '**lacks_part**' and its cognates meet the requirements of positivity. **CONCLUSION:** By expanding the OBO Relation Ontology, we were able to accommodate nearly all occurrences of negative findings in the sample studied.

Keywords: referent tracking, negation, negative findings, ontology, electronic health record

1. Introduction

A substantial part of the observations made by clinicians are entered into patient records as 'negative findings', i.e. as statements documenting that something *is not the case*. Typical examples are statements such as 'no headache', or 'no known drug allergies'. Elkin *et al* found SNOMED-CT to provide coverage for 14,792 concepts in 41 health records from Johns Hopkins University, of which 1,823 (12.3%) were identified as negative by human review [1]. Mutalik *et al* report the presence of 8,358 instances of UMLS concepts in 60 documents, of which 571 (6.8%) involved negation [2]. Such negative findings are no less important than positive ones for accurate medical decision-making, and failure to document pertinent negative findings may also have medico-legal consequences in connection with claims of malpractice. In 1998, an NHS Independent Review panel judged the record-keeping in a specific case to fall below the level of good practice because '*the notes make no reference to any other findings, nor of any negative ones which would be relevant when considering problems specific to diabetes. Thus no reference is made to the absence of a smell of ketones on Miss J's breath, nor any other negative indications*' [3]. In the US, Medicare and Medicaid compliance requires that in the patient record '*abnormal and relevant negative findings of the examination of the affected or symptomatic body area(s) or organ system(s) should be documented.*' [4].

Standardized terminologies accordingly contain many terms in which some form of negation is used. When the January 2006 version of SNOMED-CT is queried for the occurrence of the word "absence" by means of the Virginia Tech SNOMED CT® Browser [5], for example, 1137 descriptions are retrieved, examples being "absence of scapula", "absence of breast", and so forth. 1336 descriptions are retrieved involving

the term 'absent', as in "absent leg", "absent eyebrow, "bone absent", "absent skin test reaction", "absent bone in hand", "acquired absent testis", and so forth.

A similar query for "not" returns 7272 descriptions, including: "not breathing", "not constipated", "not feeling great", "kidney not palpable", etc., and for "negative" 1058 descriptions, including: "Joint stress test negative". The distribution of these descriptions over the various SNOMED-CT concept categories is illustrated in Table 1.

<INSERT TABLE 1 HERE>

Terms of this sort do not pose problems of understanding for physicians or nurses: as experts in biomedicine they are familiar with corresponding specialised usage, and as human beings they can deal with the intrinsic ambiguities of natural language. For information systems and software agents, in contrast, such terms cause problems, and they have been shown to be associated with a number of characteristic errors when used for purposes of automatic reasoning [6, 7]. There are many reasons for this. One is that reasoning systems themselves involve a logic of negation which does not gell with the uses of negation in standard terminologies. Second, the treatment of negation in popular computational idioms such as OWL DL itself involves non-trivial (and sometimes confusingly documented) features which set traps for inexpert users [8].

A more general reason is that terminologies have thus far been built primarily on the basis of what is called the concept-based paradigm. This means that terminologies are conceived as being built not out of terms but rather out of what are called 'concepts', in order, it is said, to abstract away from incidental syntactic features of the former and to focus instead on common meanings.

Unfortunately the term 'concept' is itself thereby used in a variety of conflicting ways, to refer sometimes to these common meanings, sometimes to entities which are

themselves asserted to *have* meanings, sometimes to psychological entities (for example to the ideas in the minds of those who use the corresponding terms), and sometimes to classes or properties or attributes in reality [9]. As a result of this congeries of interpretations, adequate quality control in concept-based systems is difficult to achieve [10, 11]. This means in turn that most such systems suffer from idiosyncrasies of various sorts – including, most importantly for our present purposes, misclassifications of terms containing negation.

SNOMED-CT, for example, has a large number of 'concepts' classified as procedures. We here leave aside the general issue as to whether it is appropriate to classify procedures as concepts, and focus instead on the specific problem posed by those cases where the SNOMED terms involve some form of negation. Examples are "Medication not administered", "Biopsy specimen not retrieved", "Surgical biopsy not taken", "Metabolic function not tested", and so forth. Strikingly, some of these terms are subsumed by terms which would enable us to infer that they themselves designate procedures. "Metabolic function not tested", for example, is subsumed by "Metabolic function test (procedure)". Of a similar nature is the misclassification of the term "Topography not applicable" which in SNOMED-CT is taken to be a body region.

As pointed out in [7], such misclassifications reflect in part a confusion of epistemology with ontology. Facts pertaining to what clinicians know, or do not know, about entities on the side of the patient are converted by the terminology into entities on the side of the patient. The very possibility of such conversion is however once again a consequence of the application of the concept-based paradigm, since the latter provides so little clarity as to the distinction between the realm of clinicians' statements (observations, terms, concepts, ideas, knowledge) and the realm of entities in reality to

which such statements would be addressed. Thus in particular it provides no means to distinguish ontologically between "what is done" and "what is not done", since both are of course equally respectable concepts.

Such practices are not acceptable under paradigms that adhere to a view based on unqualified realism such as Basic Formal Ontology (BFO) [12] and Referent Tracking (RT) [13]. But because of the importance in biomedicine of our being able to deal with terms that, at first sight, seem to refer to what does not exist, these systems must be able to give an account of the meanings of such terms from the realist perspective.

In this paper, we introduce the 'lacks' relation to achieve this goal. We first explain the basics of BFO and RT, and then demonstrate how 'lacks' fits into the theory underlying both systems.

1.1. Basic Formal Ontology

BFO is a framework that is designed to serve as basis for the creation of high-quality shared ontologies in the domain of natural science, and that embraces a methodology which is realist, fallibilist, perspectivalist, and adequatist [12]. This implies a view according to which: (1) reality and its constituents exist independently of our (linguistic, conceptual, theoretical, cultural) representations thereof, (2) our theories and classifications can be subject to revision motivated by what we discover about this reality, (3) there exists a plurality of alternative, equally legitimate views on reality, and (4) that these alternative views are not reducible to any single basic view. It is (1), above all, which is important for us here.

BFO subdivides reality into a number of basic categories. First, it distinguishes *particulars* from *universals*; the former are entities such as: the authors of this paper,

the surgical procedure that the first author underwent when he was 11 years old; the latter are entities such as: *person* and *appendectomy*, which have the former as their instances. Clinical practice and experimentation relate primarily to the former; scientific theories, which are concerned with what is general in reality, primarily to the latter.

Second, BFO distinguishes within the realm of particulars between *continuants* and *occurrents*. Continuants are entities – such as the first author of this paper; his dedication to the use of realist ontology in healthcare information systems – that endure continuously through a period of time while undergoing changes of various sorts. Occurrents, in contrast, *are* such changes; they are entities (otherwise called 'processes,' 'actions', 'events') which unfold over a certain time through successive temporal parts or phases. However, not all occurrent entities are segmentable in this way into temporal parts or phases, because there are beginnings and endings and other boundaries in the realm of occurrents, and the latter are instantaneous: they are analogous to the edges and surfaces of objects in the realm of continuants. Just as such spatial boundary-entities can exist only as the boundaries of three-dimensional spatially extended objects, so temporal boundary-entities can exist only as the boundaries of temporally extended processes. Typically, the beginning and ending of an occurrent, as well as everything that takes place between these two points, are parts of the occurrent itself. The beginning and ceasing to exist of a continuant, in contrast, are not parts of the continuant itself; rather, they are parts of that occurrent which is its life or history.

Third, there is the distinction between *dependent* and *independent* entities, where each dependent entity is defined as being such that it cannot exist without some independent entity as its bearer. A dedication towards some goal, for example, cannot

exist without a cognitive being that hosts this dedication. Temperatures, body weights and heights similarly cannot exist without some material entity in which they inhere.

Fourth, there is the distinction between *fiat* and *bona fide* entities, which is based on the opposition between fiat and bona fide (or physical) boundaries, the latter being exemplified by boundaries – such as the boundary of Utah, or of the 20th century – introduced via human demarcation [14]. Bona fide boundaries, in contrast, are parts of brute physical reality, and exist independently of any demarcations or decisions which we elect to make.

BFO also distinguishes three major families of relations between the entities just sketched: (1) *<p, p>*–relations, obtaining between particular and particular (for example: Werner Ceusters being **identical_with** the first author of this paper); (2) *<p, U>*-relations, obtaining between particulars and universals (for example: Werner Ceusters being an **instance_of** the universal *person*); and (3) *<U, U>*-relations, obtaining between universal and universal (for example: *scientific paper* being a *subkind_of artifact*) [15]. (We here use *italic* for relations exclusively involving universals, and *italic* for all other relations.)

The importance of this trichotomy is exemplified by the fact that relationships such as parthood have distinct properties at the particular and at the universal levels, the failure to pay to attention to which has led to a number of erroneous representations of relations crucially important in the domain of medical care [16].

1.2. Referent Tracking

Referent tracking (RT) is a new paradigm for representing and keeping track of particulars that has been introduced to support the entry and retrieval of data in

electronic health records (EHRs) [17]. Its purpose is to avoid the ambiguity that arises when statements in an EHR refer to the patient, or to entities such as disorders, lesions on the side of the patient, exclusively by means of generic terms from a terminology or ontology.

Suppose, for example, that two physicians are treating the same patient McX, and that each enters into the EHR a statement to the effect that they observed McX suffering from some problem Y. On current regimes for data entry into EHRs it is then left unspecified whether the physicians in question are referring to the same or to different entities on the side of the patient. Suppose Y is, for example, diabetes. Here only one answer is possible: a patient cannot suffer from a simultaneous plurality of diabetes, and while humans will likely face no problems should an EHR fail to conform to this constraint, for software agents programmed to make inferences from the data such failure will cause problems. Suppose, however, that Y stands in for 'fracture of the right lower arm': this failure will cause problems both for software agents and for humans. The reason is that the physicians in question might have been referring either to the same or to two different fractures, and in the latter case either to distinct fractures present simultaneously in different parts of the lower arm of the patient, or to distinct fractures in the same spot that have occurred at different times, or to any combination thereof. Referent tracking avoids such ambiguities by introducing unique identifiers, called IUIs (for 'Instance Unique Identifiers'), for each numerically distinct entity that is referred to in statements in a record. It The paradigm thus represents a radical generalization of current EHR practices, where unique identification is restricted to independent continuant physical entities such as patients, care providers, buildings, machines and so forth, in requiring the provision of unique identifiers for the entire vast

variety of clinically salient real-world instances, including fractures, polyps, seizures, and all those other entities currently referred to in EHRs in ambiguous fashion by means of general terms alone.

To effectuate this requirement in the concrete form in a Referent Tracking System (RTS) designed to serve the needs of the healthcare enterprise, we need at least:

(1) a mechanism for generating IUIs that are guaranteed to be unique strings;

(2) a procedure for deciding which particulars should receive IUIs;

(3) protocols for determining whether or not a particular has already been assigned a IUI (each particular should receive at most one IUI in order to ensure that information about particulars will exist in integrated form even where it is scattered across a plurality of information systems);

(4) rules governing the processing of IUIs in information systems in general, including rules concerning the syntax and semantics of statements containing IUIs;

(5) methods for determining the truth values of propositions that are expressed through descriptions in which IUIs are employed;

(6) methods for correcting errors in the assignment of IUIs and for investigating the results of assigning alternative IUIs to problematic cases;

(7) methods for taking account of changes in the reality to which IUIs get assigned, for example when particulars merge or split;

(8) methods for associating IUIs with general terms from terminologies specifying the types of entities to which the IUIs have been assigned.

When faced with a statement to the effect that "McX has a fracture of the right lower arm", we would assign IUIs as follows:

#1: McX

#2: the specific fracture to which the statement refers

#3: McX's right arm

#4: the lower portion of McX's right arm.

The statement itself would then be converted to a conjunction of statements of the forms:

#1 has a #2

#2 **instance_of** *fracture*

#4 **instance_of** *right lower arm*

#3 **instance_of** *arm*

#3 **part_of** #1

#2 **inheres_in** #4

#4 **part_of** #3,

Statements of this sort can easily be written as RDF-triples and are thus able to contribute to the endeavours of the Semantic Web.

Ideally, relational expressions such as '**part_of**', '**instance_of**', and so on, would then be drawn from a suitable relation ontology [15]. Terms such as 'fracture' and 'right lower arm' would come from an ontology faithful to the principles of BFO, so that the terms in question would inherit their customary meanings. But the referent tracking paradigm allows also the use of terms drawn from concept systems. Statements such as '#4 **instance_of** *right lower arm*' would then signify that, within the linguistic and scientific community in which the given concept system is used, it is acceptable to use the term 'right lower arm' to refer to the particular in question.

The proposal to enforce systematic identification of particulars is a novel idea when applied in the EHR domain; but this idea is itself not new. It has been embraced by

scholars in the domain of computer science, for example in [18], which argues that problems in database schema integration, schema evolution, and interoperability are precisely the consequence of the ambiguities brought on by the use of general terms with no adequate attention to the underlying particulars. At the heart of the problem, according to [18], is the erroneous assumption adhered to in database design circles according to which entities can be referred to in every case as instances of pre-specified classes. The authors term this *the assumption of inherent classification* and make the case that this assumption violates philosophical and cognitive guidelines on classification.

1.3. The problem of negative findings

In [17] we have described a formal framework that is able to deal with phenomena in reality by means of elementary statements of the sorts just described (#1 has a #2, etc.), at the same time specifying the role to be played by terminologies and ontologies in this framework (Table 2). The problem which confronts us here turns on the fact that referent tracking adheres to the realist philosophy imposed by BFO. It thus needs to take into account a constraint to the effect that only entities that exist are to be assigned a IUI. How, then, can it deal with the 'negative findings' or 'negative observations' captured in expressions such as: "no history of diabetes", "hypertension ruled out", "absence of metastases in the lung", and "abortion was prevented"? Such statements seem at first sight to present a problem for the referent tracking paradigm, since they imply that there are no entities on the side of the patient to which appropriate unique identifiers could be assigned.

INSERT TABLE 2 HERE

2. Objectives

If referent tracking is to be accepted as a viable paradigm for EHR management, it has to be able to deal with phenomena of the mentioned sort. Our objective is thus to expand the set of statements with which an RTS can currently deal, in such a way as to allow representations of those portions of reality in which something is *not* the case without violating the basic principles of realist ontology [12].

3. Material and methods

We analysed a series of negative findings encountered in 748 sentences drawn from 41 patient charts from Johns Hopkins University [1]. We assumed such findings to be descriptions of real phenomena on the side of the patient (including aspects of the patient's environment). We classified these phenomena in terms of the various top-level categories and relations defined in BFO and taking into account the role of negation in the corresponding descriptions. We also studied terms from SNOMED-CT containing one or other form of negation. We then explored ways to represent such phenomena by means of the types of representational units available on the Referent Tracking paradigm.

4. Results

4.1. Negative findings at the level of particulars

Table 3 lists the four headings under which negative findings can be classified when we take into account BFO's distinction between particulars and universals and the types of

relationships that can obtain between them. $<U, U>$-relations do not belong to the realm of particulars and are thus excluded from our purview here. The last column of Table 3 shows the distribution of the occurrence of different types of negative findings in the analysed sample.

INSERT TABLE 3 HERE

On the basis of this analysis, we argue that at least one new relationship needs to be included in the machinery of BFO, which we define as follows:

p **lacks** *U* **with respect to** *r* **at time** *t*: there obtains a relation between the particular *p* and the universal *U* at time *t* which is such that *p* stands to no instance of *U* in the relationship *r* at *t*

Relations in the **lacks** family are involved in phenomena that are described by means of negative findings of types C1 and C3 in Table 3. If, for example, a patient (an independent continuant particular) does not have a right hand (also an independent continuant particular), then no instance of the universal right hand is **part_of** that patient at the given time. For C3-type phenomena, the relation with respect to which **lacks** holds, is one of **instantiation**: if for a disorder on the side of a patient it is ruled out that it is a primary hyperaldosteronism, then that disorder (an existing particular) is not an **instance_of** the universal *primary hyperaldosteronism*.

To accommodate this new ontological relation under the referent tracking paradigm, a new type of tuple is required, which we will call U^{-}, which is a **lacks**-counterpart of the U-tuple defined in [17].

$U^-_i = \langle IUI_a, t_a, \mathbf{r}, o, IUI_p, u, t_r \rangle$ The particular referred to by IUI_a asserts at time t_a that the relation $\mathbf{r}$ of ontology o does not obtain at time t_r between the particular referred to by IUI_p and any of the instances of the universal u at time t_r

As an example, if Dr. McY asserts on May 5, 2006 that patient McX lacks his right hand since a car accident, then a U⁻ statement would be entered into the corresponding Referent Tracking System in which IUI_a is the unique identifier assigned to Dr. McY, t_a is a standardised string representing May 5, 2006, $\mathbf{r}$ the parthood relationship, o the ontology in which the parthood relationship is defined, IUI_p the IUI assigned to McX, and $t_{\mathbf{r}}$ a standardised string denoting the time at which the car accident occurred. The U-slot would be filled by the identifier assigned to the universal *right hand* in ontology o.

In addition to U-tuples, the RTS defined in [17] contains also A-tuples, which express the assignment of a IUI to a particular. There is here no need to define an A⁻ tuple (as the lacks-counterpart of the A-tuple), since referent tracking does not allow the assignment of IUIs to non-existing entities.

We think there is similarly no need for a **lacks**-counterpart of other relations between particulars and would thus accommodate for C4 types of negative findings such as "the patient does not live in his house anymore" by means of simple logical negation applied to statements about particulars. Negative findings of type C2 are special cases of C3: there is no universal that instantiates them at the indicated time t_p

4.2. Negative findings at the level of universals

There are many circumstances under which it would be useful to include in realism-based ontologies representational units that express some form of negation, primarily findings of absence. Another application would be to give a realist interpretation to terms found in terminologies that refer to absences as putative entities.

In line with [15], we can thus define:

C1 *lacks* C2 *with respect to* r = [definition] for all c, t if c **instance_of** C1 **at** t then c **lacks** C2 **with respect to** r **at time** t.

Using this relationship, we are able to describe the universal *acephalus* as being instantiated by entities that lack a head:

Acephalus *lacks* Head *with respect to* has_part.

We can also use this relationship to give a realist interpretation of terms such as 'head trauma without loss of consciousness':

HeadTraumaWithoutLossOfConsciousness *lacks* LossOfConsciousness *with respect to* associated_with.

5. Discussion

We based our study on two different sources: sentences found in electronic health records containing some form of negation, and terms retrieved from SNOMED-CT by querying for standard negative formulations.

The sentences from the first sample were extracted from the patient charts by natural language parsing software sensitive to textual clues for negation [1]. Some sentences were returned erroneously because of misleading textual clues, e.g. '*The patient actually answers yes, no, and sir to all questions*'. Furthermore, not all sentences containing negation are descriptions of negative findings, e.g. '*He has no idea why he is here*'. Finally, some sentences described a 'positive' phenomenon in a 'negative' way, such as '*Her workup showed that she had an MRI of the brain that was negative in 03/02*' (a case in which the clinician actually states that the MRI is normal). None of these sentences (3.2% of the sample) were included in the analysis.

Where sentences contained modal and intensional operators we extracted for analysis the proposition to which these operators were applied. Thus for '*He has no family history of GI malignancies that I know of*' we analyzed only the proposition '*He has no family history of GI malignancies*'. (The reason is that modal and intensional operators require a second-order treatment whose discussion falls beyond the scope of this paper.) The resultant sentences accounted for 12.3% of our sample.

5.1. A realism-based typology of negative findings

The two sets of samples together exhibited phenomena expressed by means of negation that could be classified into four categories.

Category C1: The first and largest category comprises sentences representing the non-existence of a particular, sentences which can be divided further into

- cases where an *independent continuant* is absent: the absence of a left hand in some particular patient as the result of a developmental disorder, the absence of children in a family marked by reproductive disorders.

- cases where a *dependent continuant* is absent: the absence of headache during a specified time period, the absence of the ability to walk.

- cases where a *dependent occurrent* is absent: the non-occurrence of a spontaneous abortion because preventive measures were taken, the absence of tremors.

Such statements are typically registered only when the non-existence in question has clinical significance, for example because entities of the relevant sorts ought to exist under normal circumstances (congenital absence of left hand), or because something that is likely to exist under relevant circumstances, and is thus expected to exist, fails to do so (absence of headache after serious head trauma). It might, therefore, be necessary to identify two types of cases involving negative relations – cases of **lacks** in the strict sense, where there is an implicit "in the normal case the relation holds" (the entity in question *needs* the thing it lacks), and a simpler relation, which we can call *is_without*. In the latter case, a further distinction can be made between situations in which only a limited number of particulars of a given sort are without another particular (e.g. head traumas without headache), or no particulars are associated with another particular (no invertebrates have a spinal column). Another issue that needs further inquiry are the different conditions that apply to lacks (and is_without) for occurrents than for continuants.

Non-existent entities have been a subject of debate in philosophy at least since Parmenides, whose central thesis is one to the effect that *that which is not* cannot be thought about or spoken about, so that to follow what Parmenides calls the 'way of truth' is to accept that *to think* and *to think of something existing* are one and the same [19]. Where, more recently, Kant [20] and Frege [21] held that it was a mistake to suppose that 'existence' is a predicate which could be asserted or not asserted of a given entity – an existing dollar is not a special kind of dollar – other philosophers such as Meinong embraced the notion of non-existent entities. For purposes of natural science, however, where we are dealing with entities localized in space and time, it seems wrong to include among such entities both existents and non-existents. Again: an absent leg is not a special kind of leg because it is not a kind of entity at all.

The introduction of the 'lacks' relation enables us to do justice to negative findings without stepping beyond the bounds of what exists. This is because 'lacks' allows us to perceive negative phenomena such as absences as a positive relation that holds between some existing entity and some corresponding universal, an instance of which would exist (ought to exist, is expected to exist) if it were not absent.

Category C2: The second category pertains to the absence of an entity which did previously exist in the patient in question. It is exemplified by cases where a patient did have a left hand up to certain point in time, but then lost it because of a surgical or traumatic amputation. This case is unproblematic for RT, since there has an existing particular to which a unique identifier can be assigned did indeed exist, and the identifier can still be used even after the entity itself has gone out of existence.

Category C3: The third category involves statements to the effect that an entity is not an instance of some clinically salient universal. An example is 'non-smoker'.

Although some terminologies regard constructions of this sort as unproblematic, they are out place in realist ontologies since there are no corresponding universals on the side of reality. A sentence to the effect that a particular person is a non-smoker, is, we believe, more correctly to be analysed as expressing a proposition to the effect that the person in question is not an instance of the class *smoker*.

Category C4: Finally, there are cases in which the negative phenomenon is of a sort such that some entity does not have certain clinically salient properties. Every organism with a heart has a heart that beats, and all beating hearts beat at a certain frequency. The frequency may vary, i.e. it may exhibit different values, over time. Similarly, all human bodies, as long as they exist, have a certain weight, but this weight may change over time. Properties such as heart beat frequency and weight are expected to exhibit values that fall within certain ranges. Actual weights or frequencies may be in the high or low ends of such ranges, and terms have been constructed that capture succinctly an entity's property of being in a clinically relevant range, examples being 'bradycardia', 'tachycardia', 'overweight', and so forth. If such a term is negated, it does not entail that the patient under scrutiny does not have a heart, nor that it is not beating, but only that it is not beating within the frequency range referred to by the term. Importantly, however, it depends on the ontological nature of the property in question whether any actual value can be inferred from one or other negative proposition of the given sort. For properties such as 'being dead' and 'being alive', negation of the former entails the latter, and vice versa. But if the patient is said not to be bradycardic, this does not mean that his heartbeat frequency is in the normal range, for tachycardia may still be a valid situation to take into account.

Some negative findings could be classified in one of the 4 categories of Table 3, yet still describe phenomena that require logical extensions of the Referent Tracking paradigm. An example is: '*no other complications of gastro-esophageal reflux disease*'. These form part of a larger family of cases, in which various kinds of numerical and other quantifiers (such as 'many' are involved). 'Missing finger', too, is of this type; since for a patient to have a missing finger is not for him to have no fingers, but rather to have a number of fingers that is smaller than the norm.

5.2. Coding negative findings in an RTS

If a clinician describes a phenomenon on the side of a specific patient using the phrase 'absence of metastases in the lungs', then this would be registered in a Referent Tracking System using some coding along the following lines:

$$U^-_{61092} = <\#2354, \text{'2005-12-27-18:40'}, \textbf{contains}, \#678, \#9100, \text{metastasis}, \text{'until 2005-12-27-18:40'}>$$

in which #2354 would be the IUI of the clinician, '2005-12-27-18:40' the time of the assertion, **contains** the inverse of the instance level relation **contained_in** from the OBO Relation Ontology [15], #678 the IUI of this ontology, #9100 the IUI of that patient's lungs, 'metastasis' a reference to the universal *metastasis*, and 'until 2005-12-27-18:40' a description of the time-interval during which the **lacks** relation holds (in line with the provisions of EN 12338:2005) [22]. By representing this statement in some adequate logic, now, and by applying to it associated rules of inference, further assertions can be derived, e.g. that any particular contained in that patient's lung is not

a metastasis, and that if there is a metastasis contained in some body part of the given patient, then that body part is not the lung.

6. Conclusion

By introducing the **lacks**-relation and by introducing the new tuple-type U⁻ whose semantics is based on **lacks**, we are able to represent nearly all negative findings that occur in patient charts while remaining faithful to the principles of unqualified realism. One quite general implication of our the position is indeed that negation lies outside the realm of ontology, but belongs rather to the domains of logic [23], language [24] and epistemology [25]. Claiming the opposite would be symptomatic for what Smith called '*fantology*', i.e. the false belief that the structures of logic and language (and information models) are mirrors of the structure of reality [26]. In reality, there is only what there is, and the fact that language allows us to describe what there is also by alluding to what there is not (and that logic allows us to reason about reality in corresponding ways) does not imply that reality is such as to include what does not exist. For the corresponding negative expressions do not mirror anything in reality, and to suppose that they do is a confusion of the way language works.

Table 1: Number of descriptions retrieved for the queries "not", "negative" and "absence" using the Virginia Tech SNOMED CT® Browser on the January 2006 version of SNOMED-CT.

Concept type	Descriptions retrieved		
	"absence"	"negative"	"not"
Anatomical concepts (Body Structure)	55		
Attributes	2		
Body structure	68	7	21
Clinical findings	3229	503	828
Context Dependent categories	1073	90	122
Events	1328		
Morphologies (Body Structure)	10	7	18
Observable entity	2	9	
Organism	43	71	
Pharmaceutical/biologic product		43	
Physical force		2	
Physical object	3		
Procedure	156	46	4
Qualifier value	137	21	5
Social Context	11		
Special concept	1151	250	139
Staging and Scales		6	
Substance	4	3	
TOTAL	**7272**	**1058**	**1137**

Table 2: Ontology-related tuple types in Referent Tracking

Tuple type	Phenomenon described
$A_i = <IUI_p, IUI_a, t_{ap}>$	Assignment of IUI_p to a particular at time t_{ap} by the particular referred to by IUI_a *
$R_i = <IUI_a, t_a, \mathbf{r}, o, P, t_r>$	It is asserted by the particular referred to by IUI_a at time t_a that the relationship r from ontology o obtains between the particulars referred to by P at time t_r
$U_i = <IUI_a, t_a, \mathbf{inst}, o, IUI_p, u, t_r>$	It is asserted by the particular referred to by IUI_a at time t_a that the instantiation relation as defined in ontology o obtains between the particular referred to by IUI_p and the universal u at time t_r

* the subscript 'p' stands for 'particular' and 'a' for 'author'

Table 3: categories of negative findings from the perspective of BFO

	Relation type	Type of Negative Finding	Examples	%
C1	_	Non-existence of a particular	*he denies abdominal pain; no alcohol abuse; no hepatosplenomegaly; 'he has no children'*	95,8
C2	_	Absence of previously existing particular at some given later time t	[†] no muscle pain anymore	0
C3	<p, u> *	Particular not being the instance of a class at some given time t	'… *which ruled out primary* hyperaldosteronism'; '*without any cyanosis*'	4,1
C4	<p, p>	Particular not being related to another particular in a specific way at time t	'*this record is not available to me*'	0,1

* p stands for particular, u for universal
[†] theoretical example given for completeness, although none of this sort was found in the sample

References

1. Elkin, P.L., et al., *A controlled trial of automated classification of negation from clinical notes.* BMC Medical Informatics and Decision Making, 2005. 5: p. 13.
2. Mutalik, P.G., A. Deshpande, and P.M. Nadkarni, *Use of general-purpose negation detection to augment concept indexing of medical documents: a quantitative study using the UMLS.* J Am Med Inform Assoc, 2001. 8: p. 598-609.
3. Health Service Ombudsman for England. *Errors in the care and treatment of a young woman with diabetes.* 1998 [cited 2006 December 23]; Available from: http://www.ombudsman.org.uk/improving_services/special_reports/hsc/diabetes/index.html.
4. Centers for Medicare and Medicaid Services. *Documentation Guidelines for Evaluation and Management Services.* 1997 [cited 2006 December 12]; Available from: http://www.cms.hhs.gov/MedlearnProducts/downloads/1995dg.pdf.
5. *Veterinary Terminology Services.* 2006 [cited 2006 December 25]; Available from: http://terminology.vetmed.vt.edu/default.htm.
6. Ceusters, W., B. Smith, and L. Goldberg, *A terminological and ontological analysis of the NCI Thesaurus.* Methods of Information in Medicine, 2005. 44: p. 498-507.
7. Bodenreider, O., B. Smith, and A. Burgun, *The ontology-epistemology divide: A case study in medical terminology*, in *Formal Ontology and Information Systems*. 2004. p. 185-195.
8. Rector, A.L., et al., *OWL Pizzas: Practical Experience of Teaching OWL-DL: Common Errors & Common Patterns.*, in *Engineering Knowledge in the Age of the SemanticWeb; Proceedings of the14th International Conference, EKAW 2004 Whittlebury Hall, UK, October 5-8, 2004*. 2004, Springer: Berlin/Heidelberg. p. 63-81.
9. Smith, B., *Beyond concepts: ontology as reality representation*, in *Proceedings of the third international conference on formal ontology in information systems (FOIS 2004)*. 2004, IOS Press: Amsterdam. p. 73-84.
10. Smith, B., W. Ceusters, and R. Temmerman, *Wüsteria*, in *Connecting Medical Informatics and Bio-Informatics. Medical Informatics Europe 2005*, R. Engelbrecht, et al., Editors. 2005, IOS Press: Amsterdam. p. 647-652.
11. Ceusters, W., et al., *Mistakes in medical ontologies: Where do they come from and how can they be detected?*, in *Ontologies in Medicine. Studies in Health Technology and Informatics*, D.M. Pisanelli, Editor. 2004, IOS Press: Amsterdam, The Netherlands. p. 145-164.
12. Grenon, P., B. Smith, and L. Goldberg, *Biodynamic Ontology: Applying BFO in the Biomedical Domain*, in *Ontologies in Medicine*, D.M. Pisanelli, Editor. 2004, IOS Press: Amsterdam. p. 20-38.
13. Ceusters, W. and B. Smith, *Referent Tracking in Electronic Healthcare Records*, in *Connecting Medical Informatics and Bio-Informatics. Medical Informatics Europe 2005*, R. Engelbrecht, et al., Editors. 2005, IOS Press: Amsterdam. p. 71-76.
14. Smith, B. and A.C. Varzi, *Fiat and Bona Fide Boundaries: Towards on Ontology of Spatially Extended Objects* in *Lecture Notes In Computer Science*. 1997, Springer Verlag: London, UK. p. 103 - 119.
15. Smith, B., et al., *Relations in biomedical ontologies.* Genome Biology, 2005. 6(5): p. R46.
16. Donnelly, M., T. Bittner, and C. Rosse, *A formal theory for spatial representation and reasoning in biomedical ontologies.* Artificial Intelligence in Medicine, 2006. 36(1): p. 1-27.
17. Ceusters, W. and B. Smith, *Strategies for Referent Tracking in Electronic Health Records.* Journal of Biomedical Informatics, 2006. 39(3): p. 362-378.
18. Parsons, J. and Y. Wand, *Emancipating Instances from the Tyranny of Classes in Information Modeling.* ACM Transactions on Database Systems, 2000. 25(2): p. 228-268.
19. Parmenides of Elea. *On Nature.* ca 475 BC [cited 2006 December 31]; Available from: http://www.elea.org/Parmenides/.
20. Kant, I., *The critique of pure reason.* 2003, Project Gutenberg.
21. Frege, G., *Function and Concept*, in *Translations from the Philosophical Writings of Gottlob Frege*, P. Geach and M. Black, Editors. 1966, Blackwell: Oxford. p. 21-41.
22. European Committee for Standardization, *EN 12388:2005. Health informatics - Time standards for healthcare specific problems.* 2005.
23. Poli, R., *Descriptive, Formal and Formalized Ontologies*, in *Husserl's Logical Investigations Reconsidered*, D. Fisette, Editor. 2003, Kluwer: Dordrecht. p. 193 - 210.
24. Carston, R., *Negation, 'presupposition' and the semantic/pragmatic distinction.* Journal of Linguistics, 1998. 34(2): p. 309-350.
25. Pacitti, D., *The Nature of the Negative: Towards an Understanding of Negation and Negativity.* 1991, Pisa: Giardini.
26. Smith, B., *Against Fantology*, in *Experience and Analysis*, M.E. Reicher and J.C. Marek, Editors. 2005: Wien. p. 153-170.