Referent Tracking: The Problem of Negative Findings

Werner CEUSTERS[a,1], Peter ELKIN[b] and Barry SMITH[a, c]

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

Abstract. The paradigm of referent tracking is based on a realist presupposition which rejects so-called negative entities (*congenital absent nipple*, and the like) as spurious. How, then, can a referent tracking-based Electronic Health Record deal with what are standardly called 'negative findings'? To answer this question we carried out an analysis of some 748 sentences drawn from patient charts and containing some form of negation. Our analysis shows that to deal with these sentences we need to introduce a new ontological relationship between a particular and a universal, which holds when no instance of the universal has a specific qualified ontological relation with the particular. This relation is found to be able to accommodate nearly all occurrences of negative findings in the examined sample, in ways which involve no reference to negative entities.

Keywords: referent tracking, negation, negative findings, ontology, realism, EHR

1. Introduction

Referent tracking has been introduced as a new paradigm for entry and retrieval of data in the Electronic Health Record (EHR) [1]. Its purpose is to avoid the ambiguity that arises when statements in an EHR refer to disorders, lesions and other entities on the side of the patient exclusively by means of generic terms from a terminology or ontology. Suppose that two different physicians are treating the same patient A, and that each enters into A's EHR a statement to the effect that A suffers (i) from diabetes or (ii) from a fracture of the right lower arm. Then it is in either case left unspecified whether they are referring to the same or to different entities on the side of the patient. In case (i), it is clear that only one answer is possible; yet the ambiguity as to whether each of the two physicians is referring to the same diabetes will still cause problems for software agents programmed to make inferences from the data. In case (ii) this ambiguity causes problems even for human beings, since the physicians in question might have been referring either to the same or to different fractures.

Referent tracking avoids such ambiguities by introducing unique identifiers, called IUIs or Instance Unique Identifiers, for each numerically distinct entity that exists in reality and that is referred to in statements in the record. The referent tracking paradigm thereby expands the entities uniquely identified for EHR purposes far beyond the current range, which is restricted to entities such as patients, care providers, buildings,

[1] Corresponding Author: Werner Ceusters, Ontology Research Group, Center of Excellence in Bioinformatics and Life Sciences, 901 Washington Street, Buffalo, NY 14203, USA. Email: ceusters@buffalo.edu. Internet: http://org.buffalo.edu.

machines, and so forth. A statement such as "John Doe has a fracture of the right lower arm", would be translated under the referent tracking paradigm into: '#1 has #2', the first number being the IUI for John Doe, the second the IUI for that specific fracture he is suffering from. Additional statements would then specify that '#2 is a fracture of the right lower arm' or, better, that '#2 is a #3 located in #4', together with the extra information that #3 is a fracture and #4 is John Doe's right lower arm. Expressions such as 'has', 'is a', 'located in', etc., would at the same time be replaced by the appropriate relationships from a suitable ontology [2], for which logical reasoning tools have also been defined.

In [3] we have described a framework that is able to deal with phenomena in reality that can be described by means of directly depicting statements of the sorts just described, at the same time specifying the role to be played by terminologies and ontologies in this framework. We also discussed there how information entered into an EHR system by clinicians in the usual way could be translated automatically into statements of a Referent Tracking System (RTS). One specific problem thus far left untouched is how to represent phenomena commonly called 'negative findings' or 'negative observations' within an RTS. Example statements describing such phenomena are: "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 there are here no entities on the side of the patient to which unique identifiers can be assigned.

2. Objectives, Materials and Methods

If referent tracking is to be accepted as a viable paradigm for the EHR, it has to be able to deal with phenomena of the mentioned sort. Our objective is thus to expand the repertoire of statements with which an RTS can deal in such a way as to allow representations of those relevant portions of reality in which something is not the case. We must do this, however, without violating the principles of Basic Formal Ontology (BFO) [4] upon which referent tracking is built. These principles counsel unqualified realism as a basis for the creation of high-quality shared ontologies in the biomedical domain. This means, most importantly, that our representations can acknowledge only those entities which exist in biological reality, and must reject all those types of putative negative entities – absences, non-existents, possibilia, and the like – which are postulated merely as artefacts of specific logical or computational frameworks.

We analysed 396 negative findings encountered in 250 sentences out of 18 patient charts from Johns Hopkins University [5]. We assumed such findings to be descriptions of real phenomena on the side of the patient and sought to classify the underlying structures and processes in terms of the various top-level categories and relations defined in BFO, taking careful account of the role of negation in the corresponding descriptions. We then explored ways to represent such phenomena by means of the types of representational units available on the referent tracking paradigm.

BFO subdivides reality into a number of basic categories. First, it distinguishes *particulars* from *universals*, the former being entities such as John Doe or the left arm fracture he suffered from last year, and the latter entities such as *person*, *fracture* and *arm*.

Table 1: Ontology-related tuple types in Referent Tracking

Tuple type	Phenomenon described
$A_i = <IUI_p, IUI_a, t_{ap}>$	Act of 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 in the set of IUIs 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'

Second, it distinguishes continuants from occurrents. Continuants are entities, such as John Doe and his left arm, that endure continuously through time. Occurrents, in contrast, unfold over a certain time span through successive temporal parts, examples being entities such as processes, actions and events. Thirdly, there is the distinction between dependent and independent entities, the former being such that they cannot exist without some instance of the latter: John Doe's height or weight, for example, cannot exist without the existence of John Doe himself.

3. Results

BFO distinguishes three major families of relations between the entities just sketched: (1) <p, p>: from particular to particular (for example: John Doe's nose being part of John Doe); (2) <p, u>: from particular to universal (for example: John Doe being an instance of the type *person*); and (3) <u, u>: from universal to universal (for example: *person* being a subkind of *organism*). [2]

Referent tracking applies BFO to the domain of EHRs, requiring: (1) that particulars are referred to by means of unique identifiers (IUIs), (2) that each particular should receive maximally one IUI, and (3) that only entities that exist are to be assigned a IUI. Real world phenomena are then represented in an RTS [3] by means of tuples of the sorts outlined in Table 1.

Table 2 lists the four headings under which negative findings can be classified when account is taken of BFO's distinction between particulars and universals and of the different types of relationships that can obtain between them. The last column of Table 2 shows the distribution of the occurrence of negative findings in the analysed sample. On the basis of our analysis we now argue that there must be included in the machinery of BFO new relations, a new family of formal <p,u>-relations which obtain whenever a given particular does not stand in some given <p,p> relation to any instance of a given universal. The relations in this family we can define more formally as follows:

p **lacks** u **at** t **with respect to identity**
=def. there is no x such that: x **identical_to** p at t and x **instance_of** u
p **lacks** u **at** t **with respect to part**
=def. there is no x such that: x **part_of** p at t and x **instance_of** u

and similarly for other <p,p> relations such as **quality_of, located_in**, **derives_from**, **has_participant**, and so on. Note that the **lacks**-relations are formal relations, analogous to instantiation or parthood. This means that they are not extra ingredients in being, but rather that in virtue of which existing entities are joined together to form larger wholes.

Table 2: categories of negative findings from the perspective of BFO

Relation type	Type of Negative Finding	Examples	%

C1	<p, u> *	A particular is not related in a specific way to any instance of a universal at some given time	*he denies abdominal pain; no alcohol abuse; no hepatosplenomegaly; he has no children, without any cyanosis*	85.4
C2	<p, u>	A particular is not the instance of a given class at some given time	*which ruled out primary hyperaldosteronism, nontender, in no apparent distress, Romberg sign was absent , no palpable lymph nodes*	12.4
C3	<p, p>	A particular is not related to another particular in a specific way at some given time	*this record is not available to me; it is not the intense edema she had before; he has not identified any association with meals.*	2.2

* 'p' ranges over particulars, 'u' over universals

It is **lacks** that is involved in the phenomena described by means of negative findings of types C1 and C2 from Table 2. An example of type C1 arises when a patient (an independent continuant) does not exhibit a headache (a dependent continuant); on our analysis this means that the patient and the universal *headache* (both of which are from the BFO perspective full-fledged entities) stand to each other at a given time in a certain relation, namely: **lacks with respect to the relation has_quality**. C2-type phenomena receive an identical analysis, except that here the relevant relation is **lacks with respect to the relation identical_to**. If, for example, it is ruled out, for a given disorder (p) on the side of a patient, that it is a case of *primary hyperaldosteronism* (u), then it is asserted that at the given time (t) no instance of u is identical to p. Negative findings of type C3 suggest the need for a relation analogous to **lacks**, but holding not between a particular and a universal but between one particular and another. We are not yet sure, however, whether there is a need for a relation of this sort, since the corresponding cases may perhaps be dealt with in terms of the simple logical negation of straightforward statements about the corresponding particulars.

To accommodate the new **lacks** relations in referent tracking, a further tuple type is required, which we will call U^-:

$U^-_i = <IUI_a, t_a, r, o, IUI_p, u, t_r>$ The particular referred to by IUI_a asserts at time t_a that the relation 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

4. Discussion

A substantial fraction of the clinical observations entered into patient records are expressed by means of negation. 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 [5]. Mutalik *et al* report the presence of 8,358 instances of UMLS concepts in 60 documents of which 571 (6.8%) were negations [6]. This is because negative findings are as important as positive ones for accurate medical decision making, and failure to document pertinent negative findings may have medico-legal consequences in case of allegations 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*' [7]. In the US, Medicare and Medicaid compliance requires that the patient record should document '*specific abnormal and relevant negative findings of the examination of the affected or symptomatic body area(s) or organ system(s)*'. [8]

The sentences we studied were extracted from the patient charts by natural language parsing software sensitive to textual clues for negation [5]. Some sentences

were retained erroneously because textual clues were misleading, as in: '*The patient actually answers yes, no, and sir to all questions*'. Furthermore, not all sentences containing negation are descriptions of negative findings; thus '*He has no idea why he is here*' may either refer to the positive finding of being mentally disoriented or be simply a non-clinical statement. A clear example of a sentence describing a positive phenomenon in a negative way is: '*Her workup showed that she had an MRI of the brain that was negative in 03/02*', which in fact states that the MRI was normal. Such sentences (8.3% of the sample) were not included in our analysis. Modal and similar operators were left aside in this analysis, so that for example only the italicized portion of the sentence '*He has no family history of GI malignancies* that I know of' was analyzed. This is because referent tracking has been designed to give modal aspects a second-order treatment, the discussion of which falls beyond the scope of this paper.

Some negative findings could be classified in one of the 3 categories, but describe phenomena that currently cannot be dealt with under the referent tracking paradigm. Examples are: '*no other complications of gastro-esophageal reflux disease*'.

With the introduction of the new **lacks** relations – an expanded version of the rationale for which is provided in [9] – we defend, in effect, the thesis that *negation is outside the realm of ontology* but belongs rather to the domains of logic [10], language [11] and epistemology [12]. Denial of this thesis is symptomatic of what Smith has called '*fantology*', i.e. the false belief that the structures of logic, language and information are mirrors of the structure of reality [13]. In reality, there is only what there is. Language and logic allow us to talk and reason about what there is by using negation. But the corresponding negative expressions do not mirror anything in reality.

Thus, if a clinician describes a phenomenon on the side of a patient using the phrase 'absence of metastases in the lungs', then the corresponding assertion would be registered in an RTS using some coding along the following lines:

U^-_{61092} = <#23, '2005-12-27-18:40', **contains**, #678, #91, *metastasis*, 'until 2005-12-27-18:40'>,

in which #23 would be the IUI of the clinician, '2005-12-27-18:40' the time of assertion, **contains** the inverse of the <p,p>-relation **contained_in** from the OBO Relation Ontology [2], #678 the IUI of the OBO-ontology, #91 the IUI of the 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 12338:2004 [14]). By representing this statement in some adequate logical form and by applying the corresponding inference rules further derivations can then be made, for exampe to the effect that, whatever particular there is in that patient's lung, it is not a metastasis, and that if there is a metastasis contained in some body part of the patient, then that body part is not the lung, and so forth.

Finally, it is possible to define a lacks relation that holds between universals. This would be useful for statements of the sort that all relatives of a patient are disease free or that none of his white blood cells in an examined sample exhibit a certain anomaly.

5. Conclusion

By introducing **lacks** relations of the <p,u> sort together with the new U^- tuple type, we were able to represent 99.9% of the negative findings that occur in the analysed sample (and thus, we believe, of the vast majority of negative findings that occur in EHRs in general) in such a way as to remain faithful to the principles of unqualified realism within an EHR regime based on the idea of faithfulness to clinical reality. Further research is required to assess the need for two other families of **lacks**-like relations holding, respectively, between particulars and between universals.

References

[1] Ceusters W. and Smith B. Referent tracking in Electronic Healthcare Records. In: Engelbrecht R. et al. (eds.) Medical Informatics Europe, IOS Press, Amsterdam, 2005, p. 71-76.

[2] 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.

[3] Ceusters W, Smith B. Strategies for referent tracking in Electronic Health Records. Journal of Biomedical Informatics. In press.

[4] Grenon P, Smith B, Goldberg L. Biodynamic ontology: applying BFO in the biomedical domain. In DM Pisanelli (ed.), Ontologies in Medicine, Amsterdam:IOS Press, 2004, p. 20-38.

[5] Elkin PL, Brown SH, Bauer BA, Husser CS, Carruth W, Bergstrom LR and Wahner-Roedler DL. A controlled trial of automated classification of negation from clinical notes. BMC Medical Informatics and Decision Making 2005;5:13.

[6] Mutalik PG, Deshpande A, Nadkarni PM: 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, 8:598-609.

[7] Health Service Ombudsman for England. Errors in the care and treatment of a young woman with diabetes. http://www.ombudsman.org.uk/improving_services/special_reports/hsc/diabetes/index.html (last accessed: December 27, 2005).

[8] Centers for Medicare and Medicaid Services. 1997 Documentation Guidelines for Evaluation and Management Services. http://www.cms.hhs.gov/MedlearnProducts/downloads/1995dg.pdf. (Last accessed: December 27, 2005).

[9] Simon J, Fielding JM and Smith B. Using philosophy to improve the coherence and interoperability of applications ontologies. In Büchel B, Klein B and Roth-Berghofer T (eds.), Proceedings of the First Workshop on Philosophy and Informatics. DFKI, Cologne: 2004, 65-72.

[10] Smith B. Logic, form and matter, Proceedings of the Aristotelian Society, Supp. Vol. 1981; 55: 47–63.

[11] Carston R. Negation, 'presupposition' and the semantic/pragmatic distinction, Journal of Linguistics 1998;34 (2):309-50.

[12] Pacitti D. The nature of the negative: Towards an understanding of negation and negativity. Giardini, Pisa, 1991.

[13] Smith B. Against fantology. In: Reicher ME, Marek JC (eds.), Experience and Analysis. Vienna 2005, p. 153-170.

[14] EN 12388:2005. Health informatics - Time standards for healthcare specific problems.