International Journal of Medical Informatics (2006) **75**, 224—231

ELSEVIER

International Journal of
Medical Informatics

www.intl.elsevierhealth.com/journals/ijmi

Formal ontology for natural language processing and the integration of biomedical databases

Jonathan Simon[a,d,*], **Mariana Dos Santos**[b], **James Fielding**[a], **Barry Smith**[a,c]

[a] *IFOMIS, Saarland University, Saarbrücken, Germany*
[b] *Language and Computing nv., Ghent, Belgium*
[c] *Department of Philosophy, University at Buffalo, Buffalo, NY, USA*
[d] *Department of Philosophy, New York University, New York City, NY, USA*

KEYWORDS
Medical natural language understanding;
Medical terminologies;
Biomedical systems integration;
Formal ontology;
Biomedical data-mining

Summary The central hypothesis underlying this communication is that the methodology and conceptual rigor of a philosophically inspired formal ontology can bring significant benefits in the development and maintenance of application ontologies [A. Flett, M. Dos Santos, W. Ceusters, Some Ontology Engineering Procedures and their Supporting Technologies, EKAW2002, 2003]. This hypothesis has been tested in the collaboration between Language and Computing (L&C), a company specializing in software for supporting natural language processing especially in the medical field, and the Institute for Formal Ontology and Medical Information Science (IFOMIS), an academic research institution concerned with the theoretical foundations of ontology. In the course of this collaboration L&C's ontology, LinKBase®, which is designed to integrate and support reasoning across a plurality of external databases, has been subjected to a thorough auditing on the basis of the principles underlying IFOMIS's Basic Formal Ontology (BFO) [B. Smith, Basic Formal Ontology, 2002. http://ontology.buffalo.edu/bfo]. The goal is to transform a large terminology-based ontology into one with the ability to support reasoning applications. Our general procedure has been the implementation of a meta-ontological definition space in which the definitions of all the concepts and relations in LinKBase® are standardized in the framework of first-order logic. In this paper we describe how this principles-based standardization has led to a greater degree of internal coherence of the LinKBase® structure, and how it has facilitated the construction of mappings between external databases using LinKBase® as translation hub. We argue that the collaboration here described represents a new phase in the quest to solve the so-called ''Tower of Babel'' problem of ontology integration [F. Montayne, J. Flanagan, Formal Ontology: The Foundation for Natural Language Processing, 2003. http://www.landcglobal.com/].

* Corresponding author at: Institute for Formal Ontology and Medical Information Science, Universität des Saarlandes, Postfach 151150, D-66041 Saarbruecken, Germany. Tel.: +49 681 302 64 770.
E-mail address: jonathan.simon@ifomis.uni-leipzig.de (J. Simon).

1386-5056/$ — see front matter
doi:10.1016/j.ijmedinf.2005.07.015

1. Introduction: L&C's LinKBase® and Basic Formal Ontology

For millennia, when human beings have encountered difficulties in understanding reality, they have turned to philosophers for solutions. Why should we not do likewise today? The return to philosophy means in our context a return to those realist ideas rooted in the works of Aristotle and since subjected to some 2000 years of refinement. The return to philosophy however in no way requires that we abandon our pragmatic perspective. In his *Physics*, Aristotle writes, ''When the objects of an inquiry, in any department, have principles, conditions, or elements, it is through acquaintance with these that knowledge, that is to say *scientific knowledge*, is attained.'' Our thesis is that we would do well to keep these words in mind when we seek to design an adequate ontological inventory of those basic elements that belong to the structure of reality.

LinKBase® is a biomedical domain ontology that has been designed to integrate terminologies and databases with applications designed for natural language processing and information retrieval. The ontology contains 543 different relation types (linktypes), reflecting often subtle differences of meaning. They are divided into different groups, including spatial, temporal and process-related linktypes. In addition the ontology currently contains over 2,000,000 medical concepts with over 5,300,000 links instantiated between them. Such instantiated links are also called axioms. Both concepts and links are language independent, but they are cross-referenced to some 3,000,000 terms in various natural languages. LinKBase® provides a central hub of terms with fixed structured definitions into which various external medical terminologies and databases, such as Swiss-Prot, SNOMED®, and the Gene Ontology (GO), may be embedded. Such embedding is a complex task, not least because the different terminologies or databases that are to be integrated are often internally and mutually inconsistent. Our assumption, however, is that these terminologies must all essentially speak about the same reality, so that there must be a common thread that runs through them. The LinKBase® methodology is thus based on the idea that external terminologies and databases can be integrated on the basis of a sound understanding of the basic categorical distinctions that are common to them all.

Basic Formal Ontology is a philosophically inspired top-level ontology which has been developed in order to provide a coherent and unified representation of these basic categorial distinctions and which is currently being implemented as a top-level open source backbone ontology for LinKBase®. BFO is both an ontology in the sense of a partially ordered terminology system, and also a collection of first-order logical axioms providing additional information about the terms and relations represented in the terminology. The terminology of BFO is comparatively small, consisting of less than 100 concepts, but its axiomatics are quite sophisticated, representing the culmination of over 3 years of joint effort by a team of computer scientists, logicians, philosophers, informaticians and medical professionals [1]. BFO will provide a framework for mapping external ontologies, terminologies and databases onto LinKBase® in a way that is designed to provide for successful integration as well as to provide a useful guide for the development of future algorithms that allow for cross-ontology navigation.

Much of our work is closely related to that of the researchers associated with the Laboratory for Applied Ontology (LOA), based in Rome and Trento, who have published widely in this field, in the sense that both we and they operate on the belief that philosophical principles can aid informatics applications, in particular in natural language processing and data integration operations [2,3]. Our methods differ primarily in our philosophical foundations, but for many purposes these differences may be neglected [4].

2. Methods: term definition standardization

2.1. The purposes of standardization

As biomedical ontologies and terminologies have expanded they have become increasingly susceptible to a variety of different types of errors [5]. Many of these errors result from the lack of a coherent understanding of those basic categories and relations that structure our reality. The BFO formal ontology provides application ontologies with a set of standardized definitions for these fundamental ontological elements, definitions which can be exploited by reasoning applications, including applications designed for natural language understanding. By analysing and restructuring the ontologies underlying current applications, we can create principled formalizations that can improve coherence and adaptability both within and between ontologies and thus aid in the transmission of domain knowledge between users and software agents.

The resultant restructuring reflects an implementation of philosophical rigor along two dimensions. First, it establishes internal consistency on the basis of precise analyses of the top-level concepts and relations involved. Ontologies such as LinKBase®, SNOMED® and GO are understood as systems of concepts joined together by binary relations such as ''is-a'' and ''part-of''. The problem is that the top-level concepts and relations structuring these ontologies are specified for the most part only imprecisely via natural language formulations which lead to various ambiguities. Our project is then one of defining a common deep structure to which all of the top-level concepts and relations of the separate ontologies can be mapped on the basis of a sound conceptual analysis. Resultant standardization serves not only to link together external ontologies but also to identify and repair internal inconsistencies and ambiguities in LinKBase®.

The second dimension of rigor requires the use of the standard first-order logical language in which also the concepts of BFO are defined and axiomatized. By adopting this logical framework the rigor of the BFO classification system can be imported into an ontology from the outside. But this importation is meta-ontological, in the sense that changes are not made directly to the external ontology. Rather, their place in the BFO re-articulated domain ontology, in this case LinKBase®, is marked via an external mapping algorithm in a way that provides the degree of consistency required to navigate between different third-party ontologies such as GO and SNOMED®.

2.2. The specification of the standardization of LinKBase® terms

The provision of a deep structure of first-order logical definitions for concepts, relations and axioms is specified as follows:

(1) Each LinKBase® concept is associated with a class of individuals, namely the class of individuals to which the concept applies. Thus the concept ''Heart'' may be associated with the class of all individual hearts. We define a LinKBase® concept *C* by mapping it to a pair whose first element is a set containing the class with which *C* is associated, and whose second element is the set of all of the instances of that class.

(2) A LinKBase® relation is capable of joining any two LinKBase® concepts into an axiom. To define a LinKBase® relation *R* we associate it first with a relation R^* in the formal language of BFO. If the LinKBase® relation is ''part-of'' then its BFO associate will be the formal parthood relation. To complete the characterization of a LinKBase® relation, we employ schematic letters *X* and *Y* ranging over concepts. A LinKBase® relation such as ''part-of'' cannot be understood on its own, but is rather understood as setting out an axiom scheme ''*X* part-of *Y*'' (in general ''*X R Y*'', or $R(X,Y)$) which yields an axiom when *X* and *Y* are replaced by genuine LinKBase® concepts. A final detail involves adding, for each schematic concept letter, a binary parameter determining which member of the pair associated with that concept is to be considered. Thus for some relations the element will be the set containing all instances of the associated class, and for other relations the element will be the set containing the class itself (crucially, this parameter is fixed for each relation; the role of the parameter is not to conceal a contextual variable). The latter eventuality is rare, but an example is given in the section entitled ''Absences in LinKBase®'' below. The full definition of ''*X R Y*'' consists in a mapping to a logical formula of the form: 'For all *x* such that *x* is an element of the determined member of the pair associated with *X*, there is a *y* such that *y* is an element of the determined member of the pair associated with *Y*, and $R^*(x,y)$'.

(3) A LinKBase® axiom is a LinKBase® relation instantiated by two LinKBase® concepts. An axiom will thus be of the form ''*C R D*'', where *R* is a LinKBase® relation and *C* and *D* are LinKBase® concepts. Exploiting the associations articulated above, we may specify the definition of an axiom ''*C R D*'' as a logical formula of the form: 'For all *c* such that *c* is an element of one member of the pair associated with *C*, there is a *d* such that *d* is an element of one member of the pair associated with *D*, and $R^*(c,d)$'. For example, for the axiom ''Heart part-of Circulatory System'' the definition reads: 'for every individual heart there is some individual circulatory system of which it is a part.'

In the remainder of this essay we first summarize the ways in which the philosophical insights afforded by this standardization have allowed us to remove ambiguities and inconsistencies within the LinKBase® ontology itself. Second, we discuss how the standardization has assisted in the task of integrating external biomedical ontologies with LinKBase® and each other. To this end we shall describe the way that the standardization facilitated the development of the MaDBoKS ontology mapping software.

3. Results: part 1

3.1. Objects and processes in LinKBase®

In philosophical circles it is well understood that the universe contains two types of entities that relate differently to time. There are, on the one hand, objects: tables, chairs, countries, and people. These entities are said to *endure* through time, which means that they do not have temporal parts, but rather are wholly present at every moment in which they exist. On the other hand are processes: brain surgeries, heart attacks, lives. These entities are said to *perdure* through time, which means that they do have temporal parts, or phases, such as the first half of the surgery, the last phase of the heart attack, or one's childhood. This distinction is not always drawn clearly in natural linguistic practice. Thus there are ambiguous terms like 'injury', 'dilation', and 'dislocation', each of which corresponds to two distinct concepts. Thus we speak both of an injury as a perdurant (''when did that injury occur?'') and as an endurant (''That injury looks terrible''). Likewise with kinds of injuries, like dislocations: ''The dislocation of his shoulder occurred yesterday'' versus ''The doctor reduced the dislocation.'' Indeed, in the medical domain it is commonplace for a sort of process and the state resulting from that process to share the same name.

'Dilation' may stand for the process of dilation, i.e. of becoming broader: ''Once in place, a small balloon tip is inflated for a few seconds to *dilate* the artery.'' Or, it may stand for the dilated, broadened structure: ''*Dilation* of the posterior mitral ring was corrected.''

Here the philosophical distinction between endurants and perdurants allows us to maintain the separation of concepts which would otherwise be, and often are, conflated. By implementing this distinction into the LinKBase® top level, we have been able to recognize these instances of homonymy when they appear. We thereby avoid modeling errors that emerge in contexts where relationships between links and concepts are inferred.

3.2. Absences in LinKBase®

It is a tenet of contemporary philosophy that absences are not entities, but the lack of entities. Yet LinKBase® must represent natural medical language concepts like ''absence of bacteriuria (bacteria in the urine)'', and ''sputum without blood''. Further, while less common, medical texts may feature reference to absences without a specified location of absence, because the location is determined by context.

The straightforward approach, and the approach that LinKBase® formerly used, violated the philosophical tenet mentioned above, and construed these absences as special kinds of entities, namely processes of absence. With this approach, it was necessary to specify more about the processes in question. What kind of process is an absence? What is its duration? Who are its participants? How do we know when two descriptions of absences actually refer to the same entity?

Processes are perdurants, entities located in spacetime. They thus have boundaries, volumes, and locations (''the surgery took place in the operating room''). An adequate inference engine will know various things about bounded objects: it will know, for example, that if the boundary of object *x* is different from the boundary of object *y*, then *x* cannot be the same object as *y*. Now it is clear that in a natural language data extraction application, information about the boundary of an absence would be specified via a description like ''an absence *in the liver*.''

Philosophical scrutiny (one of whose functions is to test adaptability by demanding responses to creative counterexamples) tells us that the treatment of absences as processes is unstable, in that a reasoning engine attempting to handle and infer information about absences so construed runs the risk of deriving contradictions. This possibility arises when we wish to recognize the identities of differently described absences. 'The book was absent from my apartment' and 'The book was absent from my bedroom' seem to refer to the same absence. However, as soon as we instruct our inference engine to consider the two absences here described as identical, we will encounter inconsistency. For the system will record both that the absence has as boundary: my apartment, and that it has as boundary: my room. But this is a contradiction, since if *x* is the same as *y* then the boundary of *x* must be the same as the boundary of *y*. Thus if a treatment of absence concepts and relations in LinKBase® is to be perfectly general, and is not to rely on every absence concept coming with its own preset location, then we cannot construe absences as processes. So how do we treat them?

Another tenet of philosophy is: distinguish the particular from the universal (or class). When we say ''There is an absence of bacteria in the patient's urine'' we clearly are not saying *of* the bacteria in the urine, that *it* is not there. Rather, we are saying *of the universal*: bacteria in the urine, that *it* has no *instances* in the patient's urine. Following the intuition here, the current modeling eliminates concepts of absence themselves. Rather, relations of absence (like the absence of bacteria in the urine) are construed as links between the relevant bacteria concept, and the urine concept, but here it is the universal of the former that is involved: ''if x is the bacteria universal, and y is an instance of urine, then x has no instance located in y.'' This technique allows us to make inferences very naturally that would be artificial and error prone with the absences-as-entities model. We no longer need to answer the question of whether the absence of the book from my apartment is the same absence as that of the book from my room. We may naturally infer that there is an absence of the book from my room, given that there is an absence of the book from my apartment. This follows from our general knowledge of location and parthood.

Along with improving our reasoning power, this solution improves our representation structure, rendering applications involving absences more elegant and simple. The old representation of absences as processes blocked us from directly linking two entities where one entity is ''absent in'' the other entity, but rather a third concept had to be created, the process of ''absence of entity'' which related both entities. Thus, to represent the concept ''sputum without blood'' the concept ''absence of blood'' had to be created to be related to ''blood'' (the absent entity) and to ''sputum without blood'' (the location of the absence).

By representing absence as a relation between the ''absent entity'' and the ''entity from which the related entity is absent'' we avoid creating a third unneeded concept, and reduce the distance between the related concepts to one relation instead of two (e.g. the concept ''sputum without blood'' can be represented with a direct link to the concept ''blood'', which will be interpreted in formal language as: ''The blood universal has no instance located in (the patient's) sputum''). The distance between concepts, and between links, on parent child trees, is relevant to many LinKBase® applications [6].

4. Results: part 2

4.1. SNOMED® and the ''parthood'' relation

Identically named concepts and relations often have very different denotations. The degree of internal consistency required to apply the BFO standardization accurately to an ontology requires that these terms be disambiguated. One common variety of disagreement within a taxonomic system centers on divergent uses of the relation ''parthood.'' In SNOMED®, for example, the concept ''amputation of toe'' is a special case of the concept ''amputation of foot'' [7]. But while the toe certainly is a part of the foot, the amputation of the toe certainly is not an amputation of the foot. The former ought to be represented either as a part of an amputation of the foot, or alternatively, as an amputation of part of the foot. Depending on the context, these are two very different sorts of things [8].

SNOMED® here runs together endurants and occurrents. It runs together that element of parthood associated with the foot, an entity that endures in time, with that parthood associated with an amputation, an event that occurs in time. It is for reasons such as these that these two dimensions of parthood must be kept apart.

4.2. Endurants and perdurants within GO

GO is divided into three disjoint hierarchies: the *cellular component*, *biological processes*, and *molecular function* ontologies [9]. The first, equivalent to that of anatomy in the medical domain, is an ontology of endurants. It allows users to access the physical structure with which a gene or gene product is associated. A biological process, on the other hand, is defined in GO as ''a phenomenon marked by changes that lead to a particular result, mediated by one or more gene products.'' This ontology is therefore a hierarchy of occurrents.

There are, however, some confusions over the role of the molecular function hierarchy. While GO defines molecular function as ''the action characteristic of a gene product,'' it is clear that functions do not occur, but rather endure; the function of a gene or gene product exists identically for as long as its bearer exists and is present at all times, even if that function is never realized. Even mutant genes retain their function. Thus for example, ''signal transducer activity'' remains the function of the EPO_HUMAN protein even though the latter is incapable of performing the signal transduction process.

Molecular functions and biological processes are obviously closely related. The function ''signal transducer activity'' certainly *involves* performing ''signal transduction'' in some sense; yet in GO this relationship is undefined. The authors of GO have attempted to clarify this relationship, stating, ''a biological process is accomplished via one or more ordered assemblies of molecular functions,'' in order to suggest that the relation is one of agency. Here, functions *initiate* biological processes, but this would suggest that they share in a relation of parthood, which GO on the other hand explicitly rules out. For GO's authors insist, correctly in our view, that parthood only holds between entities of the same hierarchy. So long as the associated relations continue to conflate the distinct categories of function and process within the ontology, however, GO's architecture will continue to constrain the sorts of reasoning systems which it can support.

4.3. Mapping ontological elements: applying external consistency

The Mapping Databases onto Knowledge Systems tool (or MaDBoKS) is an extension of the LinkFactory® ontology management system that administers and generates mappings from external (relational) databases onto LinKBase® [10]. This mapping mediates the data contained in the external database in a manner that simulates an expansion of the hub ontology, leaving the structure of the foreign system (and LinKBase® itself) untouched. The MaDBoKS system creates a two-fold mapping from a database onto an extension of LinKBase®. First, ''column data'' is mapped, in the sense that column headings (of the external database) are represented as concepts in the expanded ontology, and implicit relationships between column headings are represented as links between these concepts. Additionally, ''cell record data'' is mapped, in the sense that the row/column intersection fields of the database are mapped to concepts in the extended LinKBase®. If preexisting concepts of the relevant sort exist then the new entries are subsumed there, otherwise they are subsumed under the entries added from the column mapping operation.

The mapping of cell record data is important when more than one database is simultaneously mapped into LinKBase®. It then allows for the identification of similar, corresponding records across databases. When the data in a given cell is mapped to a preexisting concept in the ontology (or a ''virtual concept'' introduced by a previous mapping of another database), this new concept ''inherits'' all of the information already assigned to the preexisting concept (terms, links and definitions) which may not have been present in the database from which the cell was mapped. The same procedure, founded in part on the BFO definition standardization, renders the MaDBoKS system a facilitator of translation between databases.

4.4. Mapping SNOMED® to LinKBase®

LinKBase® understands not only the relation of ''parthood'', but also ''proper parthood'', ''part-of'', ''part-for'' and ''has-part''. These refinements allow us to build an accurate representation in which various distinctions in the conception of ''amputation of foot'' discussed earlier are recognized as distinct and their relation to each other can be mapped. The distinctions rest on the formal notion of parthood, along with an understanding of the interplay of classes and their instances crucial to the modeling of this formal notion and its relatives. Class X is part-of Class Y whenever every instance of Class X is a part of some instance of Class Y. Class Y has-part Class X whenever every element of Class Y has some element of Class X as a part. Class X is part-for Class Y whenever Class X is part-of Class Y, and Class Y has-part Class X. The further distinction between parts and proper parts lies on the instance level: individual x is a proper part of individual y whenever x is a part of y, but x is not identical to y. Where the toe is both a part *and* a proper part of the foot, the foot is a part, and not a proper part, of itself. In LinKBase®, these distinct parthood relations are captured, with ''part-of'' as the root relation. Further, LinKBase® contains a concept named 'structure', designed to be relativized to embed information about parthood in the concept space, as well as in the relation space. If X is a class, then there is a concept ''X structures'' which is such that it subsumes all and only those classes that stand in the part of relation to X. For example, both the toe and the foot itself are subsumed by the concept ''foot structures.''

This configuration is then mapped to the SNOMED® ontology, where ''amputation of foot'' is related to the concept ''foot structure'' (any part of the foot including the foot itself), which subsumes two further concepts ''complete amputation of foot'' and ''partial amputation of foot'' (related to the concept ''proper part of foot''). In this way we maintain a hierarchical structure that subsumes both the toe and the foot without reducing either one to the other, thus allowing each to be related to different, and possibly incommensurable concepts without the problematic inconsistencies derived through inherited criteria.

4.5. Mapping GO to LinKBase®

During the conceptual analysis phase, we carefully investigated the top-layer concepts of the three GO sub-domains that act as our gateway between the LinKBase® concepts and GO terms. We identified the more general concepts of GO in LinKBase® and created new concepts in those cases where suitable equivalents were not already recognized. In this way we are able to relate GO's molecular function hierarchy to the two other GO hierarchies by integrating all three simultaneously into the expansion of BFO motivated by the formal-ontologically extended top level.

If we return to the EPO_HUMAN protein example from earlier, we see now that LinKBase® is able to appropriate this example and model the relations with a greater degree of clarity, essentially mirroring the BFO defined structure. The connection between a GO protein and its activity in LinKBase® is captured by a ''has-function'' relation, and the connection between an activity and its corresponding processes is captured by the LinKBase® ''realization'' relation. The former reflects the relation between a substance and its function, and the latter, that between a function and its actualization. Clearly, this latter relation is skew to the whole/part relation, which is properly left exclusive to each hierarchy.

In this manner not only is GO consistently mapped to LinKBase®, but the expressiveness of GO itself has been expanded without any major alterations required in its core structure.

5. Discussion

Our LinKBase® ontology is a representation of the medical domain. By mapping more specialized information sources like GO and protein databases, we were able very quickly to expand the reach of our ontology and hence achieve a database warehousing system within which all mapped databases stand automatically in the right sort of relation to each other in such a way that a global view of the dispersed information is made possible. The MaDBokS system can be used to graft databases onto the ontology and thereby make the latter useable for a variety of applications. The flexibility of the MaDBoKS system and the speed with which databases can be integrated allows the prototyping of different integration protocols in relation to different sets of databases and hence enables a fine-tuning of the integration process for specific applications such as data-mining and information extraction.

The BFO-driven restructuring of LinKBase® is still in its infancy, yet we already have examples demonstrating increased adaptability through the application of philosophical knowledge and techniques. We have demonstrated examples in which changes were made leading to an enhanced internal consistency, allowing the level of access necessary for a general database translation hub.

If early successes (like the integration of GO into a MaDBoKS extension of LinKBase®) are any indicator, we have great reason to expect that the thoroughgoing integration of BFO and LinKBase®, of which the above results are preliminary groundwork, will greatly enhance the capacity of LinKBase® to effect direct integration between foreign ontologies such as SNOMED® and GO. For the results cited here are not isolated instances but rather illustrations of a general pattern. There are reasons for the ad hoc features of many biomedical ontologies, the main cause of the so-called ''Tower of Babel'' problem of interoperability. These features have developed because ontologists and terminologists were forced, in moving from printed dictionaries and nomenclatures to digital systems, to make a series of uninformed decisions about complex ontological issues, indeed about the very same issues that philosophers have been pondering for millennia. To date, the importance of philosophical scrutiny in software application ontologies has often been obscured by the temptation to seek immediate solutions to localized problems. In this way the forest is lost for the trees, and larger integration problems are rendered unsolvable. Ad hoc solutions foster further ad hoc problems.

It is thus a tangled web we weave when we seek to create application ontologies without a basis in philosophically sound formal theories. The philosophically sound formalism of LinKBase® enables it to support the integration (and thereby, the untangling) of data from different external data sources in a transparent way, capturing the exact intended semantics of the database terms, and filtering out erroneous synonyms.

Acknowledgements

We are grateful for the helpful comments of Dirk Siebert, Werner Ceusters Jean-Luc Verschelde, and three anonymous reviewers. This work has been supported by the Language and Computing Research Division, the Wolfgang Paul Program of the Alexander von Humboldt Foundation, the Network of Excellence in Semantic Interoperability and Datamining in Biomedicine of the European Union, and the project Forms of Life sponsored by the Volkswagen Foundation.

References

[1] http://www.ifomis.org.
[2] N. Guarino, C. Welty, An overview of OntoClean, in: S. Staab, R. Studer (Eds.), Handbook on Ontologies, Springer-Verlag, 2004.
[3] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, L. Schneider, Sweetening ontologies with DOLCE, in: Proceedings of the EKAW 2002, Siguenca, Spain, 2002.
[4] A. Kumar, B. Smith, D. Pisanelli, A. Gangemi, M. Stefanelli, An ontological framework for the implementation of clini-

cal guidelines in health care organizations, in: D. Pisanelli (Ed.), Ontologies in Medicine, Proceedings of the Workshop on Medical Ontologies, Rome, October 2003, IOS Press, Amsterdam, 2004.

[5] W. Ceusters, B. Smith, A. Kumar, C. Dhaen, Mistakes in medical ontologies: where do they come from and how can they be detected? in: D. Pisanelli (Ed.), Ontologies in Medicine, Proceedings of the Workshop on Medical Ontologies, Rome, October 2003, IOS Press, Amsterdam, 2004.

[6] L. Van Geyt, P. Martens, B. Terzic, J. Flanagan, Get More Out of Your Unstructured Medical Documents, 2002. http://www.landcglobal.com/.

[7] SNOMED® (Systematized Nomenclature for Medicine). http://www.snomed.org/.

[8] B. Smith, Mereotopology: a theory of parts and boundaries Data Knowl. Eng. 20 (1996) 287—303.

[9] GO (Gene Ontology General Documentation). http://www.geneontology.org/doc/GO.doc.html.

[10] J.L. Verschelde, M. Dos Santos, T. Deray, B. Smith, W. Ceusters, Ontology-assisted database integration to support natural language processing and biomedical data-mining, J. Integr. Bioinform., in press.