Ten Theses on Clinical Ontologies

Stefan Schulz and Holger Stenzhorn
Department of Medical Informatics, Freiburg University Hospital, Germany

Abstract: We present ten principles for clinical ontologies that describe the authors' opinion about what should be understood by the notion of clinical ontologies and what not. In contrast to clinical terminology systems, clinical ontologies are considered to be semantic reference systems and for that – first of all – strive to account for the properties of the domain entities themselves and their proper formal definitions – rather than just linking clinical terms together.

Keywords: Clinical Ontologies, Knowledge Representation

Introduction

Issues concerning the representation of knowledge in clinical contexts have been intensively discussed in the Medical Informatics community over the last two decades [1]. The challenge to provide semantic reference has thus been met by a variety of different systems of classifications, terminologies and ontologies, covering the broad range of clinical disciplines and also outreaching to the realm of molecular medicine and genomics. The impressive growth of the Unified Medical Language system (UMLS) [2] and the development of the Open Biomedical Ontologies (OBO) [3] give particular witness to this effort. However, most of these terminologies and ontologies have been created in purely application- and purpose-driven contexts and are therefore not geared towards semantic interoperability. The latter, however, constitutes a pressing requirement to the integrated care paradigm where patient data should be easily interchangeable across institutional boundaries.

This lack of interoperability is – in our opinion – at least partly due to a lack of agreement on what terminology systems are actually supposed to encode. So far, there is only a vague understanding on what the notions of "terms", "classes" and "concepts" are really referring to in current terminology systems.

Only during the last couple of years, a more principled discussion about some of the more fundamental issues has been under way, mainly driven by philosophers and computer scientists. Although this discussion has brought about some controversy, it has fertilized and ameliorated the way biomedical terminology systems have been conceived, built, maintained and used.

The main purpose of this paper is now to focus on the role of biomedical ontologies as representational artifacts being a special kind of terminology systems. Aware of the difficulties, the use of the notions "ontology" and "terminology system" can cause (for either, different and contradicting definitions can be found in literature [4]), this paper tries to streamline and clarify the notion of "ontology" in the context of integrated clinical care.

Ten Principles

The nature, purposes and limitations of clinical ontologies are presented here as a sequence of ten principle. They represent the current view of the authors and are being formulated to stimulate a clarifying discussion. They neither claim to represent a consensus among medical informaticians or terminologists, nor do they constitute a fully consistent system of thinking.

I. Terminology systems provide semantic reference

It is generally accepted that terminology systems (in a broad sense, cf. [5]) should provide some kind of semantic reference. This means that they support the relation of term meanings. Terms are the entities of language in a given domain, they may be simple or complex words, as well as multi-word strings. Terminology systems relate terms that share the same meaning (synonymy), broader and narrower, as well as related meanings. It should be emphasized here that terminologies are term-centered, i.e. that terms constitute their basic elements.

II. Ontologies are semantic type hierarchies to support organizing domain entities

Among the numerous definitions of what an ontology constitutes, the following definition is preferred for our context: Ontologies are representational artifacts whose representational units are intended to designate classes or types in reality and to relate them to each other [6]. It is important to highlight here that the single entities represented in an ontology are not terms. Nor are concrete objects, i.e. individuals (instances, particulars) represented in the ontology proper. However, one of the main purposes of ontology is to provide a means to classify exactly those entities by defining and organizing their semantic type.

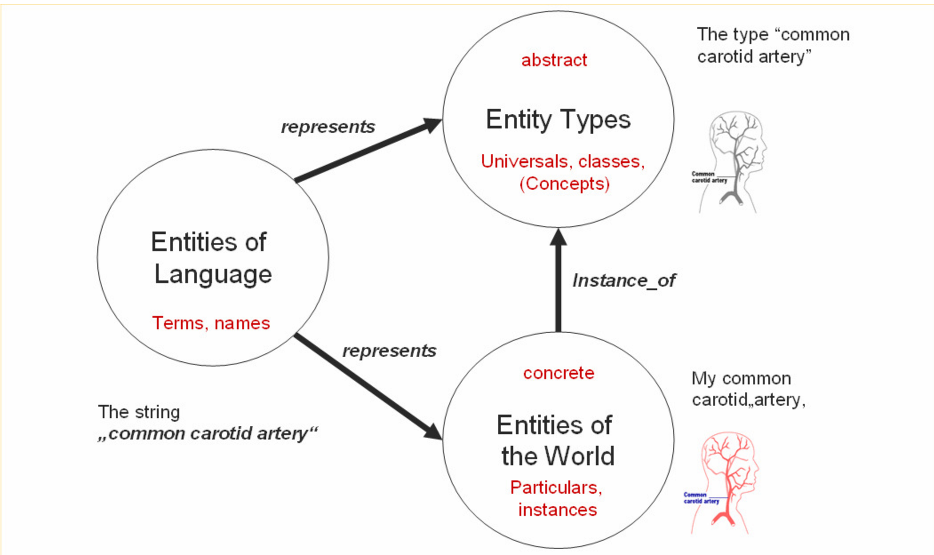

Fig. 1: World entities, their types and terms denoting them

The main classification principle of ontologies is the taxonomic order. Taxonomies relate types with their superordinate types. This hierarchy-forming relation is generally named *is_a*. As an example, the type "artery" is a superordinate one of "carotid artery" and the latter stands in turn to "common carotid artery" in a supertype relationship. Note that *is_a* must not be mistaken for the relation *instance_of*, which relates a concrete object to its type. Although it usual to state in natural language that "Tibble *is a* cat.", the individual entity with the name "Tibble" is related to the entity type cat by the relation *instance_of* (and hence the sentence would be "Tibble *is an instance* of the type cat.").

Fig. 1 depicts the relations between entities, their types and their denoting terms.

III. Ontologies represent universal truths

One of the major misconceptions about ontologies is that they are directly suited to represent large parts of clinical and scientific knowledge, i.e. the symptoms of a disease, the probability of a certain risk or the side effects of some drugs. But on the contrary, this is not the case: (Formal) ontologies should only represent what is assumed to be universally true, i.e. that it always the case that an artery is a blood vessel or that human blood contains erythrocytes.

Formally, ontologies provide and combine universal statements about all instances of a given type such as:

- All Xs are Ys
- For all Xs, there is some Z that…

For example, all instances of the type carotid artery are instances of the type artery. Or, for all instances of the type common carotid there is some instance of the type aorta which it is connected to. So it becomes obvious that ontologies are not the right place to represent probabilistic, vague or uncertain knowledge. An important corollary of these universality assumptions is the well-known principle of inheritance. Each property defined for some type in an ontology is inherited by all of its subtypes, as well as of its instances (and the instances of the subtypes). Wherever this principle does not seem to hold there is evidence of a major misunderstanding of what an ontology is supposed to be. For instance, one could add a prevalence value to disease type such as "0.5%" to the disease type "Schizophrenia". It is obvious that the same value cannot be inherited by its subtype "Bipolar Schizophrenia". But this is not an exception to the inheritance rule. The error is rather to consider prevalence as an inherent property of a disease. In contradistinction, prevalence is a property of a population with regard to a disease but not of the disease proper.

IV. Ontology types extend to classes of world entities

Types in ontologies are often referred to as classes. But as Fig. 1 depicts, classes are indeed different from types: A class is the collection of all entities in the world that instantiate a certain type. But therefore, classes and types are actually closely related. The class of entities is also considered to be the extension of a given type. In contrast to mathematical sets, classes can vary in respect to their members across time. For example, the class of E.coli (i.e. the extension of the type E.coli) remains the same although it constantly gains and loses instances. Due to the direct dependence between classes and types, usually the just described distinction between the two is not made in practical ontology engineering.

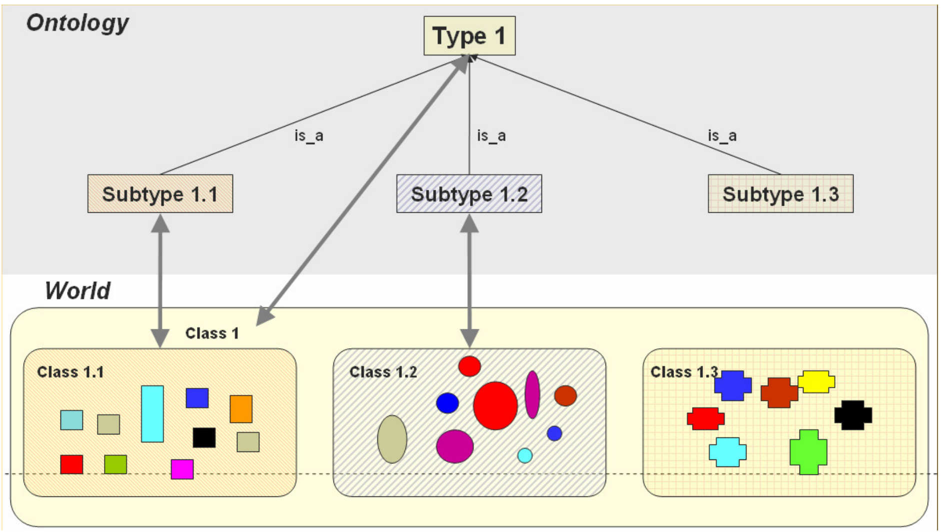

Fig. 2: Types (above), Classes (below), Instances (rectangles, ellipses, polygons)

V. Ontologies organize individual entities – not concepts

Whereas the use of the notions "type" and "class" is more recent in medical informatics, the notion of "concept" has a much longer tradition. The problem with the latter is that it is applied in several different senses [7]. On the one hand "concept" is meant to stand for "entities of thought" which is mostly considered to represent word meanings, such as discussed in the first section and is, therefore, characteristic for language-centered terminology systems. On the other hand it is also used as a synonym of what has been introduced above as "type". Finally, from a computer science point of view, "concept" often stands for unary predicates in knowledge representation languages such as OWL-DL [8]. For the sake of clarity we avoid the use of the word "concept" when referring to ontologies in the sense we have defined them.

VI. Ontologies represent what is – information models represent what we know about

The task of representing clinical facts requires that the following two aspects are kept strictly separated.

1. Universal truths about entities of the world as referred to by domain terms
2. Known facts about concrete clinical cases.

Whereas the first task corresponds to what is understood by ontologies, the second one is to be embarked upon by information models. It involves not only the facts as they are in the world but also the knowledge about these facts. In terminology systems as currently used, these epistemological aspects are often mixed up [9]. In ICD 10, for example, there is a distinction between the classes "*Tuberculosis of lung, confirmed by culture only*" and "*Tuberculosis of lung, confirmed histologically*". This reveals the difference between ontology and epistemology: The very nature of a tuberculosis in a patient does not depend on how it is diagnosed. Nevertheless, for clinical reasoning (in which the certainty of a diagnosis matters) the source of knowledge a diagnostic statement is based upon is of utmost importance. Another example is how to encode the sex of a patient. Whereas in an information model it makes perfect sense to discriminate between "male", "female" and "unknown", this distinction is nonsensical on the level of clinical ontologies because every patient has a sex, which might be known or unknown but obviously does not change according to the fact of it being known or unknown. The relation between clinical information models and ontologies has recently received an increased attention in the context of clinical archetypes, openEHR, HL-7 Version 3 and SNOMED CT [10].

VII. Practical requirements may justify controlled deviations from the "true path"

In many cases, user requirements for clinical application ontologies can make it difficult to fully follow the "pure doctrine" of ontology design, since the restriction to universal truth obviates the representation of many shared assumptions that are important in a clinical context. For instance, it may be required to classify certain health related states as risk factors for diseases. If we declare "Hypertension" as a "Risk factor for aneurysm rupture" in an ontology using an *is_a* link, we certainly go beyond the representation of universal truths, since not every hypertension causes a myocardial infarction, and "Risk factor" is rather a role ascribed by humans than a universal type. Mistaking roles for subtype relations, also called is_a overloading is a common error in ontology design [11]. For practical reasons such a routine may be, however, justified. Nevertheless it is recommended that such assertions be strictly separated from the ontology proper. In the @neurist ontology [12] this dichotomy has been made explicit by introducing a separate branch with the root node "Particular in context", as depicted in Fig. 3.

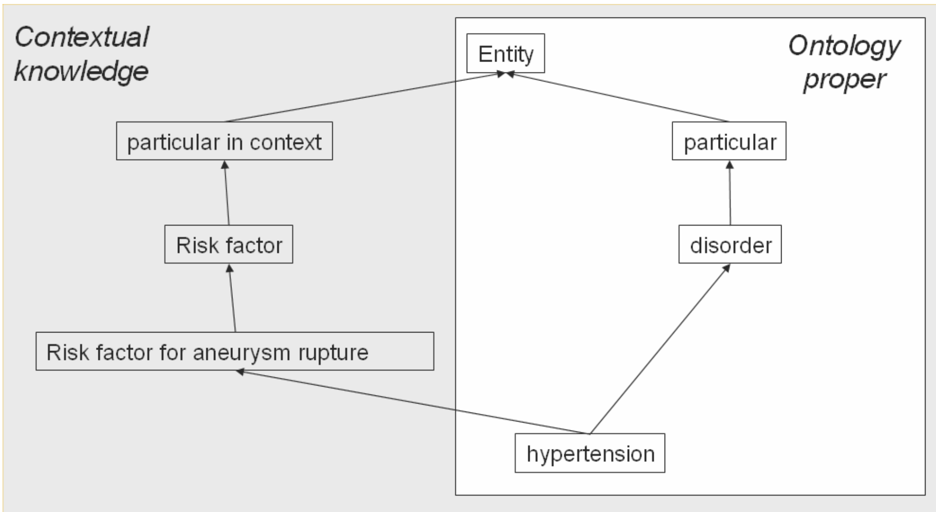

Fig. 3 Epistemological knowledge as separate branch in the @neurist ontology

VIII. Ontologies need to be linked to dictionaries

Ontologies in a strict sense (i.e. as understood as taxonomies of types) do not incorporate any lexical or terminological information at all. Still the naming of the ontology nodes should be self-explanatory and employ terms commonly used in the domain. However ambiguous formulations that regularly exist in the domain should be avoided. Furthermore, special characters should be avoided and the naming conventions of the underlying language must be followed. This means that ontology labels are not to be mistaken for actual terms (even if they are named "terms" in several ontologies). Since each node can have one name only, synonyms cannot (and should not) be managed at this level. So it is necessary to provide a link between an ontology and a dictionary where each dictionary entry corresponds to a domain term and is linked to one (in the case of polysemous terms, to two or more) nodes in the ontology. Synonyms are linked to the same node. Such a dictionary is a separate data structure and it is not an integral part of the ontology itself.

IX. Ontology users need not to see the whole "engine"

From the user perspective, the internal structure of an ontology can easily lead to confusion and misunderstandings. As users are used to browse simple hierarchical trees from top to bottom, they may be overwhelmed by the complexity of the structural relations between all the types of the ontology. They may also get puzzled by the terms characterizing the upper level, such as "dependent continuant", "quality region" or "fiat object part" as provided by upper level ontologies such as BFO [13,14] or DOLCE [15]. Still, such upper ontologies are necessary for the organization of the ontology into well-defined categories, necessary for enforcing consistency of its content. However, these categories do not correspond to terms commonly used in the domain and could therefore – at a first sight – deemed superfluous by ontology users. It is therefore neither necessary nor desirable that the whole internal "machinery" be visible for ontology users (i.e. system developers and domain experts that link to ontologies when creating clinical information models, entry forms, etc.). They should only see those parts of the ontology they need for their respective work and be provided with customized user interface to support them with their tasks.

X. Ontologies should provide tailored terminology services

The concept of terminology server / service has become popular in the nineties where specifications and use cases have been formulated [16,17,18]. Under the new viewpoint of clinical ontologies this concept should be taken up and adapted to the current requirements. End users should benefit from tailored ontology based terminology services without the need to any access to the actual ontology (cf. the last section). End users just create and select terms. The terminology service in turn then takes care of linking these terms to the ontology and of the provision of controlled terms to the user (cf. Fig. 4).

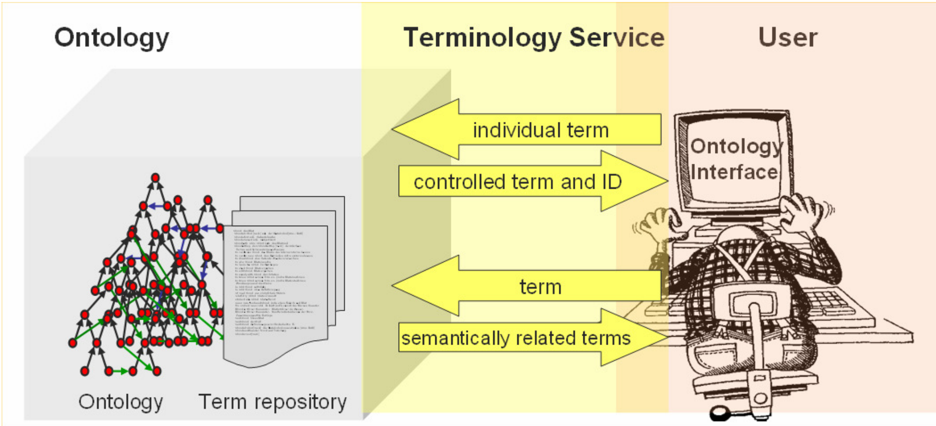

Fig. 4 Terminology services mediating between the ontology and the end user

Conclusion

In this paper we presented ten principles which we believe can clarify what (formal) clinical ontologies are, what they are not and how they should be used. These principles should be applied to any artifact using formal means for representing world entities in the domain of interest, thus providing a logically and philosophically founded basis for the meaning of clinical terms. This principle should be applied especially to SNOMED Clinical Terms (CT) [19] as an emerging terminology covering whole range of clinical medicine. SNOMED CT has still a long way to go in order to fulfill common ontological standards [20]. To this end the ten principles presented here could be useful cornerstones and guidelines.

Acknowledgements

This work was carried out in the framework of the @neurIST Integrated Project, which is co-financed by the European Commission through the contract No. IST-027703 (http://www.aneurist.org).

Address of Corresponding Author

Stefan Schulz, Department of Medical Informatics, Freiburg University Hospital, Stefan-Meier-Str. 26, 79104 Freiburg, Germany. stschulz@uni-freiburg.de

References

1. J. J. Cimino. Desiderata for controlled medical vocabularies in the twenty-first century. Methods of Information in Medicine, 37(4/5):394–403, 1998.
2. UMLS. Unified Medical Language System. Bethesda, MD: National Library of Medicine, 2007.
3. Open Biological Ontologies (OBO) [http://obo.sourceforge.net], 2007. Last accessed March 6th, 2007.

4. W. Kuśnierczyk. Nontological Engineering. FOIS 2006 – Proceedings of the 4th Intl Conf on Formal Ontology in Information Systems; 2006. p. 39-50.
5. R. Cornet R, N. F. De Keizer, A. Abu-Hanna. A Framework for Characterizing Terminological Systems. Methods of Information in Medicine, 2006; 45: 253-266.
6. S. Schulz and I. Johansson. Continua in biological systems. The Monist, 2007, accepted for publication.
7. B. Smith, "Beyond Concepts, or: Ontology as Reality Representation", In: A. C. Varzi and L. Vieu, editors, Formal Ontology in Information Systems. Proceedings of the 3rd International. Amsterdam: IOS Press, 2004, 73–84.
8. I. Horrocks, P. F. Patel-Schneider and F. van Harmelen, "From SHIQ and RDF to OWL: The making of a Web Ontology Language", Journal of Web Semantics, 1(1), 7–26, 2003.
9. O: Bodenreider, B: Smith, and A: Burgun. The ontology-epistemology divide: A case study in medical terminology. In Achille C. Varzi and Laure Vieu, editors, *Formal Ontology in Information Systems. Proceedings of the 3rd International Conference - FOIS 2004*, pages 185–195. Amsterdam etc.: IOS Press, 2004 .
10. A. Rector, R. Qamar, T. Marley. Binding Ontologies & Coding Systems to Electronic Health Records and Messages. In: Bodenreider O, editor. Formal Biomedical Knowledge Representation (KR-MED 2006) CEUR; 2006. p. 11-19.
11. N. Guarino and C. A. Welty. An overview of ONTOCLEAN. In Steffen Staab and Rudi Studer, editors, Handbook on Ontologies, International Handbooks on Information Systems, pages 151–171. Berlin: Springer, 2004.
12. S. Hanser, M. Boeker, K. Kumpf, P. Bijlenga, S. Schulz. Design of an Ontology on Cerebral Aneurysms: Representing the Conceptual Space of the @neurIST Project. Medinfo 2007 Congress, 20-24 August 2007, Brisbane, Australia. Accepted for publication.
13. P. Grenon and B. Smith. SNAP and SPAN. Towards dynamic spatial ontology. Spatial Cognition and Computation, 4:69–103, 2004.
14. P. Grenon, B. Smith, and L. Goldberg. Biodynamic ontology: applying BFO in the biomedical domain. In D. Pisanelli, editor, Studies in health technology and informatics, Vol. 102, pages 20–38, 2004.
15. A. Gangemi, N. Guarino, C. Masolo, and A. Oltramari. Sweetening ontologies with DOLCE. In A. Gómez-Pérez and V. R. Benjamins, editors, Proceedings of the 13th International Conference – EKAW 2002, volume 2473 of Lecture Notes in Artificial Intelligence, pages 166–181. Berlin: Springer, 2002.
16. A. L. Rector, W. D. Solomon, W. A. Nowlan, and T. Rush. A terminology server for medical language and medical information systems. Methods of Information in Medicine, 34(1/2):147–157, 1995.
17. C. G. Chute, P. L. Elkin, D. D. Sheretz, and M. S. Tuttle. Desiderata for a clinical terminology server. In N. M. Lorenzi, editor, AMIA'99 – Proceedings of the 1999 Annual Symposium of the American Medical Informatics Association. Transforming Health Care through Informatics: Cornerstones for a New Information Management Paradigm, pages 42–46. Washington, D.C., November 6-10, 1999. Philadelphia, PA: Hanley & Belfus, 1999.
18. J. Ingenerf and T. Diedrich. Notwendigkeit und Funktionalität eines Terminologieservers in der Medizin. Künstliche Intelligenz, 11(3):6–14, 1997.
19. SNOMED Clinical Terms. Northfield, IL: College of American Pathologists, 2007.
20. S. Schulz, B. Suntisrivaraporn, F. Baader. SNOMED CT's Problem List: Ontologists' and Logicians' Therapy Suggestions. Medinfo 2007 Congress, 20-24 August 2007, Brisbane, Australia. Accepted for publication.