



MIE 2005 Tutorial #35

Title: Ontology Design

List of instructors: Barry Smith, Institute for Formal Ontology and Medical Information Science, Saarland University, Saarbrücken, Germany and Department of Philosophy, University at Buffalo

Werner Ceusters, European Centre for Ontological Research, Saarland University, Saarbrücken, Germany

Date, time and place: Sunday, 28 August 2005, from 9am to 12noon, room 2150

Abstract: With the development of modern formal disciplines (formal logic, and the computational disciplines which have arisen in its wake) we have learned a great deal about the criteria which must be satisfied if an ontology is to be designed in such a way that the information expressed by its means can be extracted via automatic procedures in a maximally effective way. Unfortunately, existing biomedical ontologies have been developed in large part without concern for these design criteria. Both the classes they contain and also the relations between these classes (including *is_a*, *part_of*, *located_in*, *derives_from*, and so forth) are poorly defined. Also the rules for formulating definitions are themselves inadequately formulated. Moreover the organization of the ontologies as a whole leaves much to be desired, and too little effort is devoted to the design of ontologies in such a way that compatibility with other ontologies will be assured.

In this tutorial we present a methodology designed to confront these problems that has been developed and tested by IFOMIS, the Digital Anatomist Group and the Open Biological Ontologies consortium, and which is currently being applied in a series of biological and medical domains. We first explain the basics of the approach, and then demonstrate how it been successfully applied thus far in areas such as anatomy and embryology. Finally we show how electronic health records can be integrated with ontologies built in this way and illustrate prototype applications that show some of the reasoning power of the resultant system.

Description: This tutorial is part of an on-going series organized under the auspices of OBO and other bodies, which is designed to enhance awareness, among those involved in ontology development in different areas of the life sciences, of current developments and best practices in ontology. These workshops serve the goal of creating the conditions under which ontologies can be developed which are marked by high a degree of usability, reliability and interoperability.

The tutorial is divided into four parts, which can be briefly summarized as follows:

1. Realism as a basis for ontology design: work on biomedical ontologies and terminologies hitherto has been dominated by a top-down methodology based on (often poorly defined) relations between concepts. We shall present a new methodology for ontology design which starts not from concepts but from individuals as they are related together in reality. Examples of individuals are: my heart and my blood pressure, entities that are referred to in my medical record when I consult a cardiologist. In Part One of the tutorial we explain the foundations of this bottom-up methodology and show in what ways it yields a new type of ontology design.
2. Ontology and anatomy: in recent years, much work has been done on constructing formal theories that support reasoning about qualitative spatial relations among individuals. A mereology is a formal theory of parthood and of relations such as overlap (having a common part) and discreteness (having no common part) defined in terms of parthood. Since such relations apply directly to concrete individuals and require neither quantitative data nor mathematical abstractions (points, lines, etc.), a mereology is a natural basis for qualitative spatial reasoning in medicine. In medical contexts, of course, a more complicated form of qualitative spatial reasoning – reasoning about relations among classes of individuals – is also common. In canonical anatomy for example we find assertions such as "the stomach is continuous with the esophagus", "the right ventricle is part of the heart" or "the brain is contained in the cranial cavity". It is important to distinguish these sorts of assertions from claims about relations among individuals (e.g. "patient X's right ventricle is part of patient X's heart" or "my stomach is continuous with my esophagus"). In Part Two of the tutorial, we explain how class-level relations can be defined formally – in accordance with our bottom-up methodology – in terms of relations among individuals. We demonstrate that different versions of the class relations have significantly different logical properties. (The failure to distinguish between these different versions has led to errors in existing systems.) We show how precise and consistent characterizations of these relations would improve the clarity of the information embodied in ontologies such as GALEN and FMA, and how they lead to more reliable coding and to stronger automated reasoning capabilities. Consistency leads also to enhanced interoperability of the ontologies which result, including interoperability which crosses granularities, for example from the molecule to the cell or organ.
3. Realism and biological databases: in Part Three of the tutorial, we advance a suite of ten relations (including `is_a` and `part_of`) which have been adopted for use in the

construction and maintenance of OBO (Open Biological Ontologies) and similar ontologies in the future. Each relation is provided with a formal definition that is designed to establish the meaning of the corresponding relational expression in an unambiguous way, and thus to assist the users and compilers of biological ontologies in avoiding errors in coding and annotation by providing them with a more coherent understanding of both the relations and the relata which such ontologies involve. The resulting framework is designed to enhance usability and interoperability of ontologies in the life sciences, and also to support new types of automated reasoning with biological data, including reasoning about the spatial and temporal dimensions of biological phenomena. We show how the relations can be used to integrate ontologies at different levels of granularity, for example in such a way as to provide a unified treatment of phenomena such as embryological development and tumor growth.

4. Practical implementations of realism-based ontologies: current Electronic Health Records (EHRs) are organized around two kinds of statements: those reporting observations made, and those reporting acts performed. In neither case does the record involve any direct reference to what such statements are actually about. They record not: what is happening on the side of the patient, but rather: what is said about what is happening. We show how the ontology design methodology described above supports the move to a new type of EHR regime in which all the particulars to which reference is made in clinical statements are uniquely identified. This will allow us to achieve interoperability among different systems of records at the level where it really matters: in regard to what is happening in the real world. It will allow us to keep track of particular disorders and of the effects of particular treatments in a precise and unambiguous way. And, with the help of our rigorous definitions of the corresponding ontological relations, it will allow us to engage in new types of reasoning and error checking in relation to the data encoded, at the level of both particulars and general classes. In Part Four of the tutorial we will show a prototype implementation of an EHR/terminology system conforming to our methodology for ontology design, focusing on how such an implementation can be used to verify data entry in the EHR, to reason with the data, and to use the total EHR/terminology system for statistical and other purposes.

Who should attend ? This tutorial does not require any prior knowledge of ontology, though some familiarity with these topics will make it easier to understand the deeper issues involved.

Attendees who might find this tutorial worthwhile include:

developers and users of biomedical ontologies, developers and users of electronic patient record systems (including those focusing on terminology services), physicians interested in the possibilities of modern ontologies.

All participants will receive from their attendance in this tutorial hands-on training in ontology design and in the formulation and use of simple logically clear definitions.