Ontologies and Medical Terminologies

James Geller
New Jersey Institute of Technology, USA

INTRODUCTION

Ontologies

The term "Ontology" was popularized in Computer Science by Thomas Gruber at the Stanford Knowledge Systems Lab (KSL). Gruber's highly influential papers defined an ontology as "an explicit specification of a conceptualization." (Gruber, 1992; Gruber 1993). Gruber cited a conceptualization as being "the objects and concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them." (Genesereth & Nilsson, 1987). The term "Ontology" has been used in computer science at least since (Neches, 1991), but is derived from philosophy where it defines a "systematic account of existence," usually contrasted with "Epistemology."

Gruber's work is firmly grounded in Knowledge Representation and Artificial Intelligence research going back to McCarthy and Hayes classical paper (McCarthy & Hayes, 1969). Gruber's work also builds on frame systems (Minsky, 1975; Fikes and Kehler, 1985) which have their roots in Semantic Networks, pioneered by (Quillian, 1968) and popularized through the successful and widespread KL-ONE family (Brachman & Schmolze, 1985). One can argue that Gruber's ontologies are structurally very close to previous work in frame-based knowledge representation systems. However, Gruber focused on the notion of knowledge sharing which was a popular topic at KSL around the same time, especially in the form of the Knowledge Interchange Format (KIF) (Genesereth, 1991).

Ontologies have recently moved center stage in Computer Science as they are a major ingredient of the Semantic Web (Berners-Lee et al., 2001), the next generation of the World-Wide Web. Ontologies have also been used in Data Mining (see below) and in (database) schema integration.

Medical Terminologies

In the field of Medical Informatics a rich set of Medical Terminologies has been developed over the past twenty years. Many of these terminologies have as their backbone a taxonomy of concepts and IS-A (subclass) relationships. This IS-A hierarchy was pioneered in the semantic networks and frame systems mentioned above. With this structural commonality of ontologies and Medical Terminologies in mind, we will treat both kinds of knowledge representation systems together. Some of the largest existing ontologies have been developed in Medical Informatics, which makes this field especially interesting. For example, the Unified Medical Language System (UMLS; Humphreys et al., 1998) Metathesaurus contains information about over 1.5 million biomedical concepts and 7.2 million concept names.[1]

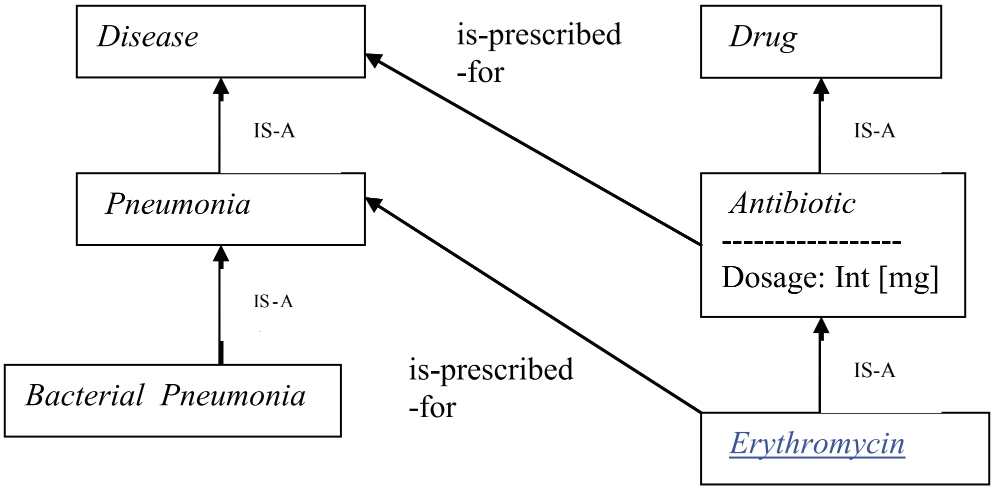
BACKGROUND

The easiest way to understand ontologies is to look at them from a structural perspective. Quillian's original semantic network was a computer data structure that mimicked a dictionary. In the KL-ONE implementation of a semantic network, the IS-A relationship took center stage. Thus, an ontology is (visually) a network of nodes (boxes) and links (arrows) connecting the nodes. Figure 1 shows a tiny excerpt of an ontology.[2]

The basic unit of knowledge in a terminology is a concept. *Drug, Pneumonia, Antibiotic* and *Erythromycin* are concepts. For each concept various kinds of attributes may be specified, e.g., name, ID number, synonyms and other alphanumeric attributes. These attributes provide additional information about the given concepts. In Figure 1, Dosage is an attribute.

Concepts may also refer to other concepts by relationships. In Figure 1, relationships are shown as thin arrows. Relationships have labels attached, such as "is-prescribed-for" and are called "semantic relationships." Semantic Relationships are distinct from the

Figure 1. Example of Medical Terminology applied to the treatment of bacterial pneumonia



special purpose IS-A relationships (bold arrows), which form a specialization/generalization hierarchy. Thus, *Erythromycin* IS-A *Antibiotic* says that *Antibiotic* is more general than *Erythromycin*. In other words, all real-world instances of *Erythromycin* form a subset of all real-world instances of *Antibiotics*. Similarly, *Pneumonia* IS-A *Disease*. The IS-A relationship allows simple forms of reasoning with the concepts of the terminology. The most popular form of reasoning is inheritance. Thus, *Erythromycin* has the attribute Dosage, even though it is not shown in the figure. Dosage is inherited (automatically) along the IS-A link (against the direction of the arrow) from the parent concept *Antibiotic.*

Many ontologies also include *individuals* and/or support different forms of logic-based reasoning. Axioms may be attached to concepts in the taxonomy. The most popular logic formalism is called First Order Predicate Logic (FOPL). Unfortunately, FOPL is neither decidable nor tractable, which means that finding true results may be impossible or may take exponential amounts of time (e.g., millions of years). Thus, weaker logical formalisms have been invented. Indeed, many older and all modern members of the KL-ONE family itself were reconceived as "Description Logics" (Baader et al., 2003).

Description Logics maintain the basic KL-ONE structure and aim for well-defined tradeoffs between computability and expressiveness (Brachman & Levesque, 1984). Meanwhile improved First Order Logic (FOL) provers have also been developed. Tsarkov and Horrocks (2003) presented a comparison between DL and FOL systems. Given the long history of *building* ontologies it is somewhat surprising that *using* ontologies is still not a straightforward process, a phenomenon referred to as *knowledge use paradox* in (Geller et al., 2004).

MAIN FOCUS

Ontologies come in many different shapes and forms. An early categorization of ontologies can be found in (Noy & Friedman, 1997). They distinguished between ontologies based on generality, domain, size, formalism used, etc. John Bateman's comprehensive ontology portal[3] lists ontologies in linguistics, medicine, geography, translation, information reuse, business, general knowledge, engineering, software, etc. Ontologies may also be categorized into terminological versus axiomatized ontologies. Fensel (2004) provides a good, compact review of ontologies and how they will be useful in Electronic Commerce. An easy introduction to ontologies in Bioinformatics is (Baclawski & Niu, 2006). A comprehensive account of many aspects of ontologies is (Gomez-Perez et al., 2004).

One of the most widely used and successful lexical ontologies is WordNet (Fellbaum, 1998). It has inspired similar projects in other languages, such as GermaNet (Hamp and Feldweg, 1997). The CYC project also