Requirements for Ontology Indexed Knowledge Resources: Clarifying the relation of Universals, Prototypes, and Kinds of Metadata

Alan Rector and Robert Stevens
Department of Computer Science, University of Manchester, UK
rector@cs.man.ac.uk rstevens@cs.man.ac.uk

Abstract

Our goal is to provide trusted re-usable knowledge resources for users in specialized domains such as biomedicine. Such users' requirements go beyond 'ontologies' in the narrow sense of the word – and certainly what can be expressed simply in OWL-DL or even OWL-Full. They include defaults and exceptions that can only be met by taking the notion of 'prototype' seriously, and both metaknowledge about the representation and higher order knowledge about the concepts represented. Such users also need information more naturally queried under closed world semantics than under OWL's open world semantics.

1 Introduction

Great expectations have been raisd for the role of "ontologies" in the Semantic Web and more broadly for information integration and re-use. However, users are vague in their understanding of what constitutes an "ontology". Our experience is that many users are looking for something that goes far beyond the narrow definitions of ontologies as models of the concepts in their domain. There need might better be described as:

> *Trusted re-usable resources for the reference knowledge needed to understand, build and maintain sharable knowledge intensive applications and agents.*

Requirements we have encountered in developing such resources include:

1. *Rich metadata* – to understand the knowledge to be used, edited, maintained and shared
2. *Prototypical knowledge* with defaults and exceptions – to hold the knowledge of the 'fail safe' options and to index further knowledge resources
3. *Higher order information about the things represented* – clearly distinguished from metadata about the representations
4. Clear distinctions between information intended for *open and closed world reasoning*
5. *Linguistic information* clearly separated from the conceptual information

Regrettably, users' requirements do not always cleave naturally along the lines laid down by technologies. On the one hand the requirements go well beyond even OWL-Full in their need for prototypes and default reasoning; on the other they are difficult meet on any scale without the rigorous logical underpinnings of a formal language such as OWL-DL.

The examples in this paper are drawn from experience with a large drug ontology [Solomon et al. 1999] developed in a hybrid system based on an earlier description logic technology[Rector et al. 1997]. Analogous reference information on drugs forms part of the US National Cancer Institute's Thesaurus [Goldbeck et al. 2004] which has been subject to a widely circulated critique[Fischer 2004].

2 Basic Distinctions

Our contention is that this can only be achieved by taking key distinctions more seriously between

1. Representations and the things represented
2. Definitional and prototypical information about concepts
3. Open world reasoning about all possible worlds and closed world reasoning about the specific world represented.
4. Linguistic vs conceptual knowledge

2.1 Representation and Things Represented

The most important distinction is between representation and thing represented. Our representations are artifacts in the world. Typically they consist of "classes" representing concepts or universals and "instances" representing the individual things in the world. A key part of any knowledge resource is the knowledge about the classes and instances themselves - *e.g.* authorship, authority, scope, dependency on external resources, etc This is what we refer to as "metadata" – data about the representation itself

By contrast, there are also requirements for *higher order knowledge* about the "concepts represented" – *e.g.* that a concept such as "pill" is vague or that the species "Lion" was first described by Linnaeus Such statements are not metadata – they are not knowledge about our representation but about the things represented.

Both metadata and higher order knowledge bring with them the danger of the paradoxes of self-reference and so have been carefully limited in the OWL language specification. However, clear discussion of the use cases and examples make possible approximations for many cases that are both safe and practical although rarely complete.

2.2 Prototypes and Definitions

Historically, prototypical representation with defaults and exceptions dominated knowledge representation through the 1970s and 1980s and persists as a major topic in cognitive science[Barsalou et al. 1998]. However, representation of defaults and exceptions has proved difficult to incorporate into logic-based formalisms based primarily on universals and definitions. Any representation of prototypical knowledge with defaults and exceptions is entirely missing from OWL.

However, many applications have strong use cases for fail-safe defaults – *e.g.* for drug interactions and contraindications where the worst must be assumed until proven otherwise. Defaults and exceptions have also proved extremely useful in developing highly flexible tailored interfaces[Nowlan 1994]. Ironically, experience in both drug information and user interfaces has shown that the cleaner 'normalised' representations[Rector 2003] made possible by logic based systems such as OWL-DL, minimize the conflicts ("Nixon diamonds") which are inevitable in such systems. Hybrid systems that clearly distinguish between statements that are 'prototypically true' and statements that are 'universally true' have work surprisingly well in both drug information and user interface domains.

2.3 Closed and Open World Knowledge

Most reasoning about instances in databases uses closed world reasoning. Open world reasoning (true A-Box reasoning) has proved extremely expensive. Furthermore, most users assume that reasoning about instances use closed world reasoning. While open world reasoning is natural for the universals represented in the ontology itself – those things true in any world – it is often unnatural for things in other parts of a knowledge resource – those things true in this world. The NCI Thesaurus in which the information on drug licensing lists only the conditions for which each drug is licensed provides a typical example of the confusion. While one solution is to 'close' each statement explicitly, a more practical solution would be to be able to represent that this property is intended to be used under a closed world assumption.

2.4 Language vs Concepts

Many of the issues in ontology development end up turning on what linguistic forms different communities wish to use for the same concepts. While OWL provides support for labels, the richness of notions such as 'preferred terms', 'variants', let alone the complexities of style and usage, require more specialized tools.

3 Summary

Given the above requirements and the pressure of current user demands, we propose a layered hybrid architecture, as shown in Figure 1, which can address most of these needs. Most of this architecture has already been realized in the GALEN project using an earlier generation of description logic [Rector, Bechhofer *et al.* 1997]. Developments covering many of these proposals are already well advanced within the context of OWL

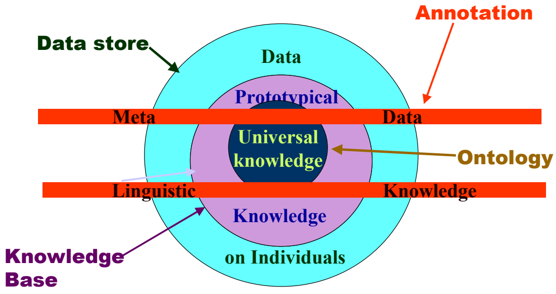

Figure 1: Proposed layered Architecture

Acknowledgments

This work supported in part by the CO-ODE grant from the Joint Information Services Committee (JISC) and the HyOntUse GR/S44585 from the UK EPSRC.

References

[Barsalou, Huttenlocher *et al.* 1998] Barsalou, L. W., J. Huttenlocher and K. Lamberts (1998). "Basing categorization on individuals and events." Cognitive Psychology **36**: 203-272.

[Fischer 2004] Fischer, D. H. (2004). Converting a thesuarus to OWL: Notes on the paper "The national cancer institute's thesuarus and ontology", Fraunhofer Inwstitute Integrated Publication and Information Systems, Darmstadt, Germany.

[Goldbeck, Fragoso *et al.* 2004] Goldbeck, J., G. Fragoso, F. Hartel, J. Hendler, et al. (2004). "The National Cancer Institute's thésaurus and ontology." Journal of Web Semantics **1**(1): 32-36.

[Nowlan 1994] Nowlan, W. A. (1994). "Clinical workstation: Identifying clinical requirements and understanding clinical information." International Journal of Bio-Medical Computing **34**: 85-94.

[Rector 2003] Rector, A. (2003). Modularisation of Domain Ontologies Implemented in Description Logics and related formalisms including OWL. Knowledge Capture 2003, Sanibel Island, FL, ACM.

[Rector, Bechhofer *et al.* 1997] Rector, A., S. Bechhofer, C. Goble, I. Horrocks, et al. (1997). "The GRAIL concept modelling language for medical terminology." Artificial Intelligence in Medicine **9**: 139-171.

[Solomon, Wroe *et al.* 1999] Solomon, W., C. Wroe, J. E. Rogers and A. Rector (1999). "A reference terminology for drugs." Journal of the American Medical Informatics Association((Fall Symposium Special Issue)): 152-155.