Problems of Scale in Building, Maintaining and Using Very Large Formal Ontologies

Doug LENAT
Cycorp, USA

Abstract. Though Cyc is a formal ontology, the process of building it, over the past 22 years, has been a passionately empirical process. We have had several surprises along the way, some of them scientific, some engineering, and some socio-logical. For instance, the requirement to represent arbitrary pieces of commonsense knowledge led us, in the mid-1980’s, against our intuitions, to move to an increasingly expressive formal representation language. By 1990, we had to admit that the dream of a “Final Encyclopedia” of correct knowledge was a chimera, and what we needed to focus on was a tapestry of locally-consistent “micro-theories” containing contextualized knowledge. Since then, we have begun to work out the fine structure of these micro-theories, their important attributes and ways in which they related to each other, and to appreciate the surprising complexity of the calculi required to formally reason across them. We have also experienced a tipping-point, methodologically, over the past few years, as the ontology has grown large enough to serve as an inductive bias for further knowledge acquisition. I.e., Cyc increasingly actively helps with its own continuing expansion, and by now almost all the activity going on at Cycorp is related to semi-automatic learning from corpora (including the Web) of text and structured sources, whereas as recently as three years ago the majority of the activity here was a cadre of ontological engineers manually writing more axioms to expand the Cyc Knowledge Base. We’ve also developed and used — and in most cases discarded — a series of interfaces, training paradigms, and so on, as the ontology has grown. In the talk, I shall survey what we used, and when, and why we moved on. Most of the reasons have to do with the ontology outgrowing the tools, or increasing variety among the types of users and ontological engineers. Finally, I will discuss some of our ongoing research efforts, and ongoing interface efforts, which are becoming increasingly intermingled — and why that is perhaps inevitable.

Keywords. Ontology Design, Ontology Maintenance