All You Can Eat Ontology-Building: Feeding Wikipedia to Cyc

Catherine Legg

Department of Philosophy and Religious Studies
University of Waikato, New Zealand
clegg@waikato.ac.nz

Abstract

In order to achieve a genuinely intelligent World Wide Web, it seems that building some kind of general machine-readable ontology is an inescapable task. Yet the past 20 years have shown that hand-coding formal ontologies is not practicable. A recent explosion of free user-supplied knowledge on the Web has led to great strides in automatic ontology-building (e.g. YAGO, DBpedia, recent work at EMLR), but here quality-control is still a major issue. Ideally one should automatically build onto an already intelligent base. I suggest that the long-running Cyc project can finally come into its own here, describing methods developed at the University of Waikato over the past summer whereby 35K new concepts mined from Wikipedia were added to appropriate Cyc collections, and automatically categorized as instances or subcollections. Most importantly, Cyc itself was leveraged for ontological quality control by 'feeding' assertions to it one by one, allowing it to 'regurgitate' those that are ontologically unsound. Cyc is arguably the only ontology currently sophisticated enough to be able to perform such a 'digestive' function, using its principled taxonomic structure and purpose-built inference engine. It is suggested that a traditional fixation of AI researchers on realizing the intelligence of the brain has perhaps caused us to overlook more humble yet genuine steps towards the AI vision which might be gained by realizing the intelligence of the stomach.