Panel: Large Knowledge Bases

Adam Pease (Teknowledge, chair), Chris Welty (Vassar College),
Pat Hayes (U. West Florida), Anthony G. Cohn (U. Leeds), Ken Murray (SRI)

Abstract

It is estimated that 1-2 exabytes of data is now being generated each year, almost all of it in purely digital form (Lyman et. al. 2000). Properly structured, this information could form a global knowledge base. Currently however, this information exists in many different forms, many of which are only suitable for human consumption, and which are largely opaque to computer based understanding.

Major efforts to build large formal ontologies or address issues in their construction have been undertaken funded by the government in the US such as the DARPA Knowledge Sharing Effort (Patil et al, 1992), High Performance Knowledge Bases (Cohen et. al., 1998), Rapid Knowledge Formation (RKF, 2002) and in Europe including Advanced Knowledge Technologies (Shadbolt, 2001) and OntoWeb (OntoWeb, 2002), as international standards efforts such as the IEEE Standard Upper Ontology (Niles & Pease, 2001) and by private corporations including Cycorp (Lenat, 1995) and VerticalNet (Roddy et. al., 2000).

We will discuss

- the applicability of these efforts to question answering and how they might augment information retrieval.
- progress of these efforts and the rate at which competent knowledge bases will be able to support question-answering tasks of various kinds.
- how upper ontologies, particularly spatial and temporal ontologies can impact question-answering tasks.
- Is the rate of progress and cost of building ontologies going to pay off?
- Can ontology-based systems work synergistically with IR-based systems?
- Are small ontologies better than large ontologies?

References


Lenat, D., 1995, "Cyc: A Large-Scale Investment in Knowledge Infrastructure". Communications of the ACM 38, no. 11, November. See also http://www.cyc.com

http://www.sims.berkeley.edu/research/projects/how-much-info


