

KEYNOTE ADDRESS

Computers versus Common Sense

Dr. Douglas B. Lenat
Principal Scientist
MCC
Austin, Texas

Suppose that it is the year 2020, and the ‘enterprise wide knowledge utility’ is a reality. Is it data- or inference- intensive? Where does it sit on the expressiveness versus efficiency trade-off curve? What was the “AI bottleneck” and the “DB bottleneck” to building it? How local/guaranteed/distributed/heterogeneous/parallel is it? Looking back thirty years from now, where should we have spent our limited resources (people and dollars) in the 90’s, to have accelerated its coming into existence?

I’ll give representative answers from today’s Data Base and Artificial Intelligence communities, and my own opinions which differ at times from both of those. The upshot of that is a mandate to AI, to roll up their sleeves and begin to build and experiment with very large knowledge bases.

I came to this conclusion in 1984, and with no little trepidation took my own advice. I left the field of machine learning to embark on a decade-long effort to construct a Knowledge Base with on the order of ten million assertions, spanning all of human consensus reality, down to some reasonable level of depth.

Why that particular KB, instead of, say, one that focused on engineering electromechanical devices? Intelligent behavior, especially in unexpected situations, requires being able to fall back on general knowledge, and being able to analogize to specific but far-flung knowledge. Natural language understanding, even in seemingly narrow domains such as “Wall St. Journal articles about mergers and acquisitions,” also requires broad general knowledge, to resolve word sense ambiguities, ellipses, pronominal referents, and of course metaphors and analogies. And machine learning (automated discovery) occurs at the fringe of what one already knows, hence also would be empowered by such a KB.

Aside from the magnitude of the task, the person-centuries it would take to build the KB, my concern stemmed from the dangers of getting stuck on representation thorns along the way (representing and reasoning with time, space, substances, belief, intention, emotion, causality, parts, awareness, contexts, counterfactuals, and so on.) Our approach had to be a pragmatic one: for each of those issues, we found a set of partial solutions which, in aggregate, efficiently covered the common cases. Notice this is a violation of the 1-2-infinity taboo which has so dominated most sciences in the past century (i.e., separately handling 27 special cases of phenomenon x is considered unaesthetic, unscientific, etc.)

Some of the other interesting issues in building the large KB are: how we decide what knowledge to include; how it accesses, deduces, or guesses answers; how it deals with uncertainty and contradictions; and how we coordinate a large team of knowledge enterers without having them “diverge.”

We’ve gotten pretty far along already, in the process of building up an already-large KB (over a million rules and general assertions). I’ll present some of those solutions, giving the flavor of our representation language, ontology, and methodology. Finally I’ll relate this back to the general issue of producing a global information infrastructure.

I invite those of you who wish additional detail to read the article by R. V. Guha and myself in the August ’90 *CACM*.