Editorial: Ontology Challenges: A Thumbnail Historical Perspective

James Geller[1], Yehoshua Perl[1,*], Jintae Lee[2]

[1]CS Department, New Jersey Institute of Technology, Newark, NJ, USA
[2]Leeds School of Business, University of Colorado at Boulder, USA

Rarely in the history of computer science has there been an idea that was so obviously right and that has failed for so long to deliver on its promise. In its simplest form, this idea can be phrased as follows: People use hierarchically organized knowledge about domain objects and relations as a major ingredient of their intelligent behavior. Thus, one might reason abductively, if we equip computers with this kind of knowledge and appropriate processing algorithms, we should get intelligent behavior.

The earliest history of artificial intelligence concentrated on the processing algorithms. But, after many unexpected difficulties, research turned more and more toward the knowledge part of the *knowledge plus reasoning algorithms* framework. Yet, paradoxically, even though large amounts of knowledge have been amassed in some areas, notably medical informatics, and even though knowledge structures have been studied for over 35 years, the expected big breakthrough toward computational intelligence has not happened. This leaves us with the initially mentioned paradox: Why has an idea that is so obviously right not led to the expected results?

In 1968, Quillian (Quillian 1968) published his famous semantic network paper. His network was essentially an electronic representation of a dictionary of real-world terms. Using that dictionary and spreading activation reasoning, Quillian's program was able to generate simple natural language expressions of stored semantic relationships. Given the extremely limited hardware (in today's view) that Quillian had available, his achievement was a major milestone that inspired research in semantic networks (Woods 1975; Lehmann 1992; Sowa 1991), cognitive psychology (Collins and Quillian 1969; Collins and Loftus 1975; Anderson 1983) and a field that was to become known as knowledge representation (Brachman and Schmolze 1985; Cercone and McCalla 1987; Sowa 2000).

By the mid-1980s, knowledge base building entered a crisis. Many researchers argued that reasoning issues had to be addressed first, before building sizable representations of real-world knowledge. What could one do with a large representation of knowledge if the correct way of nonmonotonic reasoning (Ginsberg 1987) or tractable reasoning (Brachman and Levesque 1984) was not known yet? Mainstream research in knowledge representation became extremely formal and knowledge base building became unfashionable, except in the field of expert systems. There, knowledge had to be built in a way that provided an immediate pay off as a working system. For that purpose, production rules were the best available mechanism. Building large knowledge bases for the simple purpose of representing the world as it is was deemed undesirable, as it had neither the formal rigor of mainstream knowledge representation research nor the deliverables quality of expert systems research.

Luckily, there were a few exceptions. We will focus on two of these. The first exception has been the field of medical informatics, which, uninhibited by the fashions of knowledge representation, built massive medical knowledge bases. The best known among those is the Unified Medical Language System (UMLS) (Tuttle et al. 1991; Humphreys et al. 1998; Lindberg et al. 1993; Humphreys and Lindberg 1989). The paper by Jurisica, Mylopoulos and Yu in this special issue uses the medical domain as its application area and contains some material about the UMLS. For a deeper perspective regarding UMLS research, see McCray and Miller (1998).

The second exception was the CYC research group (Guha and Lenat 1990; Lenat 1990; Lenat and Guha 1990). The CYC project elicited as much excitement as skepticism in the hallways of conference venues and research institutions. Regrettably, the skeptics were closer to being right, and their counterparts (including the editors) are still wondering why. With the partial failure of the CYC project (defined by not reaching some of the very well-defined and published goals of CYC's own designers), the early 1990s appeared to be the time when practical knowledge base building was completely at the ropes. With the resurgence of interest in subsymbolic representations, such as neural networks, sparked by the PDP group in 1986 (Rumelhart and McClelland 1986), there seemed to be no direction left to turn. Then, around 1993, people like Thomas Gruber reinvigorated the knowledge representation enterprise by suggesting to look at knowledge base building from the new viewpoint of crafting ontologies (Gruber 1993).

Ontology building deals with modeling the world with shareable knowledge structures. The additional constraints of both formal knowledge representation and expert systems are considered secondary compared with the main goal of building the appropriate knowledge structures. Yet, the ontology enterprise has not brought the long-expected breakthrough either, and this is part of the *knowledge use paradox*. One stands to conclude that the saying "a little bit of knowledge is a dangerous thing" must be at work here. But CYC has a large amount of knowledge, and it still did not deliver everything that was expected.

With the emergence of the Semantic Web, the development of ontologies and research in ontology integration have moved center stage (Berners-Lee et al. 2001; Guarino and Poli 1995; Hendler 2001; Gruninger and Lee 2002). Ontology researchers cannot hide in the ivory tower anymore. If the success of the next generation of the World-Wide Web depends on progress in ontology research, then we really need to overcome the knowledge use paradox. The editors believe that the four papers in this special issue describe four directions of how to attack problems in ontology research.

One research direction is to use better methodologies when building ontologies from scratch or from smaller ontologies. Pinto and Martins describe several ontol-

ogy building methodologies, using approaches developed in software engineering as their guidelines. Another research direction is to attack the knowledge acquisition bottleneck head on. Richards' approach is to use a combination of two existing methodologies. One of these results in expert system rules, which are then reverse engineered in the second step into a concept representation. A third direction for research deals with evolving existing ontologies. Like Pinto and Martins, Noy and Klein also derive lessons from outside of artificial intelligence, namely from the database theory of schema evolution. The last research direction in this special issue is to define a landscape of different kinds of ontologies. Knowing the properties and the kind of the ontology that one is trying to build should be helpful in building it. Jurisica, Mylopoulos and Yu propose such a landscape of ontologies.

In more detail, the authors offer the following contributions to this special issue:

Pinto and Martins review several ontology building methodologies. They impose a framework on ontological engineering (OE) that is derived by a comparison with existing methodologies in software engineering. Their survey includes both techniques for building ontologies from scratch and for building ontologies from smaller components. For the latter, they distinguish between fusion/merging, where several ontologies from the same subject area are combined, and composition/integration, where an ontology for a larger domain is built from several ontologies of subdomains. They conclude their paper with a review of future directions in ontological engineering relevant to the Semantic Web.

Richards describes an approach for dealing with the difficult problem of building ontologies. This approach consists of a combination of case-driven rule acquisition resulting in ripple-down rules with a knowledge representation technique known as formal concept analysis (FCA). Formal concept analysis is used to distill a concept lattice out of the attributes used in the incrementally created base of ripple-down rules. The methodology presented is well grounded in a real expert system task. Furthermore, Richards describes studies with graduate students that show improved problem-solving behavior when a diagram of a concept lattice is available for the test domain, compared with a textual description of the same knowledge.

Noy and Klein start out with the observation that useful ontologies need to change in order to adapt to new situations and that the evolution of ontologies is related to the evolution of database schemas. They then analyze in detail how ontology evolution is similar to schema evolution, thus making it possible to use previous research on the topic, and how ontology evolution is different. They find that the distinction between evolution and versioning does not hold for ontologies, and that some additional complications in ontology evolution are the result of the need to manage operations, especially composite operations. The paper clearly establishes the fact that ontology evolution is more difficult than schema evolution and discusses several open research issues.

Contrary to the assumptions of some researchers, there is to date no universal agreement on how the landscape of ontologies looks. Jurisica, Mylopoulos and Yu present a big step toward such a landscape by considering four kinds of ontologies. A static ontology describes objects that exist, their attributes and relationships. A dynamic ontology describes states and processes. An intentional ontology describes agents and their propositional attitudes (i.e., what they believe, want, etc.). Finally, a social ontology covers social settings and organizational structures. The authors cover ontologies from the view points of artificial intelligence, information science and information systems and review ontologies in the area of medical informatics, thus giving the paper a wide basis for their analysis.

The editors thank the reviewers of this special issue for their diligent work. We also extend our special thanks to Xindong Wu and E.K. Park for their continued support. We hope that this special issue will be successful in defining directions that will lead to overcoming the knowledge use paradox.

References

Anderson JR (1983) A spreading activation theory of memory. J Verbal Learn Verbal Behav 22:261–295

Berners-Lee T, Hendler J, Lassila O (2001) The semantic web. Sci Am 284(5):34–43

Brachman RJ, Levesque HJ (1984) The tractability of subsumption in frame based description languages. Proceedings of AAAI-84, pp 34–37

Brachman RJ, Schmolze J (1985) An overview of the KL-ONE knowledge representation system. Cogni Sci 9(2):171–216

Cercone N, McCalla G (1987) The knowledge frontier. Springer, Berlin Heidelberg New York

Collins AM, Loftus EF (1975) A spreading activation theory of semantic processing. Psychol Rev 82:407–428

Collins AM, Quillian MR (1969) Retrieval time from semantic memory. J Verbal Learn Verbal Behav 8:240–248

Ginsberg M (1987) Readings in nonmonotonic reasoning. Morgan Kaufmann, San Mateo, CA

Gruber TR (1993) A translation approach to portable ontology specifications. Knowl Acquis 5:199–220

Gruninger M, Lee J (2002) Special issue on ontology applications and design. Commun ACM 45(2):39–65

Guarino N, Poli R (1995) Special issue on formal ontology in conceptual analysis and knowledge representation. J Hum Comput Stud 43:625–640

Guha RV, Lenat DB (1990) CYC A midterm report. AI Mag 11(3):32–59

Hendler J (2001) Special issue on agents and the Semantic Web. IEEE Intell Syst 16(2):30–37

Humphreys BL, Lindberg DAB (1989) Building the Unified Medical Language System. Proceedings of the 13th annual symposium on computer applications in medical care, pp 475–480

Humphreys BL, Lindberg DAB, Schoolman HM, Barnett GO (1998) The Unified Medical Language System: an informatics research collaboration. JAMIA 5(1):1–11

Lehmann F (ed) (1992) Semantic networks in artificial intelligence. Pergamon, Tarrytown, NY

Lenat DB (1990) CYC: Toward programs with common sense. Commun ACM 33(8):30–49

Lenat DB, Guha RV (1990) Building large knowledge-based systems: representation and inference in the CYC project. Addison-Wesley, Reading, MA

Lindberg DAB, Humphreys BL, McCray AT (1993) The Unified Medical Language System. Meth Inf Med 32:281–291

McCray AT, Miller RA (1998) Making the conceptual connections: the UMLS after a decade of research and development. J Am Med Inf Assoc 5(1):129–130

Quillian MR (1968) Semantic memory. In: Minsky ML (ed) Semantic information processing. MIT Press, Cambridge, MA, pp 227–270

Rumelhart DE, McClelland JL (1986) Parallel distributed processing: explorations in the microstructure of cognition, vol 1 and 2. MIT Press, Cambridge, MA

Sowa JF (ed) (1991) Principles of semantic networks. Morgan Kaufmann, San Mateo, CA

Sowa JF (2000) Knowledge representation. Brooks/Cole, Pacific Grove, CA

Tuttle MS, Sherertz DD, Erlbaum MS, et al (1991) Adding your terms and relationships to the UMLS metathesaurus. In: Clayton PD (ed) Proceedings of the 15th annual SCAMC, pp 219–223

Woods WA (1975) What's in a link: foundations for semantic networks. In: Bobrow DG, Collins AM (eds) Representation and understanding. Academic, New York, pp 35–82

Author Biographies

James Geller received an Electrical Engineering Diploma from the Technical University Vienna, Austria, in 1979. His MS degree (1984) and his PhD degree (1988) in Computer Science were awarded by the State University of New York at Buffalo. He spent the year before his doctoral defense at the Information Sciences Institute (ISI) of USC in Los Angeles. James Geller is currently professor in the Computer Science Department of the New Jersey Institute of Technology, where he is also the Director of the Semantic Web and Ontologies Laboratory. He spent a sabbatical in 1999–2000 at the NEC Research Institute in Princeton, NJ, and is currently a consultant there. Dr. Geller has published numerous journal and conference papers in a number of areas, including knowledge representation, parallel artificial intelligence, medical informatics, and object-oriented databases. His current research interests concentrate on medical vocabularies, ontologies, the Semantic Web and Web mining.

Yehoshua Perl received his PhD in Computer Science in 1975 from the Weizmann Institute of Science, Israel. He was appointed lecturer (1975) and senior lecturer (1979) at the Bar-Ilan University, Israel, and spent a sabbatical at the University of Illinois (1977–78). From 1982 to 1985, he was visiting associate professor at Rutgers, New Brunswick. Since 1985, he has been at the CS Department at the New Jersey Institute of Technology (NJIT), where he became professor in 1987. He received the Harlan Perlis Research Award in 1996. Dr. Perl has published more than 80 papers in international journals and conferences in object-oriented databases, medical informatics, medical vocabularies, design and analysis of algorithms, design of networks, sorting networks, graph theory, and data compression. Highlights include the shifting algorithm technique for tree partitioning, analysis of interpolation search, the design of periodic sorting networks and the Object-Oriented Healthcare Vocabulary Repository project supported by NIST.

Jintae Lee is currently an Associate Professor at the Leeds School of Business, University of Colorado at Boulder. He received the BA in Mathematics from the University of Chicago, the MA in Psychology from Harvard University and the PhD in Computer Science from MIT. His research interests are in knowledge management technologies from both systems and behavioral perspectives. He was a guest editor of the Communications of the ACM 2002 special issue on Ontology Design and Applications. He has published in various journals, including Management Science, Information Systems, Human Computer Interaction, and the Communications of the ACM.

Correspondence and offprint requests to: James Geller, CS Department, New Jersey Institute of Technology, Newark, NJ 07102, USA. Email: geller@oak.njit.edu