How the Semantic Web will change KR: challenges and opportunities for a new research agenda*

Frank van Harmelen
Department of AI
Vrije Universiteit Amsterdam
Frank.van.Harmelen@cs.vu.nl

The Semantic Web as the new habitat for agents?

Currently, the Web is the largest available environment for the deployment of agents, and much work in agent research is driven by Web-based applications ([Luke et al., 1997, Joachims et al., 1997, Bollacker et al., 1998, Doorenbos et al., 1997] are just some examples; see also the May 2000 special issue of the Artificial Intelligence Journal on Intelligent Internet Systems, vol. 118, no 1-2). However, such applications of agent technology are hampered by the fact that the Web is not geared towards agent use, but is rather designed for human use. Current Web-resources are lacking in explicit, machine accessible descriptions of their contents: they are only fully accessible to agents with a competent grasp of English (i.e. limited to human agents only).

The advent of the Semantic Web is rapidly making available a semantically much richer layer of machine accessible contents on top of the existing infrastructure. This will provide a much richer habitat for agents then the current WWW. However, many of the assumptions that underly current Knowledge Representation (KR) techniques as used in agent technology are no longer valid when KR techniques are deployed in a large scale and open environment such as the Semantic Web promises to be.

In this short note we discuss some of the assumptions underlying current KR technology that will have to be revised when applied to the Semantic Web, and we discuss some of the research challenges and new opportunities that arise out of such a revision. We accompany each point below with some references to relevant agent literature, without claiming these to be either exhaustive or the most authoritative references.

Scale: It is almost a platitude to mention the size of the Internet in general, and of the Web in particular. Google now indexes well over a billion pages, and the number of connected hosts runs into the millions. These numbers are orders of magnitude larger than any traditional single knowledge-base for which much of current KR technology has been designed [Turner and Jennings, 2000].

Change rate: Many portions of the Internet display a very high change rate, with information changing on the timescale of days (e.g. news sites), hours (e.g. auction sites), or even minutes (e.g. stockmarkets). On the other hand, KR techniques, such as those from knowledge engineering, typically have been designed for update rates in the order of months, or even slower. (e.g. [Schut and Wooldridge, 2000])

Lack of referential integrity: One of the major departures that the Web took from traditional hypertext systems in the early '90s was to drop referential integrity: links were no longer guaranteed to point to anything, and were allowed to be "broken". Rather than problematic, this decision enabled the Web to

*This material is directly based on a workshop session during the Dagstuhl Seminar on the Semantic Web in the spring of 2000, updated by many discussions in the period since that memorable week. Although all those present at the Dagstuhl seminar have contributed to what follows, I want to acknowledge in particular the contributions of Stefan Decker, Jim Hendler and Deborah McGuinness.

Many thanks also to Martijn Schut for pointing me to relevant agent literature.

scale well beyond anything that traditional hypertext systems had been able to achieve. In the same vein, the Semantic Web will have to cope with semantic forms of the "broken link": portions of knowledge-bases that are simply missing (e.g. [Sichman and Demazeau, 1995, Lin, 1994].

Distributed authority: Traditional KR has largely avoided the matter of trustworthiness of statements in a KB: statements in a KB were simply generally held to be true. Clearly, this assumption will have to be revised in a context where the typical KB consists of many sections imported from many different sources, any not under one's own control. Questions of trust become predominent in such a new setting [Jonker and Treur, 1999].

Variable quality of knowledge: Somewhat related to the previous point is the fact that in a distributed environment like the Semantic Web, knowledge from different sources is likely to differ greatly in quality, up to the point where some of these sources will even be internally inconsistent. As a result, such different sources of varying quality should not all be treated on the same footing. Techniques for the local containment of inconsistencies will have to become much more important than they have been in current KR research.

Unpredictable use of knowledge: Typically, knowledge bases are built with a particular usage in mind: it is known beforehand whether a knowledge base will be used for diagnostic reasoning, or for planning, etc. This allows knowledge engineers to make certain design decisions in the KB that are justified by the intended use. In an open environment like the Web, knowledge bases are likely to be used by third parties for purposes entirely different from the one for which they were originally designed. This puts a much higher premium on the old adage of task-independent formulation of knowledge. ([Garland and Alterman, 1995, Musen, 1992, Eichmann, 1992] and many other references in the Knowledge Engineering literature).

Multiple knowledge sources: In an open environment like the Web, knowledge will no longer be provided by a single (team of) engineer(s), but rather by linking or importing knowledge from many existing knowledge-sources. Problems such as inhomogeneous vocabularies and different conceptualisations of the same domain will become much more urgent than they are today.

Diversity of content: Typical knowledge bases deal with a narrow focus of interest, and assume a certain homogeneity of vocabulary. Again, because of the open and diverse nature of the Web, this assumption can no longer be upheld. Again, the question of how to reconcile multiple vocabularies on the same topic becomes of primary importance.

Linking, not copying: Because of the sheer size of the Web, it will be impossible to physically copy the contents of other knowledge sources when they are used. Instead, mechanisms will have to be devised for linking to such remote knowledge bases (the Semantic Web equivalent of HREF). At the same time, we will have to find solutions for the increase in access time that this entails. Suddenly, the cost of accessing a single axiom (possibly located many Internet hops away) becomes a factor to be reckoned with, very much unlike the current situation where access to the axioms of one's theory is counted as a zero-cost operation.

Justifications as first-order citizens: In an environment where one's conclusions may crucially depend on knowledge provided by unknown third parties, justifications of these conclusions become of prime importance. This is of course closely connected with a long tradition of work in KR on "explainable expert systems", with a crucial difference: the justifications are no longer primarily intended for human consumption, but rather as the basis for machines that verify the conclusions (or at least the inference chains along which they were reached). This means that such justifications will have to be passed around and inspected as "first-order citizens" on the Semantic Web, very unlike the rather derivative role played by explanations in current KB systems. [Parsons et al., 1998, Kuhnel, 1999]

Robust inferencing: In an environment the size of the Web we must abandon the classical ideal of sound and complete reasoners. Our reasoners will almost certainly have to be incomplete (no longer guaranteeing to return all logically valid results), but most likely also unsound: sometimes jumping to a logically unwarranted conclusion. Furthermore, the degrees of such incompleteness or unsoundness must be a function of the available resources. Answers will often have to be approximate (where ideally, the reasoner can give us an indication of the quality of such approximations) [Zilberstein and Russell, 1995, Russell et al., 1993, Lesser et al., 2000].

Agent research leading KR?

Of course, not all of the above research challenges are new to KR, and many of them have been on the research agenda to some extent. Examples are approximate reasoning, trust, task-independent formulation of knowledge, and reconciling multiple vocabularies. Nevertheless, from our analysis of "assumptions that break KR on the Web", it appears that many of these items will have to become much more prominent on the research agenda then they currently are.

In fact, many of the items above that are underdeveloped in current KR research are actually actively investigated in the agent community, as is shown in the references mentioned above.

It would seem from this that the agent community has an important role to play in many of the issues that will be key to the success of the Semantic Web.

References

[Bollacker et al., 1998] Bollacker, K., Lawrence, S., and Giles, C. L. (1998). CiteSeer: An autonomous web agent for automatic retrieval and identification of interesting publications. In Sycara, K. P. and Wooldridge, M., editors, *Proceedings of the Second International Conference on Autonomous Agents*, pages 116–123, New York. ACM Press.

[Doorenbos et al., 1997] Doorenbos, R. B., Etzioni, O., and Weld, D. S. (1997). A scalable comparison-shopping agent for the world-wide web. In Johnson, W. L. and Hayes-Roth, B., editors, *Proceedings of the First International Conference on Autonomous Agents (Agents'97)*, pages 39–48, Marina del Rey, CA, USA. ACM Press.

[Eichmann, 1992] Eichmann, D. (1992). Supporting multiple domains in a single reuse repository. In *Proc. Fourth International Conference of Software Engineering and Knowledge Engineering (SEKE'92)*, pages 164–169.

[Garland and Alterman, 1995] Garland, A. and Alterman, R. (1995). Preparation of multi-agent knowledge for reuse. In Aha, D. W. and Ram, A., editors, *Working Notes for the AAAI Symposium on Adaptation of Knowldege for Reuse*, Cambridge, MA. AAAI.

[Joachims et al., 1997] Joachims, T., Freitag, D., and Mitchell, T. M. (1997). Web watcher: A tour guide for the world wide web. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI-97)*, pages 770–777.

[Jonker and Treur, 1999] Jonker, C. M. and Treur, J. (1999). Formal analysis of models for the dynamics of trust based on experiences. In Garijo, F. J. and Boman, M., editors, *Proceedings of the 9th European Workshop on Modelling Autonomous Agents in a Multi-Agent World : Multi-Agent System Engineering (MAAMAW-99)*, volume 1647, pages 221–231, Berlin. Springer-Verlag: Heidelberg, Germany.

[Kuhnel, 1999] Kuhnel, R. (1999). Reaching agreements through argumentation: a logical model and implementation. *Artificial Intelligence*, 104(1):1–69.

[Lesser et al., 2000] Lesser, V., Hornling, B., Klassner, F., Raja, A., Wagner, T., and Zhang, S. (2000). Big: An agent for resource-bounded information gathering and decision making. *Artificial Intelligence*, 118(1-2):197–244.

[Lin, 1994] Lin, J. (1994). Consistent belief reasoning in the presence of inconsistency. In Fagin, R., editor, *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the Fifth Conference (TARK 1994)*, pages 80–94. Morgan Kaufmann, San Francisco.

[Luke et al., 1997] Luke, S., Spector, L., Rager, D., and Hendler, J. (1997). Ontology-based web agents. In Johnson, W. L. and Hayes-Roth, B., editors, *Proceedings of the First International Conference on Autonomous Agents (Agents'97)*, pages 59–68, Marina del Rey, CA, USA. ACM Press.

[Musen, 1992] Musen, M. (1992). Dimensions of knowledge sharing and reuse. *Computers and Biomedical Research*, 25:435–467.

[Parsons et al., 1998] Parsons, S., Sierra, C., and Jennings, N. (1998). Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292.

[Russell et al., 1993] Russell, S. J., Subramanian, D., and Parr, R. (1993). Provably bounded optimal agents. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI-93)*, pages 338–344, Chambery, France.

[Schut and Wooldridge, 2000] Schut, M. and Wooldridge, M. (2000). Intention reconsideration in complex environments. In *Proceedings of the 4th International Conference on Autonomous Agents*.

[Sichman and Demazeau, 1995] Sichman, J. S. and Demazeau, Y. (1995). Exploiting social reasoning to deal with agency level inconsistency. In *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95)*. AAAI Press/The MIT Press.

[Turner and Jennings, 2000] Turner, J. and Jennings, N. (2000). Improving the scalability of multi-agent systems. In Wagner, T. and Rana, O., editors, *Infrastructure for Agents, Multi-Agent Systems, and Scalable Multi-Agent Systems*, volume 1887 of *Lecture Notes in Artificial Intelligence*, pages 246–262. Springer Verlag.

[Zilberstein and Russell, 1995] Zilberstein, S. and Russell, S. (1995). Approximate reasoning using anytime algorithms. In Natarajan, S., editor, *Imprecise and Approximate Computation*. Kluwer Academic Publishers.