Knowledge Maintenance and the Frame Problem

Christopher BREWSTER
Department of Computer Science,
University of Sheffield,
Regent Court, 211 Portebello Street,
Sheffield, S1 4DP
C.Brewster@dcs.shef.ac.uk

Abstract

Knowledge maintenance is a major challenge for both knowledge management and the Semantic Web. Operating over the Semantic Web, there will be a network of collaborating agents, each with their own ontologies or knowledge bases. Change in the knowledge state of one agent may need to be propagated across a number of agents and their associated ontologies. The challenge is to decide how to propagate a change of knowledge state. The effects of a change in knowledge state cannot be known in advance, and so an agent cannot know who should be informed unless it adopts a simple 'tell everyone – everything' strategy. This situation is highly reminiscent of the classic Frame Problem in AI. We argue that for agent-based technologies to succeed, far greater attention must be given to creating an appropriate model for knowledge update. In a closed system, simple strategies are possible (e.g. 'sleeping dog' or 'cheap test' or even complete checking). However, in an open system where cause and effect are unpredictable, a coherent cost-benefit based model of agent interaction is essential. Otherwise, the effectiveness of every act of knowledge update/maintenance is brought into question.

Introduction

Knowledge Management (KM) has become an increasingly important area of research in Artificial Intelligence, but there are a number of challenges facing its use in this area, including knowledge acquisition, knowledge modelling and knowledge maintenance. In this paper, we will focus on the problem of knowledge maintenance, particularly from the perspective of the Semantic Web (SW) and the use of agent technologies.

The label 'Semantic Web' refers to a vision of a future Internet, which is more effective, more user-friendly, more intelligent, and one which (Berners-Lee *et al.* 2001) sees the Internet as a far more sophisticated structure than the present one of hyperlinked documents designed to be visually presentable. By using a number of technologies, a 'semantic' dimension will be added which will allow the Web to consist not only of documents but also of machine-readable structured data. This structured data will serve both human and machine needs, allowing two key areas of application (EU-NSF 2001):

a) Business or organisationally oriented applications such as b2b electronic commerce, or 'grid' applications where data and processing facilities are distributed and seamlessly integrated.
b) Intelligent personal assistants which will collect information and notify other assistants of events permitting the efficient management of personal data, personal interests and other activities. These assistants are conceived of as agents much like a personal travel agent[1].

It is particularly in the development and deployment of such intelligent software agents that many writers claim we will see the real power of the Semantic Web in our everyday lives (Hendler 1999, Kumar *et al.* 2002).

In order to perform a variety of 'intelligent' tasks, these agents must have internal representations of (part of) the world, or more accurately, of the domain of interest to them. These

[1] This concept was the AI challenge explicitly made by McCarthy at the time of McCarthy and Hayes (1969).

will take the form of ontologies or some other knowledge base format. Whenever the agents interact with other agents or acquire some information from the Web, a process of updating their representation must occur. In this context, the ability to undertake knowledge maintenance successfully is central to the Semantic Web and the deployment of agents in this environment.

In a limited domain, the limitations of the complexity of both the world modelled and the knowledge base itself, enables such systems to function effectively. As one scales up, however, one encounters a new version of the traditional Frame Problem, which can be viewed as the problem of systematically updating any knowledge representation. Given an external event or piece of new information, a system has to determine what the consequences of accepting that information are: Perhaps that every item of knowledge in the knowledge base must be checked for a potential inference? Or should a system draw no inferences unless told to do so? Neither approach is adequate. This is the traditional challenge of the Frame Problem.

With a large number of agents communicating with each other, the Frame Problem takes on a further dimension, in that decisions have to be made as to which agents to inform of a given event or piece of information. Unless events or information are suitably categorised in advance, then which agents are interested or affected by an event is unknown. This means that, for all unknown or new events, either all agents must be informed, or none, or decision criteria have to be established, something which researchers have not fully considered as yet.

It is our claim in this paper that the issues of knowledge maintenance, the functioning of the Semantic Web, communication between agents and the Frame Problem are tightly connected. With increasing complexity and more open systems, solutions to one of these areas will show the way towards solutions to the others.

This paper is organised as follows. In Section 1, we consider knowledge management and particularly knowledge maintenance, something about which relatively little has been written until recently. In Section 2, we review the Frame Problem, a classic issue in AI for over 20 years, and note the correspondence between the need to update a robot's representation of the world and knowledge maintenance. We then describe a scenario for the interaction of intelligent agents and consider the consequences of scaling up from the toy world of a research laboratory to the full Internet.

1 Knowledge Management and Maintenance

The past decade has seen recognition of the importance of Knowledge Management (KM). The growth of interest in KM is due to the recognition that commercial assets lie in the tacit knowledge of the workforce rather than in the bricks, mortar and equipment, the traditional means of valuing corporate assets. Furthermore, the rapid turnover of staff in many modern institutions has led to recognition of the need to manage corporate knowledge in a more effective manner. Even more important has been the fact that we live in a period of information surfeit due to the Web – terms such as 'information overload' or 'infosmog' are common. This necessitates an effort to turn excessive information into focussed knowledge, in the AI sense of 'knowledge as usable information' (O'Hara 2002), and manage the knowledge we have more efficiently.

Knowledge is now seen as an integral part of the resources a business has, and as a key to maximising value and obtaining competitive advantage. The Advanced Knowledge Technologies project (www.akt.org) has identified six challenges for the use of knowledge and the technologies involved in its management:

1) Acquiring knowledge
2) Modelling knowledge
3) Reusing knowledge
4) Retrieving knowledge
5) Publishing knowledge
6) Maintaining knowledge

In this paper, we are concerned with the sixth challenging 'Maintaining Knowledge' i.e. updating the knowledge repository dynamically and identifying parts which have become out of date. It is now frequently assumed that knowledge is modelled and stored in structures called 'ontologies' which represent the 'shared conceptualisation' of a specific domain. Whether knowledge is stored in ontologies, prepositional knowledge bases or simple databases, it must be maintained and kept up to date. The world changes at an ever-increasing pace and as knowledge models the world in a manner which permits decisions to be made, so it must keep up with this changing world. This might appear at first sight an easy process – data storage is cheap, so just keep adding to the knowledge base. This is what has happened to the Web from a certain perspective in that it has continued to grow without limits and there is relatively little that is removed from it. We all suffer the consequences of this infinite process of addition.

Knowledge changes in a number of ways. There is knowledge to be added to the knowledge base, due to changes in the world or our understanding of it. There is knowledge to be removed from the knowledge base because it is out of date, untrue or merely irrelevant. There is knowledge whose accessibility needs to be changed, i.e. it has become background knowledge rather than foreground knowledge. We could categorise knowledge maintenance activities as follows:

A. **Knowledge Acquisition:** The addition of a proposition to the knowledge base, e.g.
   (1) *Two planes have hit the WTC*
B. **Knowledge realisation:** The result of processing existing knowledge, and inferring new knowledge, e.g. given the new information in (1) and the existing information in (2) we conclude (3):
   (2) *Fred Bloggs works at the WTC.*
   (3) *Fred Bloggs might have been hurt or killed.*
C. **Knowledge Foregrounding:** Certain events make a whole collection of information and knowledge come to the foreground, e.g. the destruction of the WTC brings to the foreground knowledge about the buildings, Manhattan, New York, terrorism etc.
D. **Knowledge Backgrounding:** Certain events or non-occurrence of events (including the passage of time) makes items of knowledge less accessible, less available (the process of forgetting)
E. **Knowledge deletion**(Forgeting)**:**Knowledge is deleted from the knowledge base, because it is clearly out of date, false, no longer relevant, or not used for a certain period of time, e.g. the address of WTC companies.

Each time knowledge is acquired and added to our knowledge base, its effects are not merely to add a proposition to a list. The inferences which can be drawn in the light of existing knowledge may be much more important. The challenge is how to determine what those inferences could be. Similarly, certain events place some knowledge in the background or foreground or they are forgotten/deleted, but the issue arises as to how to determine to which part of the knowledge base to perform such an action. In all these cases, essentially search and decision criteria need to be established.

2 The Frame Problem

In artificial intelligence, there has long been recognised a basic problem in the internal representations of the world known as the 'Frame Problem' (McCarthy and Hayes 1969). This was originally identified as a problem for a robot, which has to update its internal representation of the environment each time something happens. Events occur that change the state of the world and each time the robot's or system's internal representation of the world must change. The problem is that the consequences of an event are unpredictable. Thus, if a box is moved, some object may be on top of it. The movement of the box is not the only event to take into account because the movement of the other object must be accounted for as well. Yet again, one cannot easily predict this by rule (e.g. every moving object takes with it objects on top of it) because an object may be tied to the wall with string or there may be some other factor preventing the normally expected physical consequences. Clearly, a system could check every item in the world to determine if a change had (or could have) occurred or not, but this would be quite unrealistic in even a reasonably large toy world system, let alone the one we live in.

The Frame Problem is one concerning internal representations about the world. It is not a problem concerning physics or the laws of Nature. There is no question as to what the consequences of a natural event are, they just occur, and all adequate descriptions of nature must be consistent because the world is (apart from rococo quantum phenomena). The Frame Problem concerns our *descriptive knowledge* of events and the capacity (human or machine) to infer the consequences. The heart of the Frame Problem concerns where, in a model of the world, to search for consequences and when to stop. In this sense, the problem can be reduced to designing the optimum search strategy given the particular characteristics of the model of the world being used.

Clearly, given the fact that the Frame Problem concerns the update of an internal representation of the world, it is a problem very close to knowledge maintenance. Knowledge in the KM sense concerns a representation of the world, and any update to that representation potentially invokes the Frame Problem. The system has to work out which existing items of knowledge are affected by the new item of knowledge. As with the hypothetical robot, it would be impossible to check every piece of knowledge in the knowledge base, but to assume that nothing changes unless a rule requires it would be ridiculous too.

The Frame Problem has been the subject of extensive discussions in AI, yet no satisfactory resolution has been found. However, with the proposed growth of agent-based systems using the Semantic Web, this problem can no longer be safely ignored.

3 Agents Interacting

A fundamental justification for the Semantic Web, as conceived by Berners-Lee (2001) is to make it possible for software agents to operate over the machine-readable data in order to perform tasks, which could be useful for individuals and organisations. Thus, Berners-Lee *et al.* imagine a scenario where two people Pete and Lucy need to arrange a series of physical therapy sessions for their mother. Initially, Lucy's agent retrieves information from the doctor's agent, from the insurance agent, lists of therapy providers from elsewhere plus the appointment schedules from the provider's individual agents and then constructs a plan of appointments which fit the schedules of Pete and Lucy, who will drive their mother over. Pete then asks for a revision of this plan and his agent redoes the plan using stricter criteria, getting most information from Lucy's agent. This involves rescheduling some less important appointments. All this sounds idyllic (or hellish, depending on your point of view), but in order for this process to occur the Frame Problem has to be repeatedly overcome. Here are a number of problems:

1. If a therapist cancels an appointment, who will be informed? The patient obviously, but so do the agents of Lucy and Pete.
2. Lucy's agent is said to have 'complete trust' in Pete's agent 'in the context of the present task'. This implies that each task can be categorised as one to trust Pete with or not. Given a communication event from another agent requesting trusted communication, how can an agent infer that it should respond? The notion of trust is a particularly thorny one in agent research.
3. How does Pete's agent decide which appointments are 'less important'? There is nothing inherent in an appointment which enables an agent to determine its importance. Each appointment event appears to potentially trigger a Frame Problem. Human secretaries so often get it wrong, so modelling such a decision process would seem an enormous challenge.

Every time an event occurs in Lucy's, Pete's or their mother's lives, it could potentially alter something in their appointment schedule or the planned physical therapy appointments. How will the agent decide that any given event has or has not an effect? Events chain in unpredictable ways.

The situation is made more complex by imaging a situation with a potentially infinite number of agents. Given a disastrous event, to whom should an agent with knowledge of that event communicate it? There is a parallel here, between deciding which agent to communicate with, and deciding which element in one's internal representation is affected by an event. One could communicate with every agent possible, but in the real world this would be impossible in terms of communication bandwidth and time. Thus we find ourselves facing two possibilities, as with the Frame Problem, of either undertaking a 'cheap test' (all agents in a specific category e.g. 'interested in terrorist events' are informed) or using a 'sleeping dog' approach where unless my agent is specifically told in advance, it will not communicate the event to anyone.

Essentially, communication between agents is equivalent to 'communication' between items in a knowledge base, and raises the same Frame Problem issues. In the Lucy and Pete example above, if the physical therapist's agent cancels an appointment, does only Pete get informed (who would drive his mother there) or both Pete and his mother, or Lucy too? Someone (mother, Pete, Lucy) may consider the cancellation unimportant and wait for the next appointment, or may wish to reschedule and find another slot. These are decisions which it would appear difficult for an agent to make without being able to evaluate the consequences of some event or change in knowledge state.

Currently the most advanced agent based systems in laboratories have not really encountered the problem. This is partly due to limitations of size and complexity. For example, the *Electronic Elves* system of Chalupsky *et al.* (2002) consists of 15 agents in total, 9 of which are proxies for people (i.e. personal assistants). As such, the system has not 'scaled up' yet to the real world, and the authors themselves recognise as a significant research challenge "the complexity inherent in human organisations" (ibid. 2002:12). The authors specifically mention the need to adjust the autonomy of an agent in order to allow people to make 'important decisions.' The task of determining when a decision is important or not amounts to being able to resolve the Frame Problem. The authors also mention the 'co-ordination' of all the different agents as a significant research issue. This cannot be reduced to the problem of establishing an appropriate 'agent communication language', even if this is a practical prerequisite (Finin *et al.* 1998). In Kumar *et al.* (2002), a personal assistant system is presented where various agents interact in order to identify technical talks that a person might want to go to. This system includes trusted 'buddy agents' but the whole system is dependent on 'registering' one agent with another much like 'event listeners' are registered in a Java GUI. This represents a key manual intervention which circumvents the Frame Problem (what events do I treat as significant?) but naturally cannot easily scale up.

Conclusion

In this paper we have tried to show that the Frame Problem underlies key technologies in the Semantic Web. We have argued that as agents develop and attempt to communicate in a more complex and open-ended environment the Frame Problem is encountered both in updating their internal knowledge base and in deciding whom to communicate with about any change in their internal knowledge. A detailed model of the cost-benefit of exchanging information with other agents is needed, so as to guide the choice of who to contact and update for each change in the knowledge base.

Acknowledgements

Our thanks go to Professor Yorick Wilks for his comments, help and encouragement.

References

Berners-Lee, Tim, James Hendler and Ora Lassila (2001) The Semantic Web. Scientific American, May 2001,

Chalupsky, Hans, Yolanda Gil, Craig A. Knoblock, Kristina Lerman, Jean Oh, David V. Pynadath, Thomas A. Russ, and Milind Tambe (2002) Electronic Elves Agent Technologies for Supporting Human Organisations, *AI Magazine* 23:2

EU-NSF (2001) Research Challenges and Perspectives of the Semantic Web, Report of the EU-NSF strategic workshop. Sophia-Antipolis

Finin, Tim, Yannis Labrou and Yun Peng (1998) Mobile agents can benefit from standards efforts on interagent communication, in *IEEE Communications Magazine* 36:8, pp50-56

French, R. M. and P. Anselme (1999) Interactively converging on context-sensitive representations: A solution to the frame problem. *Revue Internationale de Philosophie*, 3, 365-385

Hendler, James (1999), Is There an Intelligent Agent in Your Future? *Nature* 11 March 1999

Kumar, Subhash, Anugeetha Kunjithapatham, Mithun Sheshagiri, Tim Finin, Anupam Joshi, Yun Peng, R. Scott Cost (2002) A personal agent application for the semantic web. AAAI Fall Symposium on Personalized Agents, North Falmouth, Nov 15-17, 2002

McCarthy, J. and P. J. Hayes, (1969) Some Philosophical Problems from the Standpoint of Artificial Intelligence, in *Machine Intelligence* 4, B. Meltzer and D. Michie, Eds. Edinburgh: Edinburgh University Press., pp. 463-502.

O'Hara, Kieron (2002) *Plato and the Internet,* Icon Books