The Knowledge Engineering Review, Vol. 17:1, 65–70. © 2002, Cambridge University Press
DOI: 10.1017/S0269888902000395 Printed in the United Kingdom

IEEE standard upper ontology: a progress report

ADAM PEASE and IAN NILES
Teknowledge

1 Introduction

The IEEE Standard Upper Ontology (IEEE, 2001) is an effort to create a large, general-purpose, formal ontology. The ontology will be an open standard that can be reused for both academic and commercial purposes without fee, and it will be designed to support additional domain-specific ontologies. The effort is targeted for use in automated inference, semantic interoperability between heterogeneous information systems and natural language processing applications. The effort was begun in May 2000 with an e-mail discussion list, and since then there have been over 6000 e-mail messages among 170 subscribers. These subscribers include representatives from government, academia and industry in various countries. The effort was officially approved as an IEEE standards project in December 2000. Recently a successful workshop was held at IJCAI 2001 to discuss progress and proposals for this project (IJCAI, 2001).

There are currently two ontologies that have been proposed to the group. One is an upper-level ontology (Niles & Pease, 2001; Teknowledge, 2002) that was created by merging publicly available formal content into a single, cohesive structure. This is known as the SUMO (Suggested Upper Merged Ontology). The other is a meta-ontology based on category theory (Kent, 2001), and it is known as the IFF Foundation Ontology. This paper addresses the SUO effort mainly from the perspective of the SUMO. Note also that this paper reflects the personal views of the authors and not necessarily any consensus of the IEEE SUO group.

There have been many lively discussions on the SUO mailing list regarding various philosophical, logical, mathematical and engineering issues. Some of these issues have had a direct impact on the documents, and other issues are still unresolved. In what follows, we will attempt to highlight some of the most interesting and/or relevant debates. Given the limits of space in this article, no attempt will be made to attribute views to particular participants. Interested readers are encouraged to examine the public email archive for the group at `http://suo.ieee.org/email/thrd1.html`.

2 Single ontology versus lattice of theories

One of the recurring debates on the SUO mailing list regards the overall structure of the standard ontology. According to some, the ontology should be a single, consistent structure. According to others, the ontology should be a lattice of theories that are internally consistent but may be inconsistent with theories that are not part of the same path through the lattice. The argument for creating a single ontology has four main points:

- A multiple ontology standard is more difficult to maintain, because more than one knowledge model has to be supported.
- A multiple ontology standard is more difficult to use, because it requires users to make more choices.
- The one comparable effort to create a large ontology, Cyc, does not have alternative theories (although Cyc's microtheory structure would support alternative theories).

- No one has yet shown that there are two logically inconsistent and equally valid theories that ought to be included in an engineering-focused ontology.

The argument for a lattice of theories is that:

- There are multiple, incompatible and equally valid views on how to model the world. We cannot legislate a single choice.
- There is political issue of acceptance. Groups who have one viewpoint will not use an ontology that assumes a different viewpoint.

Each of these views has found a place in one of the proposed documents. The SUMO embodies the view that we can construct a single, consistent, upper-level ontology that will be adequate for real applications. The IFF Foundation Ontology is a formal theory that is capable of representing and reasoning about relations between various ontologies.

3 Endurantism versus perdurantism

Another central debate on the SUO mailing list has concerned the representation of objects and processes. According to the so-called endurantist view, there is a basic, categorial distinction between objects and processes while, according to the perdurantist view, paradigmatic processes and objects are merely opposite ends of a contiuum of spatio-temporal phenomena. Endurantism posits that objects, unlike processes, are completely present at any moment of their existence, while perdurantism regards everything as a space-time worm (or a slice of such a worm) (Loux, 1980).

Endurantism and perdurantism have their respective advantages and disadvantages. The advantage of perdurantism is its simplicity (since everything is treated alike, as a space-time worm), and its disadvantage is that many find it counterintuitive (since objects and processes are not distinguished and thus things that are typically regarded as objects have temporal parts). On the other hand, according to many philosophers, the advantage of endurantism is that it captures the intuitive distinction between objects and processes. A disadvantage of endurantism is that it appears to be inconsistent with an almost self-evident principle about the meaning of identity, known as the indiscernability of identicals. Suppose we have an unfortunate soul named Joe who loses an arm at a point in time (which we shall call "T"). Then, since an object is wholly present at any moment of its existence (according to endurantists), we know that Joe is identical with JoeBeforeT which, in turn, is identical to JoeAfterT. However, according to the indiscernability of identicals, something A is identical with something B only if every property that can be ascribed to A can be ascribed to B and vice versa. Since Joe has the property of having an arm before T and does not have this property after T, it follows that JoeAfterT is not identical with JoeBeforeT. Thus we seem to generate a contradiction from endurantism.

There is a range of responses and counterresponses that can be offered at this point with regard to both the possible inconsistency of endurantism and the debate between endurantism and perdurantism. Another approach is to argue that, although there are deep theoretical differences between endurantism and perdurantism, these differences would not actually lead to a contradiction in any of the envisioned applications of the standard ontology.

4 Possible worlds

There is ongoing discussion on whether to include mechanisms for modal logic (Chellas, 1980), and/or contextual reasoning (McCarthy, 1993). Several alternatives have been proposed including Lewis's (1986) and Kripke's (1963) possible worlds semantics and Dunn's (1973) "Laws and facts". The primary point of contention with this subject is that, while this logical apparatus can add to the expressiveness of the SUO, it can also complicate formal reasoning and the semantics of the entire ontology. The cost and benefit trade-off here has not yet been resolved, and there are currently no modal notions in either of the SUO documents.

5 Semiotics

The discussion relating to semiotics has been less controversial than the other areas discussed in this paper, but it is worth summarising here, because it has resulted in a well-defined set of concepts that are not available in any other ontology as far as the authors of this paper are aware. This area incited less controversy after the group developed some concrete examples of the relevance of semiotics concepts to reasoning in intellectual property issues and media. A set of examples has motivated development of this section of the ontology:

- an edition of the printed play Hamlet,
- a copy of an edition of Hamlet,
- a performance of Hamlet,
- a performance of Hamlet captured on video and
- the timeless informational content of the play Hamlet.

These examples resulted in the addition of a number of relations to one document. These include:

- equivalentContentClass – a BinaryRelation that relates two subclasses of ContentBearingObject. “(equivalentContentClass ?CLASS1 ?CLASS2)” means that the content expressed by each instance of ?CLASS1 is also expressed by each instance of ?CLASS2, and vice versa. An example would be the relationship between English and Russian editions of Agatha Christie’s *Murder on the Orient Express*.
- containsInformation – a subrelation of “represents”. This predicate relates a ContentBearingObject to the Proposition that is expressed by the ContentBearingObject. Examples include the relationships between a physical novel and its story and between a printed score and its musical content.
- realisation – a subrelation of “represents”. “(realization ?PROCESS ?PROP)” means that ?PROCESS is a Process which expresses the content of the Proposition, ?PROP. Examples include a particular musical performance, which realises the content of a musical score, or the reading of a poem.
- equivalentContentInstance – a binary relation between two instances of “Representation”. “(equivalentContent ?OBJ1 ?OBJ2)” means that the content expressed by ?OBJ1 is identical with the content expressed by ?OBJ2. An example would be the relationship between a handwritten draft of a letter to one’s lawyer and a typed copy of the same letter.
- subsumesContentInstance – a binary relation between two instances of “Representation”. “(subsumesContent ?OBJ1 ?OBJ2)” means that the content expressed by ?OBJ1 contains the content expressed by ?OBJ2. An example is the relationship between a handwritten poem and one of its stanzas.

6 Groups and organisations

Another debate on the SUO mailing list has concerned the notions of organisation and position. Several criteria for any adequate formalisation of organisation have been proposed by various participants, and these criteria can be summarised as follows:

1. An organisation must have agency, e.g. it exhibits intentionality, and it has rights, responsibilities and obligations.
2. An organisation must have temporal extent. It comes into being at a certain point in time and it goes out of existence at another point.
3. An organisation may have members but it should not be required to have members. There are many examples of organisations, e.g. corporations and churches, that have assets, are liable for certain claims and so on, even though they may have no members.

In the context of the SUMO, a proposal has been developed which appears to address all of these criteria. The proposal is to create a term “OrganisationUnit”, which is borrowed from Fox *et al*. (1995), an immediate subclass of the notion of “CognitiveAgent” and then to make “Organisations” and “Positions” within “Organisations” immediate subclasses of “OrganisationUnit”.

Together with this structure of concepts, a set of predicates is defined: “occupiesPosition” (relates a person to the position they occupy within an organisation), “subOrganisations” (relates one organisation to another of which it is a part), and “employs” (relates a person to the organisation of which he/she is a member).

This structure of concepts and these predicates appear to satisfy the three criteria listed above. Consider criteria 1 and 2. “Organisation” is a subclass of “Agent” (since “CognitiveAgent” is a subclass of “Agent”) and ‘Agent’ is a space-time object, so organisations are agents and they have a position in space-time. Consider criterion 3. The sticking point here is the stipulation that organisations are not required to have members. In the structure above, we distinguish two senses of organisation: “OrganisationUnit” and “Organisation”. The latter is required to have members, because it is a subclass of “Collection”, which is the class of all aggregations of objects that have a location in space-time. However, “OrganisationUnit” is not a subclass of “Collection” because it is meant to cover both organisations and positions, so there is no requirement that an “OrganisationUnit” have any members. Accordingly, we now have a notion of organisation that covers “empty” organisations.

7 Agent communication

The development of a large standard ontology with formal semantics has the potential for significant impact on agent technology. For computer-based agents to have the same degree of autonomy, flexibility and complexity as human agents, they will need languages for communication that are comparable in scope and expressivity to those used by human beings. Agent communication languages such as KQML (Finin *et al*., 1997) have specified some of the essential parameters of communication acts. As part of the SUMO development process, types of speech acts have been incorporated into the ontology under the node “Communication”. Furthermore, the SUMO as a whole specifies the possible ontological content for those acts. It is in this content that the semantics for any intelligent communication will reside.

Part of the promise of agent technology is that agents will be able to engage in diverse and unscripted interactions with other agents in order to accomplish their respective goals. Many agent-based systems are no more advanced than “conventional” modular, distributed software systems, at least with respect to the flexibility and richness of the communicative interactions between the agents. Furthermore, the addition of new agents or new agent capabilities can require revision of the entire, embedding system. This is a situation commonly found in software integration projects, where the inclusion of new requirements can force drastic changes to the interfaces between components. The introduction of new realities can break older simplifying assumptions. By using an upper ontology like the SUMO that teases out implicit assumptions in advance, flexibility and extensibility are maximised.

8 Conclusion

So far we have discussed some of the historical issues in the development of the SUMO and the potential impact of the work. There are several significant areas of future work involving the SUMO.

One project is mapping the SUMO to the WordNet lexical database (Niles, 2002). This involves associating every node in the WordNet (Miller *et al*., 1993) lexicon with its most closely related SUMO concept. The benefits of this project are threefold. First, it will enable researchers in natural language processing and information retrieval to make use of the SUMO for sense disambiguation, term expansion to a structured vocabulary and automated document summarisation. A second benefit is that the mapping process will function as a completeness check on the content of the SUMO. Whenever we find a domain-independent node in WordNet that is not adequately represented in the SUMO, we add additional terms and/or axioms to the ontology. A final benefit is that this work should enable improved navigability and understanding of the SUMO, because it will enable the users to search using our ontology browser on almost any English word. The browser will locate the SUMO concepts most closely related to the input words and return them for the user's inspection. Although the benefits are potentially great, it is important to realise that this is a long-range task. At present about 90% of the over 60,000 WordNet noun nodes have been mapped.

A second area of future work involves creating domain-specific extensions of the SUMO. One such extension is a "Quality of Service" ontology that provides definitions for computer operating system, network and hardware concepts. The area of application is in creating contracts for delivery of a particular type and level of Internet or computing service. Another effort was a study to expand the definitions for an ontology of e-commerce services begun as part of the DARPA DAML project (DAML, 2001). Previous work (Cohen *et al*., 1999) has shown that there is a balance in any particular ontology application between the use of upper ontology and domain-specific ontology content. The development of domain-specific ontologies will help to debug the upper ontology from the bottom up, and it will provide application-ready content that, as much as possible, inherits from a common stock of general concepts.

Another area of work involves the creation of tools that use the SUMO. Some tools that have already been created include an ontology browser, a translator to the DARPA Agent Markup Language (DAML), a translator to Protégé (Musen *et al*., 2000), and tools that support collaborative ontology development.

The IFF document is also undergoing continued development. The group has not decided whether IFF or SUMO will become the standard or whether they may be merged in some way.

While completion of the formal standards process may take several years, the ontology is mature enough at this point to warrant involvement of a wider community in the development and use of the SUMO. This involvement could come in the form of criticism of the existing SUMO; development of proposed extensions to the ontology; contribution of compliant, domain-specific ontologies; implementation of the SUMO in knowledge-based reasoning applications; adoption of the SUMO as an interlingua in agent-based systems; and use of the SUMO as a reference for specifying an information integration architecture.

References

Chellas, B, 1980, *Modal Logic: An Introduction,* Cambridge University Press.

Cohen, P, Chaudhri, V, Pease A and Schrag, R, 1999, "Does prior knowledge facilitate the development of knowledge based systems" *Proceedings of the Sixteenth National Conference on Artificial Intelligence* (AAAI-1999) pp-pp.

DAML, 2001, DARPA Agent Markup Language Services Ontology, release 0.5, `http://www.daml.org/services/daml-s/2001/05/`

Dunn, J, 1973, "A truth value semantics for modal logic" in H Leblanc (ed.) *Truth, Syntax and Modality* North-Holland.

Finin, T, Labrou Y and Mayfield, J, 1997, "KQML as an agent communication language" in J Bradshaw (ed) *Software Agents* MIT Press.

Fox, MS, Barbuceanu, M and Gruninger, M, 1995, "An organisation ontology for enterprise modeling: preliminary concepts for linking structure and behavior" *Proceedings of the Fourth Workshop on Enabling Technologies, Infrastructure for Collaborative Enterprises* pp-pp.

IEEE, 2001, The IEEE Standard Upper Ontology Working Group (P1600.1) website `http://suo.ieee.org`

IJCAI, 2001, The IJCAI-2001 Workshop on the IEEE Standard Upper Ontology website `http://projects.teknowledge.com/IJCAI01/index.html`

Kent, R, 2001, "The IFF category theory ontology" *Working Notes of the IJCAI-2001 Workshop on the IEEE Standard Upper Ontology* pp. 221–226.

Kripke, S, 1963, "Semantical considerations on modal logic" *Acta Philosophica Fennica* **16** 83–94.

Loux, M, 1998, *Metaphysics: A Contemporary Introduction* Routledge.

Lewis, D, 1986, *On The Plurality of Worlds* Blackwell.

McCarthy, J, 1993, "Notes on formalizing contexts" *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence* (IJCAI-1993) 555–560.

Miller, GA, Beckwith, R, Fellbaum, C, Gross, D and Miller, K, 1990, "Introduction to WordNet: an on-line lexical database *International Journal of Lexicography* **3**(4) 235–244.

Musen, M, Fergerson, R, Grosso, W, Noy, N, Crubezy, M and Gennari, J, 2000, "Component-based support for building knowledge-acquisition systems" *Proceedings of the Conference on Intelligent Information Processing* (IIP 2000) *of the International Federation for Information Processing World Computer Congress* (WCC, 2000).

Niles, I, 2002 (forthcoming), Mapping an Upper Ontology to WordNet.

Niles, I and Pease, A, 2002, "Towards a standard upper ontology" *Proceedings of The Second International Conference on Formal Ontology in Information Systems* (FOIS-2001); available at `http://projects.teknowledge.com/HPKB/Publications/FOIS.pdf`.

Teknowledge, 2002, Ontology website `http://ontology.teknowledge.com`.