

Towards Subject-Centric Merging of Ontologies

Jack Park
SRI International
jack.park@sri.com
Patrick Durusau
patrick@durusau.net

The challenge of Cyberinfrastructure is to integrate relevant and often disparate resources to provide a useful, usable, and enabling framework for research and discovery characterized by broad access and “end-to-end” coordination

–NSF Workshop on Cyberinfrastructure
for the Social Sciences, 2005¹

Extended Abstract

We are developing a *subject-centric* approach to federation of heterogeneous representations of subjects. Subjects are represented in many ways, including through the formal structures of ontologies. Here, we will sketch our approach to semantic interoperability among heterogeneous world views through implementations of subject maps, and briefly mention that such a subject map is being considered for installation as a plugin to the Protégé platform.

Our work exists in a field rich in experience and motivation for semantic interoperability among heterogeneous ontologies. As a sketch of that field, John Madden [1] listed “three topologies for semantic interoperability:”

1. Central semantic authority
2. Hierarchical semantics
3. Federated semantics

We consider *central semantic authority*, and *federated semantics* to be of greatest contrast in our work. Since the approach we present lies in the *federated semantics* domain and involves mapping, the key points that Madden made about semantic federation are most relevant to our work. He points out that responsibility for mapping is finely divided, which, to us, implies that many individuals and groups will contribute to the final mappings, that quality depends on peer-to-peer collaboration, that there are no global guarantees, and that there is a need to support a “market” for ontology fragments

Central semantic authority reminds of “Hobson’s Choice,” where a central authority grants the right to rent any horse so long as it is the horse closest to the door. In [2], we spoke to the opportunity to determine, among different world views, which subjects are being represented, and to perform merging when subject sameness is determined. We argue that a subject map provides appropriate facilities for performing subject-centric merging and marshalling the remaining, unmerged ontological entities for reference, navigation, and completeness. In this process, we argue that no particular world view

¹ NSF: <http://vis.sdsc.edu/sbe/>

gains privilege over any other; all world views are presented uniformly and each representation is captured in any merge process without loss of information.

Consider that the XML topic maps paradigm (XTM) [3] and [4] has served the topic mapping community well, and continues to do so. Topic maps do not replace ontologies. Rather, they augment ontologies and other world views. As a brief introduction to subject maps, consider that, as we begin to apply topic mapping to the complex use cases of, for instance, bioinformatics, where subject identity is under-specified by XTM, we need a framework that facilitates a finer-grained approach to subject identity. A framework for such a specification is known as the Topic Maps Reference Model (TMRM) [5], and we have begun to label implementations of the TMRM as *subject maps* to distinguish them from their siblings. The TMRM makes no specification of the means by which subjects are identified or ways in which subject proxies, as they are called, are merged. Rather, the TMRM leaves subject map authors free to make their own design decisions, but it specifies that subject map authors are required to *disclose* the design decisions, the ontological commitments they have made such that other implementations can create means by which merging among different subject maps can be afforded. Disclosures, in the TMRM, form a *legend* for the map, much as street maps have legends to explain the artifacts represented by the map. Steve Newcomb had this to say² about disclosures:

The Topic Maps Reference Model is our attempt to set forth a checklist of things that must be disclosed about a given body of knowledge, regardless of how that body is represented, in order to enable a specific benefit to be realized. The benefit is facilitation of the task of integrating that body of knowledge with other such bodies, on a subject-by-subject basis. I like to say that the disclosures amount to descriptions of subject address spaces.

Implementations of the TMRM create ways in which determination of subject sameness can be evaluated among classes and properties found in different ontologies. When subject sameness is found, those classes or properties are merged and assertions can be made on related classes to support further merging opportunities. When ontological entities merge in a subject map, the statements made by those entities, that is their properties and relations are carried with them.

In some sense, there are ontological commitments made in subject mapping. But, they are made with the specific intent to facilitate the semantic integration of world views which are always the product of ontological commitments. The nature of this facilitation lies not in the ontological commitments made by any author; rather, merging is facilitated on the basis of determination of subject identity, which, in many cases, is underspecified in the constituent world views. This leaves open much room for the peer-to-peer collaboration mentioned in Madden's presentation. But, in the case where sufficient evidence for subject identity sameness is found, merging can be handled by the subject map engine. Steve Newcomb had this to say about ontological commitments in the TMRM:

It's in the nature of what we've been trying to do that the semantics be left undefined. Whenever we've recognized that we've been making semantic

² Steve Newcomb: Personal communication, January 24, 2006

assumptions, we've ruthlessly expunged them. We're endeavoring to create conditions favorable for discovery of relevant information expressed in terms of diverse universes of discourse, and across those universes of discourse. We need to be both inviting to knowledge resources created in the light of diverse semantic systems, cultures and communities of practice, and, at the same time, even-handed with respect to all of them. Among other things, we're looking to create a better marketplace for ideas, with more opportunities to add value for anyone with value to add.

Benefits of federation through subject-centric merging derive from viewing same-subject ontological entities together in a single subject proxy. A subject proxy is the name given to a kind of *container* for all of the properties marshaled as representations of a single subject. We have characterized [2] a prime benefit as the emergence of "worm holes" between different world views. For instance, when a particular ontology provides certain properties to a merged subject proxy not provided by another entity merged into that proxy, the fact that each property is identified with the ontology from which it comes provides cognitive links between different ontologies; Exploration of different ontologies provides opportunities for chance discovery.

Light weight implementations of the TMRM are beginning to appear. For instance, the first author is creating an implementation called TopicSpaces that is initially being applied to social bookmarking applications. It is developed using Java and is being considered for use as a plugin agent for Protégé in service of ontology federation projects.

References

- [1] Madden, John, "Sharing healthcare meaning is hard to do," Slides presented at the Face-to-Face W3C HCLSIG Meeting, January 25-26, 2006, Cambridge, MA. On the web at <http://www.w3.org/2001/sw/hcls/f2f-2006/F2F-Agenda.html>
- [2] Park, Jack, and Patrick Durusau, "Avoiding Hobson's Choice In Choosing An Ontology", Teleconference presentation to the Ontolog Community. On the web at http://ontolog.cim3.net/cgi-bin/wiki.pl?ConferenceCall_2006_04_27
- [3] XTM Authoring Group, "XML Topic Maps (XTM) 1.0," on the web at <http://www.topicmaps.org/xtm/index.html>
- [4] Park, Jack, and Sam Hunting (Editors), *XML Topic Maps: Creating and Using Topic Maps for the Web*, 2002, Addison-Wesley.
- [5] Durusau, Patrick, Steve Newcomb and Robert Barta, "Topic Maps Reference Model ISO 13250-5," available on the web at <http://www.jtc1sc34.org/repository/0710.pdf>