What do we need for ontology integration on the Semantic Web
Position statement

Natasha F. Noy
Stanford University
251 Campus Drive, Stanford, CA 94305, USA
noy@smi.stanford.edu

Ontologies, integration, and the Semantic Web

In order for the Semantic Web participants to share information, they must have some agreement on what elements in their shared domain of interest exist and how these elements can relate to one another. A formal specification of such an agreement is called an **ontology**. An ontology for a domain enumerates and gives semantic descriptions of concepts in the domain of discourse, defining domain-relevant attributes of concepts and various relationships among them. For example, an ontology describing a wines will include such concepts as vintages, wine regions, wineries, grape varieties, and so on. It will also include relations such as produced by, made from, color, year, and body of wine.

The ultimate vision of using formal ontologies is to develop a single ontology or a small set of ontologies that everyone will conform to. Alternatively, on a smaller scale, there could be single organization-wide ontologies. Semantic integration then becomes a much easier, if not a trivial, problem since everyone shares the same set of definitions.

However, such a vision seldom, if ever, becomes a reality. Just as there is more than one Web directory (e.g., Yahoo!, ODP, etc.), more than one shopping site, more than one search engine on the today's web, there will be more than one ontology even for the same domain on the Semantic Web. The reasons for this diversity are both technical and non-technical.

On the technical side, the task for which ontology is going to be used greatly influences ontology design. For example, an ontology supporting an application of pairing wines with food is unlikely to have properties describing numbers of bottles or their exact prices, which is something an ontology supporting an inventory application for a restaurant will need. Furthermore, one ontology may classify wines based on the grapes that are used to produce them and another may use the region that the wine comes from as the classification criterion.

On the non-technical side, there are often cultural, organizational, or administrative reasons why, for example, different departments in an organization might undertake their own ontology-development efforts. These reasons range from the NIH (not invented here) syndrome to practical considerations such as having current software depend heavily on a particular ontology.

Therefore, **integration of ontologies** is a major challenge and research issue on the Semantic Web.

Challenges in ontology integration

Some of the specific challenges in ontology integration that we must address in the near future are:

- finding similarities and differences between ontologies in automatic and semi-automatic way
- defining mappings between ontologies
- developing an ontology-integration architecture
- composing mappings across different ontologies
- representing uncertainty and imprecision in mappings

In the Semantic Web, there will be multiple ontologies that will be developed independently but will interact with one another. These ontologies might reuse other ontologies and therefore share some of their content and frame of reference. They may make some changes to ontologies they are reusing, declare equivalence between their terms and terms in other ontologies, and so on.

The first challenge is to find similarities and differences between the ontologies in automatic or semi-automatic way. Differences could be as simple as the use of synonyms for the same concept. For example, one ontology may use the term "vintage" and another may use the term "year". There could be differences in the way ontologies organize concepts. For instance, one ontology can classify all wines based on their color, having Red, White and Rosé as the top-level categories. Another ontology can have color as a property of the wine. It may never be possible to find all mappings between ontologies in a completely automatic way since some of the intended semantics can only be discerned by humans. However, ontology-integration on the large scale will be possible *only* if we can make significant progress in identifying mappings automatically or semi-automatically.

Researchers have already made some progress in this direction. For example, Hovy and colleagues (1998) describe a set of heuristics that researchers at ISI/USC used for semi-automatic alignment of domain ontologies to a large central ontology. Their techniques are based mainly on linguistic analysis of concept names and natural-language definitions of concepts. PROMPT (Noy & Musen 2003) uses the structure of ontology definitions and the structure of a graph representing an ontology to suggest to ontology designers which concepts may be related. GLUE (Doan *et al.* 2002) applies machine-learning techniques to instance data conforming to ontologies to find related concepts.

Once we find the mappings, we need to define a formalism for representing them that would enable and facilitate various tasks that use the mappings. These task include (but are not limited to) the following:

- answering queries posed to one ontology in terms of another ontology
- transforming instance data conforming to one ontology into another ontology
- using one ontology to drive an application developed based on another ontology

One approach to expressing the mapping information is to use the statements in the ontology language itself to express the correlation. OWL for example, has such statements as `owl:sameClassAs` and `owl:samePropertyAs` that allows one to "bridge" two ontologies. A reasoning engine can then treat two ontologies as a single theory. Another approach is to express mappings as instances of concepts in a mapping ontology. Crubezy (2003) for example have developed such an ontology, which enables specification of extremely expressive mappings, including ones that require recursive definitions. More research is needed however to determine which approaches would best support specific integration tasks.

The next research issue is finding an optimal architecture for ontology integration. One possible architecture could be similar to information-integration architectures in which there is a global ontology which serves as an interface to a number of local ontologies (Genesereth *et al.* 1997; Calvanese *et al.* 2001). Queries are posed to the global ontology which translates them to the terms in the local ontologies. The drawback of such an architecture is the need to develop and, more important, agree on the global ontology.

Another possibility is a peer-to-peer architecture in which we create pairwise mappings between ontologies (Halevy *et al.* 2003). Compared to the global-ontology architecture, the number of mappings that we need to create is n^2 where n is the number of ontologies, compared to n mappings to the global ontology. At the same time, the peer-to-peer architecture preserves the de-centralized nature of the Semantic Web. We may not always need to map between each pair of ontologies and therefore in practice the number of mapping can be significantly smaller than n^2.

Reusing the mappings leads to the problem of mapping composition. Suppose we have two mappings: one mapping is between ontologies A and B and another one is between ontologies B and C. Can we use these mappings to derive the mapping between ontologies A and C? Can we compose the mappings in computationally complete and efficient way?

In many cases, in particular when using automatic means to find mappings, we may not be able to define mappings precisely. Sometimes a precise mapping simply will not exist. For example, one classification of wines may only have red and white wines (classifying rosé wines as white wines). Another ontology may have a separate class for rosé wines. This class in the second ontology will not have an exact counterpart in the first. A precise mapping may exist but our means for finding it automatically will not be able to find it but will suggest several likely candidates instead. And in some cases, we do not need precise mappings and knowing that a class A in one ontology is a subclass of a class B in another is sufficient. Challenges in these area include not only classifying and representing different types and sources of imprecise and approximate mappings but also using this information for such tasks as discovering new mapping information or performing reasoning services across the mapped ontologies.

While researchers are actively working on some of these challenges in ontology integration, they have only scratched the surface. The unique scale, decentralization, and lack of central control in the Semantic Web require significant new advances to make ontology integration possible on the Web scale.

References

Calvanese, D.; Giacomo, G.; and Lenzerini, M. 2001. Ontology of integration and integration of ontologies. In *Description Logic Workshop (DL 2001)*, 10–19.

Crubézy, M. 2003. Mediating knowledge between application components. In *Workshop on Semantic Integration at ISWC-2003.*

Doan, A.; Madhavan, J.; Domingos, P.; and Halevy, A. 2002. Learning to map between ontologies on the semantic web. In *The 11th Intl WWW Conference.*

Genesereth, M. R.; Keller, A. M.; and Duschka, O. 1997. Infomaster: An information integration system. In *ACM SIGMOD Conference.*

Halevy, A. Y.; Ives, Z. G.; Suciu, D.; and Tatarinov, I. 2003. Schema mediation in peer data management systems. In *19th Intel Conf on Data Engineering (ICDE).*

Hovy, E. 1998. Combining and standardizing largescale, practical ontologies for machine translation and other uses. In *First International Conference on Language Resources and Evaluation (LREC)*, 535–542.

Noy, N. F., and Musen, M. A. 2003. The PROMPT suite: Interactive tools for ontology merging and mapping. *International Journal of Human-Computer Studies, in press.*