# Semantic Bridging of Independent Enterprise Ontologies

Michael N. Huhns and Larry M. Stephens
*University of South Carolina, USA, huhns@sc.edu*

Abstract: Organizational knowledge typically comes from many independent sources, each with its own semantics. This paper describes a methodology by which information from large numbers of such sources can be associated, organized, and merged. The hypothesis is that a multiplicity of ontology fragments, representing the semantics of the independent sources, can be related to each other automatically *without* the use of a global ontology. That is, any pair of ontologies can be related indirectly through a *semantic bridge* consisting of many other previously unrelated ontologies, even when there is no way to determine a direct relationship between them. The relationships among the ontology fragments indicate the relationships among the sources, enabling the source information to be categorized and organized. A preliminary evaluation of the methodology has been conducted by relating 53 small, independently developed ontologies for a single domain. A nice feature of the methodology is that common parts of the ontologies reinforce each other, while unique parts are de-emphasized. The result is a *consensus* ontology.

## 1. INTRODUCTION

Corporate information searches can involve data and documents both internal and external to the organization. The research reported herein targets the following basic problem: a search will typically uncover a large number of independently developed information sources—some relevant and some irrelevant; the sources might be ranked, but they are otherwise unorganized, and there are too many for a user to investigate manually. The problem is familiar and many solutions have been proposed, ranging from requiring the user to be more precise in specifying search criteria, to constructing more intelligent search engines, or to requiring sources to be more precise in describing their contents. A common theme for all of the approaches is the use of ontologies for describing both requirements and sources. Unfortunately, ontologies are not a panacea unless everyone adheres to the same one, and no one has yet constructed an ontology that is comprehensive enough (in spite of determined attempts to create one, such as the Cyc Project, underway since 1984). Moreover, even if one did exist, it probably would not be adhered to, considering the dynamic and eclectic nature of the Web and other information sources.

There are three approaches for relating information from large numbers of independently managed sites: (1) all sites will use the same terminology with agreed-upon semantics (improbable), (2) each site will use its own terminology, but provide translations to a global ontology (difficult, and thus unlikely), and (3) each site will have a small, local ontology that will be related to those from other sites (described herein). We hypothesize that the small ontologies can be related to each other automatically *without* the use of a global ontology. That is, any pair of ontologies can be related indirectly through a *semantic bridge* consisting of many other previously unrelated ontologies, even when there is no way to determine a direct relationship between them. Our methodology relies on sites that have been annotated with ontologies (Pierre, 2000); such annotation is consistent with several visions for the semantic Web (Heflin and Hendler, 2000; Berners-Lee, et al. 2001). The domains of the sites must be similar—else there would be no interesting relationships among them—but they will undoubtedly have dissimilar ontologies, because they will have been annotated independently.

Other researchers have attempted to merge a pair of ontologies in isolation, or merge a domain-specific ontology into a global, more general ontology (Wiederhold, 1994). To our knowledge, no one has previously tried to reconcile a large number of domain-specific ontologies. We have evaluated our methodology by applying it to a large number of independently constructed ontologies.

## 2. RECONCILING INDEPENDENT ONTOLOGIES

In agent-assisted information retrieval, a user will describe a need to his agent, which will translate the description into a set of requests, using terms from the user’s local ontology. The agent will contact on-line brokers and request their help in locating sources that can satisfy the requests. The agents must reconcile their semantics in order to communicate about the request. This will be seemingly impossible if their ontologies share no concepts. However, if their ontologies share concepts with a third ontology, then the third ontology might provide a “semantic bridge” to relate all three. Note that the agents do not have to relate their entire ontologies, only the portions needed to respond to the request.

The difficulty in establishing a bridge will depend on the semantic distance between the concepts, and on the number of ontologies that comprise the bridge. Our methodology is appropriate when there are large numbers of small ontologies—the situation we expect to occur in large and complex information environments. Our metaphor is that a small ontology is like a piece of a jigsaw puzzle, as depicted in Fig. 1. It is difficult to relate two random pieces of a jigsaw puzzle until they are constrained by other puzzle pieces. We expect the same to be true for ontologies.

Ontologies can be made to relate to each other like pieces of a jigsaw puzzle. (Top) Two ontology fragments with no obvious relationships between them. (Bottom) The introduction of a third ontology reveals equivalences between components of the two original ontology fragments

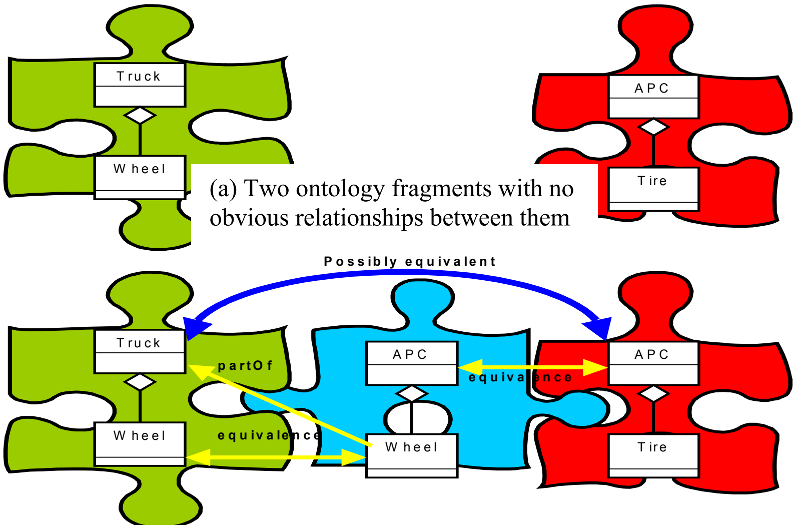
Two concepts can have the following seven mutually exclusive relationships between them: *subclass*, *superclass*, *equivalence*, *partOf*, *hasPart, sibling*, or *other*. If a request contains three concepts, for example, and the request must be related to an ontology containing 10 concepts, then there are 7x3x10 = 210 possible relationships among them. Only 30 of the 210 will be correct, because each of the three concepts in the request will have one relationship with each of the 10 concepts in the source's ontology.

The correct ones can be determined by applying constraints among the concepts within an ontology, and among multiple ontologies. Once the correct relationships have been determined, we make use of *equivalence* and *sibling* or, where those do not exist, the most specific *superclass* or *partOf*.

In Fig. 1, the ontology fragment on the left would be represented as *partOf(Wheel, Truck),* while the one on the right would be represented as *partOf(Tire, APC).* There are no obvious equivalences between these two fragments. The concept *Truck* in the first ontology could be related to *APC* in the second by *equivalence, partOf, hasPart, subclass, superclass,* or *other*. There is no way to decide which is correct. When the middle ontology fragment *partOf(Wheel, APC)* is added, there is evidence that the concepts *Truck* and *APC*, and *Wheel* and *Tire* could be *equivalent*.

This example exploits the existence of the relation *partOf*, which is common to all three ontologies. Other domain-independent relations, such as *subclassOf*, *instanceOf*, and *subrelationOf*, will be necessary for the reconciliation process. Moreover, the reflexivity, symmetry, asymmetry, transitivity, irreflexivity, and antisymmetry properties are needed for relating occurrences of the relations to each other (Stephens and Chen 1996). Domain concepts and relations can be related to each other by converse/inverse, composition, (exhaustive) partition, part-whole (with 6 subtypes), and There must be some minimum set of these fundamental relations that are understood and used by all local ontologies and information system components.

In attempting to relate two ontologies, a system might be unable to find correspondences between concepts because of insufficient constraints and similarity among their terms. However, trying to find correspondences with other ontologies might yield enough constraints to relate the original two ontologies. As more ontologies are related, there will be more constraints among the terms of any pair, which is an advantage. It is also a disadvantage in that some of the constraints might be in conflict. We make use of the preponderance of evidence to resolve these statistically.



(a) Two ontology fragments with no obvious relationships between them

(b) The introduction of a third ontology reveals equivalences between components of the original two ontology fragments

Figure 1. Ontologies can be made to relate to each other like pieces of a jigsaw puzzle

## 3. EXPERIMENTAL METHODOLOGY

We asked each of 53 graduate students in computer science, who were novices in constructing ontologies, to construct a small ontology for the Humans/People/Persons domain. The ontologies were required to be written in DAML and to contain at least 8 classes with at least 4 levels of subclasses; a sample ontology is shown in Fig. 2.

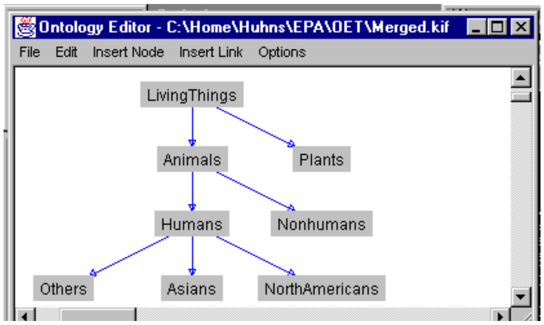
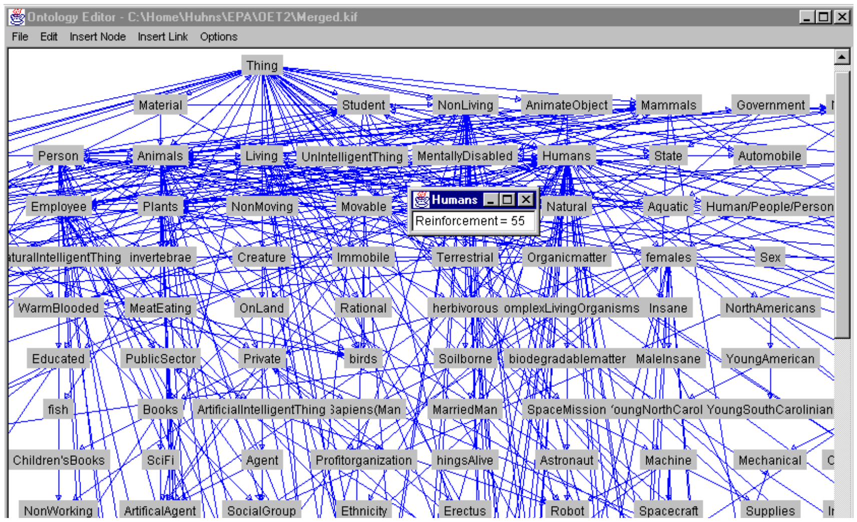


Figure 2. A typical small ontology used to characterize an information source about people (all links denote subclasses)

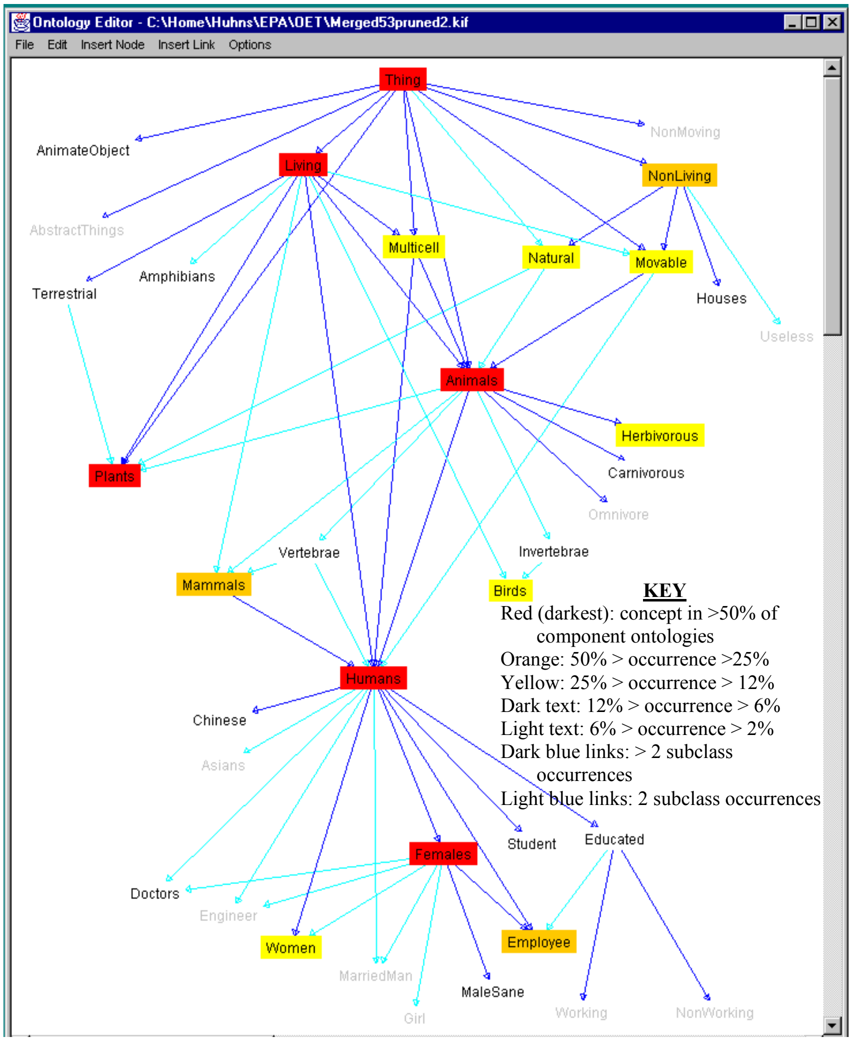
Using string-matching and other heuristics, we merged the 53 component ontologies. The component ontologies described 864 classes, while the merged ontology contained 281 classes in a single graph with a root node of the DAML concept #Thing. All of the concepts were related, i.e., there was some relationship (path) between any pair of the 281 concepts (see Fig. 3).



Reinforcement = 55

Figure 3. A portion of the ontology formed by merging 53 independently constructed ontologies for the domain Humans/People/Persons. The entire ontology has 281 concepts related by 554 subclass links

Next, we constructed a *consensus ontology* by counting the number of times classes and subclass links appeared in the component ontologies when we performed the merging operation. For example, the class Person and its matching classes appeared 14 times. The subclass link from Mammals (and its matches) to Humans (and its matches) appeared 9 times. We termed these numbers the “reinforcement” of a concept.



**KEY**

Red (darkest): concept in >50% of component ontologies

Orange: 50% > occurrence >25%

Yellow: 25% > occurrence > 12%

Dark text: 12% > occurrence > 6%

Light text: 6% > occurrence > 2%

Dark blue links: > 2 subclass occurrences

Light blue links: 2 subclass occurrences

Figure 4. The final consensus ontology formed by merging concepts with common subclasses and superclasses. The resultant ontology contains 36 concepts related by 62 subclass links.

Redundant subclass links were removed and the corresponding transitive closure links were reinforced. That is, if C has subclass A with reinforcement 2, C has subclass B reinforced *m* times, and B has subclass A reinforced *n* times, then the link from C directly to A was removed and the remaining link reinforcements were increased by 2. We then removed from the merged ontology any classes or links that were not reinforced.

Finally, we applied an *equivalence heuristic* for collapsing classes that have common reinforced superclasses and subclasses. The equivalence heuristic found that all reinforced subclasses of Person are also reinforced subclasses of Humans, and all reinforced superclasses of Person are also reinforced superclasses of Humans. It thus deems that Humans and Person are the same concept. This heuristic is similar to an inexact graph matching technique such as (Manocha et al., 2001). Fig. 4 shows the collapsed consensus ontology, now containing 36 classes related by 62 subclass links.

## 4. DISCUSSION OF RESULTS

A consensus ontology is perhaps the most useful for information retrieval by humans, because it represents the way most people view the world and its information. For example, if most people wrongly believe that crocodiles are a kind of mammal, then most people would find it easier to locate information about crocodiles if it were located in a mammals grouping, rather than where it factually belonged.

The information retrieval measures of precision and recall are based on some degree of match between a request and a response. The length of a semantic bridge between two concepts can provide an alternative measure of conceptual distance and an improved notion for relevance of information. Previous measures relied on the number of properties shared by two concepts within the same ontology, or the number of links separating two concepts within the same ontology (Delugach 1993). These measures not only require a common ontology, but also do not take into account the density or paucity of information about a concept. Our measure does not require a common ontology and is sensitive to the information available.

Although promising, our experiments and analysis so far are very preliminary. We used the following simplifications:

- We did not use synonym information, such as is available from WordNet, and so did not for example merge “meat eating” and “carnivorous.”
- We did not make use of class properties, as in subsumption.
- Our string-matching algorithm did not use morphological analysis to separate the root word from its prefixes and suffixes, and did not identify negated concepts, such as “uneducated” versus “educated.”
- We used only subclass-superclass information, and have not yet made use of other important relationships, notably *part-of*.

Our hypothesis, that a multiplicity of ontology fragments can be related automatically without the use of a global ontology, appears correct, but our investigation is continuing according to the following plan:

- Improve the algorithm for relating ontologies, based on methods for partial and inexact matching, making extensive use of common ontological primitives, such as *subclass* and *partOf*. The algorithm will take as input ontology fragments and produce mappings among the concepts represented in the fragments. It will use constraints among known ontological primitives to control computational complexity.
- Develop metrics for successful relations among ontologies, based on the number of concepts correctly related, as well as the number incorrectly matched. The *quality* of a match will be based on semantic distance, as measured by the number of intervening semantic bridges.

## 5. CONCLUSION

Imagine that in response to a request for information about a particular topic, a user receives pointers to more than 1000 documents, which might or might not be relevant. The technology developed by our research would yield an organization of the received information, with the semantics of each document reconciled. This is a key enabling technology for knowledge-management systems.

Our premise is that it is easier to develop small ontologies, whether or not a global one is available, and that these can be automatically and *ex post facto* related. We are determining the efficacy of local annotation for Web sources, as well as the ability to perform reconciliation qualified by measures of semantic distance. The results of our effort will be (1) software components for semantic reconciliation, and (2) a scientific understanding of automated semantic reconciliation among disparate information sources.

## REFERENCES

Berners-Lee, T. Hendler, J. and Lassila, O. 2001, “The Semantic Web,” *Scientific American*, May 2001.

Delugach H. S. 1993 “An Exploration Into Semantic Distance,” *Lecture Notes in Artificial Intelligence*, No.754, Springer-Verlag, Berlin.

Heflin J. and Hendler J. 2000, “Dynamic Ontologies on the Web,” *Proc. 17th National Conference on AI (AAAI-2000)*, AAAI Press.

Mahalingam, K. and Huhns, M.N. 1997, “An Ontology Tool for Distributed Information Environments,” *IEEE Computer,* vol. 30(6)

Manocha, N., Cook, D. and Holder, L. 2001 “Structural Web Search Using a Graph-Based Discovery System,” *ACM Intelligence*, vol. 12(1)

Pierre, J. M. 2000 “Practical Issues for Automated Categorization of Web Sites,” *Electronic Proc. ECDL 2000 Workshop on the Semantic Web*, Lisbon, Portugal. http://www.ics.forth.gr/proj/isst/SemWeb/program.html

Stephens L. M. and Chen, Y. F. 1996, “Principles for Organizing Semantic Relations in Large Knowledge Bases,” *IEEE Transactions on Knowledge and Data Engineering*, 8(3)

Wiederhold, G. 1994, “An Algebra for Ontology Composition,” *Proc. Monterey Workshop on Formal Methods*, U.S. Naval Postgraduate School