Research Directions to Support Next-Generation Web Visions

Michael N. Huhns
Center for Information Technology
University of South Carolina
Columbia, SC 29208
huhns@sc.edu

March 2002

1. Introduction

There are many visions for the future of the Web, but most are based on the following tenets:

- The Web will be ubiquitous
- The Web will provide not only content, but also services
- The Web will be continue to be dynamic, and will grow substantially in the short-term
- The Web will continue to have no central authority, and its components will remain autonomous
- The Web will support cooperative peer-to-peer interactions, as well as client-server interactions

Most of the visions also recognize the following problems with the Web, as it currently exists:

- Information on the Web is not organized
- Information on the Web might be inconsistent, incomprehensible, and inaccurate
- Web searches typically earn low values for precision and recall
- Web services are rigid, procedural, and strictly client-server
- State information is purely local.

If there were a central authority with a global ontology to which all Web components adhered, and *if* the components of the Web were static, and *if* the identity of the components were fixed, and *if* there were a small fixed number of component types, *then* the above problems would disappear, but the Web would no longer be the vibrant useful place on which the global economy and modern society are based. Research is needed to overcome the problems and realize the visions in a way that will foster the growth of the Web and won't compromise its utility. I believe that *the keys to the next-generation semantic Web are cooperative services, systemic trust, and semantic understanding, coupled with a declarative agent-based infrastructure*.

The size and dynamism of the Web presents problems, but it fortuitously provides a means for solving its own problems. For example, for a given topic there might be an overload of information, with much of it redundant and some of it inaccurate, but a system can use voting techniques to reduce the information to that which is consistent and agreed upon. For another example, there might be many potential service providers competing for many potential clients, and some of the providers might not be trustworthy, but a system can use a Web-based reputation network to assess credibility. Finally, there might be many different ontologies used by different sites, but a multiplicity of ontologies can be shown to yield a global, dynamically formed, consensus ontology. I will address each of these concepts in turn.

2. Trusted Autonomy

Research should be conducted to investigate how information credibility can be assessed and maintained in a large-scale environment of autonomous information sources. This is a key component of an overall research objective of increasing the *security* of information, while ensuring its availability. We have begun to investigate the role of reputations and how they can be propagated and computed efficiently. The essence of our approach is the use of large-scale systems of computational agents that interact to maintain up-to-date assessments of the reliability of information sources in specific domains.

Information systems are becoming increasingly autonomous. Automated planners are scheduling military combat missions. The Air Force routinely uses fire-and-forget armaments. NASA missions might be away from human contact and control for several years. Cruise missiles are programmed with only waypoints and destinations. Global supply chains will control the complex movement of goods from raw materials to customers without human intervention. Unmanned aerial vehicles (UAV) are being supplied with weapons to defend themselves and attack targets. Soon, soldiers will be sharing the battlefield with robotic tanks and artillery. As systems and missions become more complicated and of longer duration, the software systems controlling them will of necessity have to be large and complex. The situations to which the systems will be subjected cannot all be anticipated, so the systems cannot be fully tested. We will basically have to trust them, and there must be a principled basis for our trust.

What are some of the ingredients of trust?

1. One basis for trust is understanding, i.e., you are more likely to trust something if you understand it. Unfortunately, as systems become more complex it is harder to understand them. So, we need to describe and program and use them at higher levels of abstraction where the complexity is hidden. At this level, systems become team members and behave like agents or, where there are a lot of them, behave like societies.
2. Another basis for trust arises from philosophy and societal conventions. I suggest that agents with an explicit ethics and a philosophy that we understand and endorse can lead to systems that we trust. For example, Western philosophy teaches that there is value in each human life; so if a robotic tank embodied such a philosophy, then we could be confident it would not roll over its own troops in its zeal to carry out an assignment.

I believe that trust can be established through components based on agents, an architecture that embeds explicit philosophical principles into the agents, and organizations of agents that are akin to human social systems. Human organizations and societies have evolved ways of maintaining social order by adopting ethics, norms and conventions. Complex agent-based systems can be constrained by sets of agent societal laws similar to Asimov's Laws.

The research that has been conducted on DARPA programs such as CoABS, TASK, and ANTS so far has demonstrated that

- Agents can glue together independently developed legacy systems.
- Control of a system can be distributed among autonomous agents, but still be globally coherent.
- Coherence and capability improve greatly when systems (represented by agents) cooperate.

However, the following crucial technical issues remain unresolved, and their resolution constitutes the major thrust of a proposed future effort:

- Whether adopting human social concepts can enable agents to achieve the same flexibility and robustness exhibited by some human societies
- Whether having an explicit philosophy can enable agents and the complex systems they form to handle emergencies and unanticipated problems or circumstances.
- Whether philosophical agents are more effective members of robotic-agent-person teams.
- Whether we can raise the abstraction level at which we program and use complex systems, thus simplifying their deployment.

The results of such a program will be an evaluation of the feasibility of systems controlled by a philosophical agent society, and a roadmap for development, implementation, and deployment in future systems.

The benefits of this program to future information systems are: (1) it will support applications of longer duration and greater complexity than are possible under the current model of human control, (2) it will reduce costs by minimizing the amount of support required for applications, (3) it will essentially eliminate communication time lag as a significant factor, providing the ability to react to and take advantage of serendipitous events, and (4) it will significantly enhance application robustness.

3. Information Credibility

We are rapidly approaching a world in which computing is ubiquitous. Sensors, wearable computing devices, and portable computing and communications devices will make it possible to generate rapidly an enormous database of individual data items pertaining to both the human and equipment resources of any large organization. It will be important in both strategic and tactical situations to have a means for assessing the reputation of information in the system so that decision-making processes are enhanced.

Our suggested approach draws on solutions to several problems that resemble the credibility problem. First, we draw for a theoretical underpinning on the extensive research on the rumor problem and on Bayesian network solutions to that problem. Next, we include some of the methods used in Web search engines for estimating the quality of the information on a particular Web site. Finally, since estimation of credibility relies on interconnectivity and interrelatedness of information, and this information is subject to corruption or repudiation, we introduce some of the methods that are successfully being used to compare phylogenetic trees in the presence of mutation and change of an underlying genetic structure.

At its core, the problem of credibility reduces to a problem of quantifying parameters on the nodes of a graph. Each node on the graph represents an item of information, and directed arcs link one item to any other item whose credibility is dependent upon the first. The credibility itself is a numerical quantity or quantities attached to the node. One basic technique for determining the reputation of data has been adopted by some web search engines, most notably Google. This technique involves consideration of the number of Web links pointing *into* the site in question. Under the presumption, perhaps, that forty million Web sites are not likely to be wrong, the more links into a site, the more reputable the data. Thus, when Google ranks the sites returned from a search, a higher ranking is given to a site to which a large number of other sites point.

A similar approach can be taken with regard to reputation, with some changes. Data items that are mutually verified by *independent* sources are more likely to be reputable than are items that can be tracked back to a *single* source, regardless of how many times the item has been retransmitted by someone downstream from the original source. On the other hand, some individual sources can be more credible than others, and items generated from within a secured environment are more likely to be trustworthy than are items generated in the field from sensors or communications devices that could have been physically compromised since their initial placement.

Viewed in this way, the problem of determining the reputation of information is strongly linked to the problem of building “come-from” trees and the problem of inferring credibility of a given node from the weighted credibility of nodes from which the node in question derives information. This is a classic problem in computing, but often a difficult one to solve due to the need for substantial computing resources and an everything-against-everything computation to be performed. Even after a computation is performed to produce a starting point, a problem persists in how to add information to the credibility trees incrementally rather than being forced to do the computation over and over again.

The problem as stated also bears resemblance to the problem of phylogenetic comparisons in computational biology. Of great interest to computational biology is the determination of a path by which one genetic sequence can be transformed into another by mutation and natural selection. These transformations can be represented as trees, and one computational problem in phylogenetics is to quantify the similarity of one tree against another. Weightings exist on the arcs of the trees, since some changes are more likely than others based on the underlying biology. And, just as with the problem of reputation of information, the result of performing computation on the problem as modeled may produce wrong or irrelevant answers, since irrelevant similarities among phylogenetic trees can exist.

Finally, a theoretical and information-theoretical analysis of the problem is required. The problem of credibility is closely related to the rumor problem. Items that can be traced back to a single source are no more reliable than that source, no matter how many times the original information has been propagated further. The rumor problem has been solved by the use of Bayesian networks that not only manage the uncertainty, but also are sufficiently robust that a single garble or corruption will not produce a catastrophic ripple effect.

From an information-theoretic standpoint, trustworthiness and robustness in an information system are due to redundancy (cf. parity bits in a data word). Ubiquitous information sources, if organized appropriately, can provide the redundancy needed to detect, correct, and compensate for errors, whether the errors arise from inexact sensors, algorithmic mistakes, or disinformation. A further complexity exists in the problem in that “trust” and “mistrust” of information are not necessarily symmetric; in some instances a minimax principle obtains, but in others the possibility of any success must outweigh the possible costs. We expect our investigations to yield fundamental advances in understanding the limits of large-scale distributed reliability assessments.

The result will be an improved understanding of the credibility of information and the reputation of information sources. This will have the practical benefits of improving information utility for Internet applications, where typical searches yield huge numbers of documents with unknown credibilities, and for information-intensive military scenarios, where uncertainties and disinformation might be widespread.

4. Consensus Ontologies for Semantic Reconciliation

Organizational knowledge typically comes from many independent sources, each with its own semantics. Corporate information searches can involve data and documents both internal and external to the organization. We are investigating a methodology by which information from large numbers of such sources can be associated, organized, and merged. The central 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.

4.1 Analysis

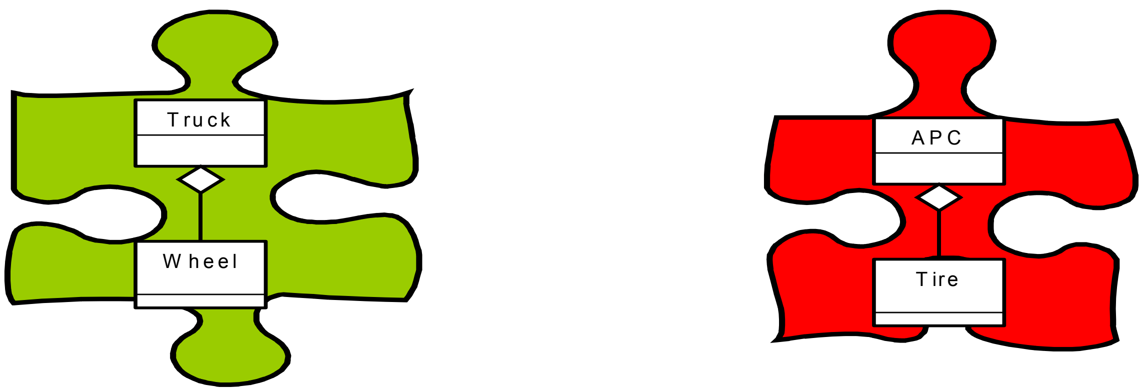
Our research targets the following basic problem: a search will typically uncover a large number of independently developed information sources—some relevant and some irrelevant; the sources might be ranked, but they are otherwise unorganized, and there are too many for a user to investigate manually. 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.

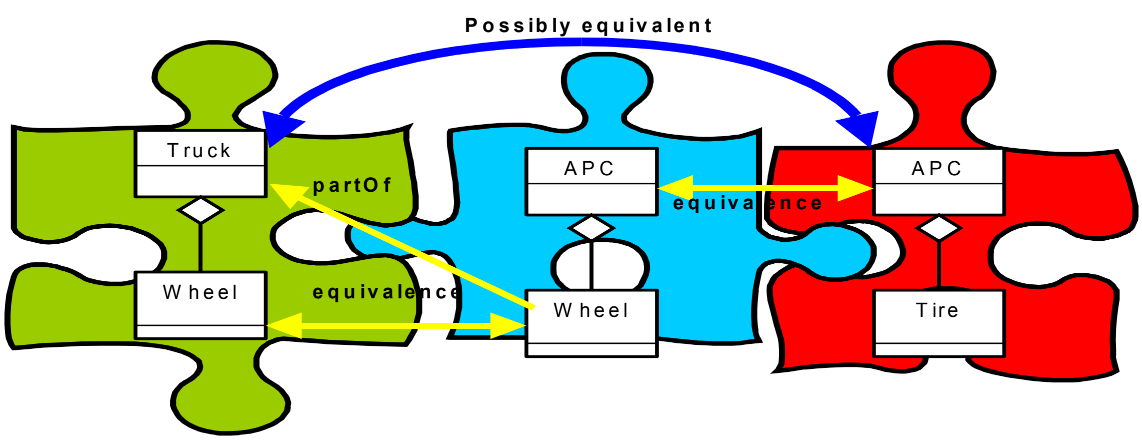
Other researchers have attempted to merge a pair of ontologies in isolation, or merge a domain-specific ontology into a global, more general ontology [Wiederhold 1994]. To our knowledge, no one has previously tried to reconcile a large number of domain-specific ontologies. We have evaluated our methodology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

4.2 Reconciling Separately Developed 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.



(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

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 Figure 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.

In Figure 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*.

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.

4.3 Discussion of Preliminary 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.

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. Constraints among known ontological primitives will 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.

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 sources.

5. Declarative Cooperative Services

As Web uses (and thus Web interactions) become more complex, it will be increasingly difficult for one server to provide a total solution and increasingly difficult for one client to integrate solutions from many servers. Web services currently involve a single client accessing a single server, but soon applications will demand federated servers with multiple clients sharing results. Cooperative peer-to-peer solutions will have to be managed, and it appears that an agent basis is what is needed. Agents can balance cooperation with self-interest. Agents also have a property of persistence, which is necessary in order to establish trust.

Moreover, agents typically interact via the exchange of declarative messages. The current models for Web services are based on the exchange of procedures via CORBA, .Net, or RMI, but the history of computing has shown that declarative approaches are ultimately favored. Just as SQL supplanted IMS procedures, so will agent communication languages, such as the FIPA ACL, supplant WSDL.

Significant research is needed to enable autonomous services to be federated on demand and with acceptable delay. Coupled with the above research advances in semantic reconciliation and the assessments of trust and credibility, the result will be a more efficient and more useful Web in an environment of ubiquitous computing.

Bibliography

[Berners-Lee et al. 2001] Tim Berners-Lee, James Hendler, and Ora Lassila, "The Semantic Web," *Scientific American*, May 2001.

[Delugach 1993] Harry S. Delugach, "An Exploration Into Semantic Distance," *Lecture Notes in Artificial Intelligence*, No.754, pp. 119-124, Springer-Verlag, Berlin, 1993.

[Heflin and Hendler 2000] Jeff Heflin and James Hendler, "Dynamic Ontologies on the Web," *Proc. 17th National Conference on Artificial Intelligence (AAAI-2000)*, AAAI Press, 2000.

[Huhns and Stephens 1989] M. N. Huhns and L. M. Stephens, "Plausible Inferencing Using Extended Composition," *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence, IJCAI-89,* Detroit, MI, August 1989, pp. 1420-1425.

[Mahalingam and Huhns 1997] K. Mahalingam and M. N. Huhns, "An Ontology Tool for Distributed Information Environments," *IEEE Computer,* vol. 30, no. 6, pp.80-83, June 1997.

[Manocha et al., 2001] Nitish Manocha, Diane J. Cook, and Lawrence B. Holder, "Structural Web Search Using a Graph-Based Discovery System," *ACM Intelligence*, vol. 12, no. 1, Spring 2001, pp. 20-29.

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

[Stephens and Chen 1996] Larry M. Stephens and Yufeng F. Chen. "Principles for Organizing Semantic Relations in Large Knowledge Bases," *IEEE Transactions on Knowledge and Data Engineering*, 8(3), June 1996, pp. 492-496.

[Wiederhold 1994] Gio Wiederhold, "An Algebra for Ontology Composition," *Proc. 1994 Monterey Workshop on Formal Methods*, U.S. Naval Postgraduate School, 1994, pp. 56-62.