White Paper (March 26, 2002)

The Value of Ontology-Based, Service-Oriented, Distributed Systems in a High Bandwidth Environment

Steven J. Gollery
Collaborative Agent Design Research Center
Cal Poly, San Luis Obispo, California

Introduction

Over the last decade, publishing information on the World Wide Web has become an accepted and even necessary way for organizations, whether commercial or governmental, to do business. Until recently, however, information on the web was intended almost entirely for communication with human beings. The small amount of information placed on the web for the use of distributed software systems fell primarily into two categories: proprietary formats for private partners, or meta-data to enhance the efforts of search engine robots.

Very recently, that situation has begun to change. Organizations are beginning to see the value of allowing programmatic access to selected information through the use of Web Services. Each web service provides operations that can invoked by other software systems across the Internet, enabling information to be shared directly between one program and another, regardless of physical location. This is made possible by two developments: one is the specification of public standards for web service deployment and use, and the other is the availability of high-speed Internet connections.

Web Services have backing from the software industry as represented by corporations including Microsoft, IBM, Sun Microsystems, and many more. There is also substantial interest from other industries, which expect major benefits, particularly in the area of supply chain management. It is very likely that Web Services will be a large part of the Internet landscape starting later this year and continuing into the future.

But Web Services as they currently stand are not enough for true interoperability and flexibility. Although more loosely coupled than any other attempt to build large-scale distributed systems, web service architectures nonetheless contain points of rigidity. Client programs cannot be implemented in such a way that they can access any web service on an ad hoc and opportunistic fashion, based on fast-changing requirements. Instead, each client program currently must be designed to assume certain characteristics about the web services it will use. Web services that do not possess these characteristics will be unavailable to that client, regardless of the potential value of the information that web service might provide.

We propose that, in order to build truly flexible Internet-based distributed systems that can respond intelligently and appropriately to unexpected problems and novel situations, service providers need to make more available than just their information: they also need to take advantage of available bandwidth to allow service consumers to have access to their information model: in other words, to the ontology of the services and the domain of

knowledge. Automating information interchange requires high capacity network communications.

The remainder of this paper discusses the use of dynamic ontologies as an enabler for flexible, high-speed information interchange.

Ontology-Based Web Service Systems

An ontology-based web service includes the ability for a remote system to learn about the information model used by the service provider to represent its services and its domain of knowledge: this is the service provider's *ontology*. Once the ontology has been made available through a web service, service consumers (client programs) can access the ontology, thereby learning about the concepts and relationships that the web service provider uses internally.

When a service consumer program has the entire model of the service provider's ontology, the consumer can reason about the value of the services offered by this provider. This is a far more powerful concept than the kind of searches allowed by today's web service specifications, which are essentially keyword based. This enhanced service discovery capability allows the consumer program to manage rapidly-changing decision-making environments without missing vital information.

An ontology-based service also provides access to the information model of its domain of knowledge. A service consumer program can therefore infer relationships between the knowledge representation that the service uses, and the representation that the consumer uses. This ability to relate the two representations provides a solution for another problem afflicting web services as currently defined: the potential mismatch between objects on the service provider and objects on the service consumer. Web service operations are defined to receive and return objects that are instances of very specific classes. In order to allow truly dynamic discovery of and access to arbitrary services, service consumer software must have the ability to translate between its own object representation and the definition that the service requires. But without an understanding of the context within which the service classes are defined, the consumer is unable to perform that translation.

By making the service provider's ontology available to remote service consumer programs, we improve the ability of the entire system to respond and adjust to unexpected and unplanned changes. When a sufficient number of organizations have exposed their information and their information models as web services, we see the beginnings of an information community.

Basing interoperability on dynamic information models instead of static implicit definitions not only allows automated reasoning, it also removes the necessity of reaching agreement on representation. The difficulty of defining a standard representation that is acceptable to a sizeable group of organizations has frequently stymied or fatally slowed past efforts to achieve interoperability. When each organization can define its own ontology, the organization remains in control of its own representation, even while making its information available to other systems.

The concept of including software-processable semantic information in a web site has been called the Semantic Web. The Semantic Web has been discussed in many books and articles; one of these is "The Semantic Web" by Tim Berners-Lee, James Hendler, and Ora Lassila in the May, 2001 issue of Scientific American Magazine. The combination of Web Service specifications with the concepts of the semantic web was described in the March/April 2001 issue of the IEEE Intelligent Systems journal in the article "Semantic Web Services" by Sheila A. McIlraith, Tran Cao Son, and Honglei Zeng, all of Stanford University.

The Information Community

As organizations deploy ontology-based web services, other organizations will follow suit. In some domains, such as emergency management, the ability to rapidly share information among organizations in a swiftly changing environment may make a measurable difference in the quality and timeliness of the response. Planners will have intelligent software assistants that use information from many sources to coordinate response to situations that are overwhelming today because information is not available when it is needed, in a form that allows software-based reasoning.

In the commercial world, a manufacturer can derive benefit from enabling potential customers to reason about products automatically, potentially creating sales that occur at electronic speeds, instead of human decision-making speeds. The value of the system increases tremendously if the customer's accounting department, shipping companies, financial institutions, and all the other components of the supply chain are also available programmatically.

In the academic world, research results, journal articles, and all forms of other documents will be available in representations that allow ontology-aware software to learn about the concepts within the documents, and to infer relationships among documents. When research is presented on the web in a form that can be reasoned upon by automatic inference engines, we expect that new forms of collaboration will become available to a wide community of scientists.

Possibilities expand further when we begin to consider cross-domain access. Service consumer software will range across such diverse domains as news stories, historical documents, corporate reports, weather predictions, scientific research, and dozens of other kinds of information to make connections, create relationships and to learn in ways that are impossible today. The Scientific American article mentioned earlier points out the potential to develop query engines that bring together information from multiple sources to produce their results. This and many other processes that must be performed by a person in several steps today will be the responsibility of automated intelligent agents in the near future.

In a community of systems sharing their ontologies and their information, there will be a natural place for the intelligent agent. In many cases, these agents will be representatives of human beings: instead of spending time simply locating information, people will describe the information that they want, or what task they need the agent to perform, and the agent will perform the actions that the user must do personally now.

Conclusion: The Role of High Bandwidth

In an interview in the March 2002 issue of Technology Review (www.technologyreview.com), Larry Smarr, director of the California Institute for Telecommunications and Information Technology at the University of California, San Diego, points out that "the 21st Century is very much about ... global virtual teaming." Smarr also foresees a future in which "the percentage of traffic on the Internet that touches a human is going to be a very, very tiny fraction....we are going to have a world of intelligent agents that...work for us during the day or night." Both of these visions of the future only become possible when high speed networks are combined with the ability to exchange semantically-rich information models.

Once an information community exists, the amount of traffic across the network can be expected to increase enormously. In fact, as each new ontology-based web service comes online, there will be more traffic as other services, service consumers, and intelligent agents learn about and use the new service's capabilities. Over time, we can expect that the traffic on a network like this will increase exponentially as new services are created.

Much of the value of information access comes from the ability to find information and to incorporate it into the decision-making process quickly. Information needs may be transient and rapidly changing. High bandwidth access to the network is the key to success for ontology-based information-sharing distributed systems.