International Semantic Web Workshop, Stanford University, July 30–31, 2001
Carolyn Sleeth, Research Analyst
SRI Consulting Business Intelligence
csleeth@bic.sri.com

The Semantic Web, Trust, and the Business Environment

Recent discussion in the business community has focused on the network economy and the change brought by the Internet-enabled enterprise. No longer can a company focus solely on itself—an independent unit—as its primary concern. Rather, it must focus on the partnerships and relationships it maintains with other companies. These relationships and their communication processes have become the keys to successful business.

In theory, the Internet makes these partnerships and relationships easier to build and maintain. The W3C's vision of the Semantic Web charts one of the many possible paths with which business uses can extend theory to reality. The Semantic Web will augment current Web content, which is primarily available in human-readable format only, to include machine-readable content. Machines can then use these data for automation and integration in a variety of applications, bypassing time-consuming human intervention. Technologies and protocols such as intelligent software agents, XML (Extensible Markup Language), RDF (resource description framework) schema, DAML (DARPA Agent Markup Language), and OIL (Ontology Inference Layer) enable this machine-readable Web. Making information on the Web machine readable ultimately means making the Web a more meaningful place for human users.

A major limitation inherent in a straightforward implementation of these technologies, however, is that they provide only for the communication of information. How do machines determine whether the source of the information is trustworthy? Within a single enterprise, sources of information are strictly under the control of the company's information technology department. For example, on a traditional factory floor, a control system knows exactly which fieldbus it is communicating with. It operates with the knowledge that the control system's information is qualified and trustworthy. Even in a closed, interbusiness supply chain, partners have considerable control in making certain that sources are reliable. But the Semantic Web will enable an expansion beyond closed communities, and devices will be able to communicate and perform transactions with other devices that they have never met. The genie that opens once-closed corporate networks to external communications also opens a Pandora's box of undifferentiated, unqualified sources. One of the major problems that most users, both business and consumer, face today is the proliferation of low-quality data and misinformation.

How can devices know that particular information is high quality—that it reliably and correctly serves the task at hand? For an intelligent device to operate usefully, its data inputs must be certified as accurate and of high quality. For many applications involving transactions between parties with no foreknowledge of each other, a trusted third party must assure that the correlation of the information to the real world is reliable. Only then does a device or system have the knowledge to proceed, reliably, in its task.

What can provide this assurance? At the XML 2000 conference, Tim Berners-Lee pointed out that the incorporation of digital signatures in the information architecture of the Web could resolve the "how" of this problem. Digital signatures, in combination with

RDF schemas and ontologies can provide the means not only to authenticate information but also to automate much of the filtering process, automatically eliminating vast quantities of data from consideration, according to the quality expectations or needs of the user. Such a system has the potential to bring an element of trust to the Web, with implications for both businesses and consumers.

The integration of digital signatures into the Web information architecture would bring a much-needed granularity to the digital-signature and public-key-infrastructure (PKI) cryptography systems on the Web. The current system operates in a binary manner, allowing only trusted or not-trusted designations for specific sources or transactions. The Semantic Web will allow users to designate trust for a certain source in only certain domains or in only specific areas of a particular ontology, or in certain types of transactions. Various combinations of trusted sources from a diverse set of information providers in different domains of expertise can result and will enrich the user's Web information and electronic-commerce environment.

Within limited, closed communities, such as intercompany supply chains, users themselves are able to define who they trust as information providers in particular domains of expertise. As communication expands beyond closed communities, however, users looking to corroborate their information will need trusted third parties to help determine whether the information is reliable or not.

Trusted third parties are already emerging in a number of industries. Businesses that have already earned trust in a corporate setting see an easy entry into the role of a trusted third party for online interactions. This role could easily extend into an authority that devices in different industries automatically and electronically turn to in order to authenticate the unfamiliar source. Organizations that are currently investigating opportunities as trusted third parties include banks, online auctioneers, dispute-resolution services, health-information banks, security-system vendors, credit-card companies, postal services, and government agencies. PKI vendors, such as Verisign, also pitch themselves as trusted authorities that can confirm identities of individuals or corporations.

Within their own industries, these enterprises have positioned themselves as authorities in certain domains. The digital-signature capabilities of the Semantic Web will allow them to communicate their authentication of information—does it have a reliable correlation to the real world—to the devices receiving it. This authentication of information—the award of a digital certificate—gives the information the backing of the trusted third party. If a device accessing the Web finds information certified by an appropriate digital certificate (as specified by the device's user), the device can proceed in interactions with a level of authority. The input from the trusted third party enables machines to interact with each other in a vastly more useful, reliable, and trusted way.

Berners-Lee's vision of incorporating digital signatures at several architectural levels of the Semantic Web will be an essential element for the continued success of the Web as unqualified, undifferentiated data proliferate more rapidly than qualified data on the Web. Digital-signature capabilities will be particularly important in the business environment as corporations attempt to break out of their private networks to build ad hoc networks and alliances and to expand their reach to unfamiliar partners and customers.