ORACLE
TECHNOLOGY NETWORK

Feature

As Published In

ORACLE MAGAZINE
May/June 2006

# Semantic Breakthrough

By David Baum

The SemanticWeb is breaking out of research labs and into the back office, as businesses of all types prove the potential of this far-reaching technology.

By any measure, the World Wide Web is an immense success, helping everybody from preschoolers to CEOs share information, simplify research, and conduct business online. Still, many researchers aren't happy with the state of Web technology. They think that something is missing: intelligence.

For example, to assemble information from several sites with today's Web, you have to visit them individually and then cut and paste the content to create a cohesive presentation. This pastime has become routine for many; yet if you ask a computer to do the same thing, it wouldn't know where to start. That's because the Web's HTML pages are designed for humans, not machines, to read.

This weakness is more obvious during complex Web transactions such as online shopping. Purchasers know that the posted item number and price refer to the product pictured nearby—Deep Blue Goggles and Snorkel Set, Item Number A-224, US$12.99. A machine can't make this same inference, because HTML wasn't designed to contain this level of meaning. HTML can say, "Make this box square, and position it next to this text string." But it has no semantic mechanism to express that these particular pieces of information are bound together—that this price refers to those items—let alone to suggest that a customer might also be interested in a pair of flip-flops.

The Semantic Web puts HTML data into a machine-readable format, so that computers can aggregate it and understand these relationships. It accomplishes this task with Extensible Markup Language (XML) and data-language standards such as Resource Description Framework (RDF) and Web Ontology Language (OWL), two World Wide Web Consortium (W3C) standards. These standards and descriptors enable Web developers to add layers of meaning to Web documents, supplying a framework for defining how data is linked and how its intended relationships are expressed.

"The Semantic Web is a great way to map unrelated sets of data together based on common criteria, and to mine data in new ways by being able to look across many types of information stores," says IDC analyst Scott Lundstrom. "In the past, it was hard for humans to make these

Ads by Google

**Oracle 10g**
Improve Performance, Lower Costs Oracle 10g Tailored for Business
www.teamsolve.co.uk/10g

**Oracle EHR**
Electronic Health Record: pharmacy; manage inventory; take credit cards
thot.com

**Oracle developer**
Find the Best-Paying $100K Jobs Join the Club & Get Hired Faster
www.TechnologyLadder.com

**Foothill College**
Earn Your Computer Science Degree & Compete For The Valley's Top Jobs!
www.Foothill.edu/CTIS

**Senior JD Edwards/E1 Jobs**
Direct to employer, Confidential Specialize in JDE/Enterprise One
www.jdejobinfo.com

associations. Now we are teaching computers to do it for us."

## Design Flaws

When the Web took off as a commercial venture, HTML became fixed in the collective development toolbox. Today's resulting hyperlinks include little information about the relationships among sites or elements on a page. This limitation restricts the Web from reaching its potential as a medium in which data can be shared and processed by automated tools, as well as by people. If Web documents were defined semantically, simple searches and queries could more easily identify related content.

For example, in the life sciences domain, scientific information related to drug discovery and development could be semantically linked to simplify data handling operations, accelerating the process of identifying safe and effective pharmaceutical compounds. IDC's Lundstrom says Semantic Web technology is a natural fit for the knowledge management challenges facing researchers and clinicians in this industry. As scientists continue to unravel the human genome, the ability to locate, navigate, analyze, and integrate vast amounts of scientific information will be essential. Yet current database management systems can't easily capture all the relationship nuances of biological and chemical objects, such as genes, proteins, and DNA sequences. Once these elements are identified and linked in a Semantic Web graph, researchers can view, analyze, and take action on distinct bits of information in a unified way. "There is a degree of flexibility in the Semantic Web that allows researchers to turn very disparate bits of data into knowledge," Lundstrom says.

Like adding metadata to a database, the Semantic Web includes information about relationships between data elements. These elements are defined in self-descriptive statements called RDF triples containing a subject, predicate, and object. By modeling data in this way, developers can build implicit relationships among data elements that were formerly treated as separate entities—or that previously had to be modeled and analyzed using dissimilar database technologies and tools. Lundstrom says that building new relationships in this way can improve business intelligence. One example: "Pharmaceutical companies amass huge volumes of experimental data that may be spread across many R&D units—from target biology and compound discovery to preclinical development and clinical trials," he says. "Semantic Web technology is a great way to connect and interpret it."

## Google, the Next Generation

Intelligent internet searches are probably the best example of Semantic Web technology in action. With typical Web searches, the exact search string must be embedded in all the documents that result. For example, a search on "yacht racing" that doesn't include the phrase "America's Cup" won't turn up articles about this world-renowned sailing event, unless those articles happen to use the phrase "yacht racing." But if Web servers understood the context of yacht racing and its common data elements, they would know that terms like America's Cup and yacht racing are interrelated, even if they're not explicitly mentioned in the search criteria. On the Semantic Web, these terms can be described in a yacht-racing ontology—a customized lexicon—designed to deliver relevant information for people doing casual searches. When coded with a language such as RDF, these ontologies become understandable to electronic agents searching for information as well.

"Ontologies created with RDF enable you to model data in such a way that machines can make sense out of it," says Mary Parmelee, an ontologist team lead at McDonald Bradley, a privately held company delivering IT solutions to defense, intelligence, homeland security, federal law enforcement, and federal civil clients. "Ontologies express the relationships between data objects and describe them by their properties, so an inference engine can interpret these relationships automatically."

An inference engine is a smart search engine that interprets the rules that describe objects in an ontology. The ontologies specify how data relates to other data, acting as a filter to enhance Web searches based on a metadata layer that has been defined by subject-matter experts. RDF supplies a framework for labeling these ontologies and embedding labels in XML documents, so an inference engine can understand them.

McDonald Bradley is using this type of inference engine as part of a U.S. Department of Defense initiative to enable decision-makers and end users to search and retrieve data from disparate data sources. For example, if military personnel found an unidentified container in an abandoned vehicle, they could perform an intelligent search that linked the chemicals in that container, the manufacturer of that container, the type of vehicle, and the location. "Intelligence experts create these ontologies because they understand how certain materials link to other materials and how manufacturers might be involved," says Parmelee.

Snapshots

Cerebra
www.cerebra.com
Location: Carlsbad, California
Industry: Software and IT services
Employees: 35
Oracle products: Oracle Database 10*g*

McDonald Bradley
www.mcdonaldbradley.com
Location: Herndon, Virginia
Industry: Defense
Employees: 320
Oracle products: Oracle Database 10*g*

Moving into the Mainstream

Semantic Web capabilities fill a unique niche in information-intensive industries such as defense and pharmaceuticals. IDC's Lundstrom says that these technologies play a practical role in many other industries as well. "The Semantic Web is prevalent in life sciences because of the economics of clinical development: Anything that accelerates the rate at which a compound is approved is extremely valuable," he concedes. "But these same concepts apply to any situation in which a customized lexicon or ontology is useful."

Robert Shimp, vice president of global technology sales support and marketing at Oracle, identifies four typical ways a business could use Semantic Web technology: search, Web services, grid computing, and content management/compliance.

Just as semantically enabled searches apply to the internet at large, these same dynamics can be applied to a particular business domain within a large company. For example, an automobile manufacturer might have many autonomous divisions, each with different part numbers for the same basic items. Division A's part number XYZ-1234 might refer to an engine block that is identical to part number ABC-9876

Leading the Way with Oracle Database 10*g*
According to Robert Shimp, Oracle vice president of global technology sales support and marketing, Oracle Database 10*g* Release 2 is the world's first mainstream commercial database to provide direct and native support for semantic technology in the form of RDF triples. In an Oracle database, RDF triples are

at Division B. Purchasing officers want to obtain volume discounts by ordering these parts in bulk, but it's difficult to sort out these differences, because parts data is stored in different locations and used by multiple systems. A simple ontology could compare the information in each division's parts database to reveal that part number XYZ-1234 is identical to part number ABC-9876, and so on.

Companies also use semantic technology to describe Web services, making it easier for developers to find the functions they need when they are incorporating them into business processes. "A developer using Oracle SOA [service-oriented architecture] Suite to create an internet shopping application might need to locate a Web service that performs a credit check. Semantic technology could make it easier for that developer to find a Web service with a specific application-programming interface or protocol—helping to make SOA practical on a much larger scale," Shimp says.

Meanwhile, system administrators managing computer grids want to be able to deploy applications without worrying about which individual servers or storage devices are involved. As many Oracle customers have demonstrated, Oracle Real Application Clusters technology enables companies to build infrastructure grids using clusters of low-cost servers, modular storage arrays, and inexpensive disk drives, lowering costs and simplifying configuration challenges. Semantic technologies can describe grid resources so that each device in the network can "understand" what's available, negotiate for resources, and execute application logic. "Oracle is using semantic technology to enable machines to understand other machines, so that they can more easily cooperate in grid environments," says Shimp.

persisted, indexed, and queried similar to other object-relational datatypes. This serves as a foundation for a new generation of enterprise-class, semantically enabled business applications.

Oracle Database 10*g* Release 2 also includes a proprietary rule language and an integrated query language to integrate RDF triples with relational tables. (As yet, there isn't a standard for Semantic Web rules. Oracle is participating in a W3C working group to develop rule standards.) These capabilities are particularly useful when it comes to modeling and using complex data sets, because many types of data can be combined and interrelated in a single instance of Oracle Database 10*g*.

"If you're using Oracle Database 10*g* Release 2 and the Oracle Spatial option in your infrastructure today, you pretty much have everything you need to get started," says Shimp. "You don't need to build an entirely new infrastructure."

"Terminology such as RDF and OWL make semantic technologies sound sort of complex, but the practical usage of these technologies is quite simple," concludes Shimp. "Anybody with a basic understanding of databases will find that they're comfortable with semantic development techniques."

Semantic technology is important to content management and compliance because it makes it easier to meet Sarbanes-Oxley requirements that call for segregation of duties between financial responsibilities and function points. "In an industry such as financial services, there are multiple databases and multiple applications that have control points," says Jeff Pollock, vice president of technology at Cerebra, an Oracle partner that provides adaptive technologies for information integration, in Carlsbad, California. "For example, [banking personnel] will not be allowed to approve a loan application if they are also allowed to change the loan rate table. Semantic Web technology allows us to build rules that specify these roles, and then to keep track of which data is accessed by which individuals."

Oracle Database 10*g* Release 2 provides the core infrastructure for creating these semantically enabled databases, while partners such as Cerebra supply useful development tools. For example, one Cerebra customer is a Fortune 500 electronics manufacturer that was having trouble reconciling product definitions among its manufacturing, financial, and content management systems. Each quarter, the company struggled to pull together its financial results because of a disparity in how its products were classified. For example, financial analysts need to be able to track revenue according to categories like computer hard drives, but that's getting trickier as these components find their way into everything from PCs to DVD players. Cerebra created an ontology that helps managers understand these product families by unifying metadata, product data, market hierarchies, and Web site product classifications. The solution is anchored by Oracle Database 10*g*, which stores data in both relational and RDF formats, along with a rule language and query language to combine the two.

"Before, creating financial reports was largely a manual process, because somebody had to reconcile how various electronics products were related," says Pollock. "We transformed a six- to eight-week reporting process involving manual code and lots of [Microsoft] Excel spreadsheets into a one- to two-week process that is highly automated by the semantic technology inherent in Oracle Database 10*g*. This effort shortened the financial reporting lifecycle and made the entire ERP ecosystem more adaptable and flexible to the business community."

Down to Earth

The mathematical foundation of Semantic Web technology was laid decades ago when computer scientists invented graph-based data languages, but only recently have these data-modeling methods entered the practical domain of business. Shimp says the technology is now readily accessible to DBAs through data-language standards such as RDF. "There is a deep reservoir of knowledge in the academic community about the underlying mathematics of semantic technology, but from a practical, day-to-day, 'How do I use this stuff?' perspective, you don't need to know much of anything about it," he notes. "You need to be familiar with XML data tagging and basic relational concepts. But RDF triples can be represented effectively in relational data, and many of the concepts that Oracle DBAs know instinctively naturally apply."

Next Steps

VISIT the Semantic Technologies Center

LEARN what Oracle brings to Semantic Technologies

READ more about the Semantic Web
"Applying Semantic Web Technologies to Drug Safety Determination"
"Semantic Data Integration in the Life Sciences"

McDonald Bradley's Parmelee agrees, saying that defining terms in an ontology is not very different from what developers do in the database world. "The same logic that goes into an ER [entity relationship] diagram or a multidimensional database allows you to distinguish between objects and relationships in an ontology," she says. "We are conducting lab experiments associated with a large data integration project in which we use the Oracle Spatial RDF capabilities of Oracle Database 10*g* for storing data in RDF format. This format preserves the true graph-based representation of the ontology model, rather than trying to fit a graph-based structure into a standard relational mold."

IDC's Lundstrom believes the ultimate value of Semantic Web technology will be directly proportional to the rate at which the W3C standards are stabilized. "The Web is becoming more diverse even as information systems are interconnected in new ways, making it harder and harder

for humans to understand these relationships," he says. "Standards like RDF and OWL make our software smarter, so we can create systems that make logical inferences for us. Once the standards are widely accepted, the Semantic Web will take off."

When that happens, instead of hard-coding relationships between data elements, as developers must now do, semantic technologies will weave this information into the very fabric of the World Wide Web. The flexibility of RDF allows developers to specify implicit relationships among data elements, leading to more complete and intelligent queries.

Lundstrom believes that Oracle is at the forefront of the database industry in its support for RDF datatypes and other essential technologies of the Semantic Web. "Oracle is leading the marketplace by embedding Semantic Web capabilities into its database, enabling computers to aggregate data and make inferences about data relationships," he says. "Oracle is in a great position here since the companies that benefit the most from the Semantic Web are large, distributed, global operations, where Oracle is already the database vendor of choice."

---

David Baum (david@dbaumcomm.com) is a freelance business writer based in Santa Barbara, California.

Send us your comments