

Publications

Archives

Search

Free Newsletters**Online Features**

Daily News

Fear

Swift Kick

Answering Machine

Reload

Five Thoughts

Machine Shop

Numbers

Learning Curve

Book Room

Executive Guides

Technology

Career

Reference Tools

Lifestyle

CMO in the Know

Research

Research Reports

Polls & Surveys

Events Calendar**White Papers****Advertising Info****Contact Us**

Editorial Staff

Service Center[Print This](#)[Email Article](#)[Request Reprint](#)[Talk Back](#)[SAVE THIS](#)**WHAT'S NEXT****Semantic Applications, or Revenge of the Librarians**

The real future of technology rests on the concept of standardizing human language, not computer programs. And that puts the customer in total control.

BY DAVID MOSCHELLA

CONSIDER THAT two of the most significant Next Big Thing concepts on the Web today are Web services and semantic applications. Both are potentially major advances in the way IT is used. However, the transition from Web Services to truly semantic applications will mark a major handoff between the supplier- and customer-driven worlds. For the first time, the establishment of major new computer capabilities is becoming primarily a customer responsibility.

Web services is part of a long tradition of major supplier-led initiatives, each characterized by a great deal of promotion and hype, along with many important new acronyms and standards. In contrast, semantic applications will be predominantly a customer challenge, with vendors playing a supporting but more secondary role. This difference is telling since, despite all the hype, Web services should really be seen as just a stepping-stone to the larger goal of developing truly intelligent semantic applications

Despite the power of Web services, emerging standards such as [XML](#) and [Simple Object Access Protocol \(SOAP\)](#) are really only capable of facilitating some of the tasks at hand. They are basically just the conduits through which information can flow. To achieve true computer-to-computer interoperability, business terminology itself has to become much more standardized. And it is here that the concept of semantic applications starts to emerge.

Intelligent applications require that two different computers effectively "understand" each other, a fundamental change from the way the Web operates today. For example, today's HTML-based Web is conceptually much like the postal system in that the computers that send and receive Web pages typically have no knowledge regarding the content that is being transmitted.

In contrast, a truly intelligent Web application could behave as if it actually understood terms such as price, product, availability, name, address and date. With these types of capabilities and standards, computers should be able to sift through information and process transactions in a much more systematic way.

The idea of widespread semantic standards has gotten a tremendous boost from none other than Tim Berners-Lee, the primary architect of today's Web, and now the director of the W3C. Since the late 1990s, Berners-Lee has been using the

phrase *semantic Web* to describe a Web increasingly capable of interpreting its vast range of content. The range of possible semantic Web applications includes improved search systems that could, for example, reliably distinguish a request for a high school photograph of Martin Luther King, Jr. from a request for a photograph of Martin Luther King Junior High School; as well as improved system interoperability, for example, a semantically-enabled Computer-Aided Design and Manufacturing (CAD/CAM) system that could understand the manufacturing requirements of a particular design and link directly to a parts-and-inventory system that could then automatically generate overall materials requirements and their related costs, all from the initial design.

The Semantics of Semantics

My thirty-year-old *Webster's Seventh New Collegiate Dictionary* defines semantic as "of or relating to meaning in language." In technology, the word *semantic* is often used in two different but important ways. Sometimes it refers to "intelligent" Web applications capable of essentially "understanding" each other. At other times, it refers to better ways to manage and use ever-rising quantities of digital information. The task of managing content is already well underway and primarily evolutionary in nature. In contrast, the development and use of intelligent applications is still much more speculative, but has the potential to become a major new source of IT value and usage.

Although managing enterprise content and building intelligent applications would seem to be very different things, they share some common terminology.

Metadata. This basically means "data about data." Today, most large-scale metadata creation efforts are automated. While many companies have yet to begin the process, there is a great deal of work being done by international standards groups, industry specific associations and even open-source communities. Indeed, over time and without much fanfare, a vast new infrastructure of information categorization standards is being built, often with little direct involvement of the IT supplier community. Accurate and consistent metadata will be one of the keys to many new forms of both improved content management and more intelligent Web applications.

Taxonomy. This term traditionally refers to the study of the general principles of classification. It is widely used to describe computer-based systems that use hierarchies of topics to help users sift through information. Many companies have developed their own taxonomies, although there are also an increasing number of industry standard offerings. Additionally, a number of suppliers, including Applied Semantics, Autonomy, Verity and Semio, provide taxonomy-building software.

Today, business taxonomies are mostly used for managing content, but they will eventually drive intelligent applications as well. However, this latter type of usage will often require information to be defined more rigorously and mutually exclusively, in the more traditional scientific sense of taxonomy. This leads us to the idea of ontologies.

Ontology. From an IT industry perspective, the word ontology was first used by artificial intelligence researchers and then the Web community to describe the linguistic specifications needed to help computers effectively share information and knowledge. In both cases, ontologies are used to define "the things and rules that exist" within a respective domain. In this sense, an ontology is like a rigorous taxonomy that also understands the relationships between the various classified items.

For example, there are basically five main types of streetlights—red, flashing red, yellow, flashing yellow and green. These colors could be said to define a simple traffic light taxonomy or classification system. However, each light is also associated with a set of rules—go, stop, slow down. This combination of a taxonomy and its predefined properties results in a simple traffic system ontology.

While this distinction is real enough, many IT industry participants do use the two terms largely interchangeably. My own view is that it would be great if we used the word ontology when referring to carefully defined intelligent applications, and taxonomy when we're just organizing and sifting through content. However, since most people can't use either term without feeling pretentious, we're a long way from having this level of generally understood language. To most of us, it's pretty much all still a matter of semantics.

Objects. For those organizations that deal with long documents—such as books, manuals, courses and training programs—semantic systems also provide a way of managing large numbers of discrete "information objects," which I like to define as the smallest unit of text, image, sound, video or data that is both useful and self-contained.

These "chunks" of content could be training in a specific area or a useful bit of legal or medical expertise. The process of breaking up a larger collection of content into directly addressable, self-contained objects is often referred to as "content chunking." Easily accessible chunks of content are an essential part of turning today's often imprecise and inaccurate Web searches into a more reliable, transactional experience. In this context, taxonomies can be used to help individuals efficiently locate and sift through large numbers of discrete objects.

Emerging Information Standards

One of the amazing things about semantic applications is that once one begins to see the world in this way, the need for information standardization begins to appear just about everywhere. Today almost every major industry has some sort of standardization effort. Groups such as RosettaNet (electronics), STEP (manufacturing), Acord (insurance), Medbiquitous (medical), Chematch (chemicals), the Open Travel Alliance (Travel) and STAR (automotive) are typical of literally dozens of industry-specific efforts to define unique information standards. Many of these efforts are either formally part of or work in close conjunction with various business exchanges, since shared terminology and information formats will be such an important part of many of these endeavors. The recent merger between RosettaNet and the retail-oriented Uniform Code Council is an important example of the need for broad industry cooperation.

Not surprisingly, there are also equally important cross-industry committees, associations and consortia. These organizations exist because many terms—such as names, addresses and phone numbers—tend to be used in just about every business, and therefore cross-industry standardization is also very important. Of particular long-term interest is the Defense Advanced Research Projects Agency's (DARPA) so-called DAML (DARPA Agent Markup Language) project. As the group most directly responsible for the initial funding that led to the Internet, DARPA's work is always worthy of special attention. DAML is envisioned as an extension to both HTML and XML, designed to support the inclusion of ontologies.

Indeed, if anything, there are too many overlapping groups, which is why lead customers in particular business areas are still so important. Many companies will be more than happy to have a clear path to follow. Unlike most IT vendor standards, customer-created metadata standards themselves are not typically seen as a major competitive battleground. On the other hand, the effective *use* of these standards can become an important area of competitive advantage.

The fact that it is now human language itself, not computer programs, that needs to be standardized lends a level of both intellectual interest and potential long-term significance. While effective development and implementation are by no means guaranteed, given the current momentum, there is good reason for a long-term sense of optimism.

Implications for IT Industry Leadership

Semantic applications, even more than Web services, are principally areas for customer responsibility. Indeed, the eventual emergence of more intelligent applications will mark a decisive shift toward a customer-driven industry. Whereas IT suppliers are actively promoting Web services, only customers can really build semantic applications. For example, the General Accounting Office has recently suggested that federal agencies might want to hold back on advanced XML implementations until an overall government standards and terminology strategy is clearly in place.

This public sector example reflects the fact that in order to take advantage of higher levels of business interoperability, it is customers that have to agree on and accept (and possibly develop) shared taxonomies and ontologies. They then have to implement these new forms of rules and syntax in a consistent and disciplined fashion for structured and unstructured data alike. Ideally, this industry terminology should be easily adaptable to new uses, widely available to relevant

participants and, ultimately, translatable into multiple national languages. All of this will require considerable time and energy, which are really just proxies for cost and commitment.

More broadly, the need for much more formal labeling and language suggests a new phase of IT industry focus. Whereas the PC and Internet industries were once described as the "revenge of the nerds," looking ahead, the emphasis on detailed classification and information management might well be described as the "revenge of the librarians."

Semantic systems present a different set of challenges and require a different set of business skills. The need for fancy words such as semantics, taxonomies and ontologies suggests a world of more rigorous thinking and information handling, a significant cultural change from today's mostly freewheeling Internet.

Importantly, much of this work will have to take place industry by industry, which makes it even more difficult for suppliers to influence the overall pace of change, other than through their services organizations. More pointedly, each industry will have to establish its own semantic dynamics in terms of business cooperation and value-chain leadership. This means that the rate of individual industry change could vary widely.

The more that IT activities become focused on content and industry-specific standards, the greater the shift away from a supplier-driven industry. And while neither Web services nor semantic applications will quickly become pervasive, and while both may never completely fulfill all the promises of their proponents, they are strong evidence that a customer-driven IT industry is now emerging. In the end, customers will have to decide which new standards, applications and systems are worth implementing, and on what timetable they will be built. In this sense, they are increasingly setting the overall pace of IT industry innovation.▣

Reprinted by permission of Harvard Business School Press. Excerpt from *Customer-Driven IT: How Users are Shaping Technology Industry Growth*. Copyright 2003 by David Moschella; All Rights Reserved.



Where are you on the semantic time-line? Still at the beginning of the learning curve? Or way out in front?

reader comments:

—[A very good overview of the landscape!](#)

Anca Mosoiu

Selected comments may also be posted in [Answering Machine](#)

* Optional - we will neither sell nor display your personal information.

  **Sponsored Links:**

 ["The Road to Authentication". A journey you won't want to miss. Webcast.](#) 