LEADING EDGE

The Semantic Web

R. Jason Weiss
Development Dimensions International

When the Web originally burst into public awareness, it offered limited functionality relative to the excitement it generated. Let's take a moment and recall the experience offered by early Web browsers:

- Simple text formatting using preset paragraph styles and font formats;
- Graphic images and links to multimedia and other types of binary files;
- Hyperlinks to content within a document and to other documents across the Internet.

It really doesn't seem like much when we look back at it—early browsers didn't even offer functionality to display tables of information. Yet the Web caught the imagination of a variety of people and groups who saw that it could be much more:

- A tool for electronic commerce: A case in point is former IBM CEO Lou Gerstner's immediate reaction upon seeing the Web. He asked, "Where's the buy button?" (Hamel, 2000, p. 159);
- A way to assemble and format information to interest and engage the viewer. This spun off into several technologies:
  - Scripting technologies such as JavaScript (also known as ECMAScript), and Active Server Pages (ASP), which helped developers create more interactive, dynamic Web pages;
  - Cascading Style Sheets (CSS), which supported sophisticated page layout similar to print media;
- A platform for complex applications written in Java to be served to vastly different types of computers;
- A base for automated systems that can independently make sense of and harness the information represented across the pages of content and databases connected to the Web.

That last bullet is where the Web departs from familiar experience and enters the realm of unfulfilled visions spun by technologists and science fiction writers. Progress on "intelligent agents" has been hampered by a lack of intelligence for the agents to apply—means for such systems to understand and work with information they might encounter across the untamed Web. There is no means for software to independently attach meaning to the words and concepts plainly expressed on Web pages. This is one reason that Web search engines can be so frustrating—at best they can match patterns of words, but cannot discern differences in meaning across similar word pat-

terns. The Web doesn't know if a "sports book" is a piece of literature or a place you go to bet on your favorite team.

Helping software to use the Web intelligently is the promise of the Semantic Web, promoted by Tim Berners-Lee, widely acknowledged as the Web's inventor. The Semantic Web is a set of technologies that graft information onto Web content that enables software to discern and make sense of it. With this newfound capability in hand, sophisticated software agents will then be able to approach their promise of performing relatively sophisticated decision-making tasks on a user's behalf. According to Berners-Lee, we can expect to see practical applications of the Semantic Web by about 2005. This edition of **Leading Edge** explores the Semantic Web and its implications for industrial-organizational psychology.

How Will the Semantic Web Work?

The Semantic Web is a cluster of three technologies that build on each other to identify key information on the Web, add useful supporting information, and relate it all to a larger, meaningful framework of concepts. These technologies are briefly summarized below. The interested reader is directed to www.semanticweb.org for a more detailed review.

eXtensible Markup Language (XML) Identifies Information

To help software use information effectively, we must first help it to identify meaningful information. For example, the five-digit number *68182* is a meaningless stream of digits until you identify it as a ZIP code. In the Semantic Web, this identification process is accomplished using XML, a language that supports tagging electronic information with meaningful, user-defined labels. (See Weiss, 2001 for a brief overview of XML and tags). XML turns the otherwise indistinct five-digit number *68182* into an identifiable piece of information by enclosing it in opening and closing tags as follows: <ZIPCode>68182</ZIPCode>. When XML tags are used on a page, it is easy for software to locate and parse the information for subsequent use.

The Resource Description Framework Attaches Detail and Meaning to Information

Knowing that *68182* is a ZIP code is better than knowing that it is a just a string of five digits, but that's still not much to work with. Once useful information has been identified, the next step is to locate associated information that may also play a useful role. This information is supplied by the Resource Description Framework (RDF), which adds detail to XML tags in the form of simple declarative statements identifying properties associated with the tags and their associated values. For example, RDF information sur-

rounding the ZIP code tag described earlier might describe the standard formats that ZIP codes may take. "Format" is the property at hand, and the numerical representation (a five-digit number) is the associated value. Any quantity of informative detail can be attached using the simple RDF scheme. Further, RDF information may link to other places on the Web where additional, related information may be gathered. For example, the RDF information in our ZIP code example may include the fact that the home page for information on ZIP codes is http://www.usps.gov.

Stepping back for a moment, it is easy to start to see the potential for the Semantic Web. XML tags locate information a computer can process from within otherwise indistinguishable content. RDF adds detail and can link to other sources which may independently offer additional information. Very impressive, but the term "Semantic Web" suggests a deeper processing of information than we have explored so far. Such processing is accomplished using ontologies.

Ontologies Define Concepts and the Relations Among Them

Where XML tags and RDF information are closely tied to particular information, ontologies take a larger view. Ontologies are taxonomies of entities and descriptions of the rules and relationships governing them. As such, ontologies supply the "meta-knowledge" required to make sense of and operate on specific information. To go back to our ZIP code example, an ontology page would describe the larger framework into which ZIP codes fit (street addresses), how ZIP codes are associated with other information in that framework, and rules surrounding their use. Ontology pages can also resolve conflicts that might otherwise derail straightforward operations, such as the differences between the terms "ZIP code" offered by one database and "postal code" offered by another.

Putting It All Together

In the future world of the Semantic Web, the human user will continue to see Web pages just as we have come to enjoy them since the Web came to prominence. In the background, invisible to the human user, XML codes, RDF information, and linked ontology pages will support automated software systems that process and act upon the information in relatively sophisticated ways. Truly intelligent agents may remain the stuff of science fiction for now, but at least the Semantic Web offers a path for software to approximate meaningful information processing and decision making.

In many respects, this feels like the road to a satisfying conclusion. It will be very nice to instruct my software agent to find that article I read in *The Economist* last year on the future of the "free agent" and have it figure out the details and follow through without further input from me. As a software con-

sumer, I look forward to the day when I will enjoy such flawless service. As an I-O psychologist, however, I wonder what our next step might be.

The Semantic Web and I-O Psychology

Unlike many other Web technologies, the Semantic Web by its nature holds particular implications for our field through its focus on constructing meaning through information processing. To take a simple example, knowing the relations among the concepts of knowledge, skills, abilities, and other characteristics permits us to work with associated information in important ways. For these concepts not to be represented and distinguished on the Semantic Web would limit our use of them in such a sphere, signifying something of a step back for our science as technology lurches forward. This would clearly be an undesirable end. On the other hand, having these concepts represented on the Semantic Web may enable us to use them in new, sophisticated ways as we leverage the strengths of automated systems. Simply, as the Semantic Web moves forward, it behooves us to take advantage of it—and it is especially important that we not be left behind.

Having communicated this sense of importance, I feel the need to step back and ask how we may best leverage the Semantic Web in I-O psychology. It is one thing to not be left behind. It is another to take some time and look at how we may make the best use of it. The technology is still years from producing its first application, and the examples I have read about seem to do fairly unexciting things—scheduling appointments and the like. Certainly, in the context of our field, we can find more important and valuable uses for automated systems that can process information in sophisticated ways. What are your ideas on where the Semantic Web can take I-O? Let me know at jason.weiss@ddiworld.com.

References

Hamel, G. (2000). *Leading the revolution*. Boston, MA: Harvard Business School.

Weiss, R. J. (2001, October). Six things you should know about XML. *The Industrial-Organizational Psychologist, 39*, 30–34.