FROM *IEEE INTELLIGENT SYSTEMS*

AI, Agents, and the Web

Knowledge Management and the Semantic Web: From Scenario to Technology

Paul Warren • BT Research

A new vision of the Semantic Web as a knowledge management environment introduces new requirements, including the ability to semiautomatically learn ontologies and extract metadata.

JAN./FEB. 2006

In 2001, Tim Berners-Lee, James Hendler, and Ora Lassila wrote a seminal article in which they envisioned the Semantic Web.[1] The article's general tenor, and particularly their scenario demonstrating how the Semantic Web might be used, was directed toward identifying and using services on the Web. The pursuit of this vision continues.This article describes a complementary vision—a vision of the Semantic Web as a knowledge management environment.

All knowledge workers—whether they're historians, physicists, financial advisors, or simply individuals researching a topic for personal reasons—share certain characteristic activities. One of the chief among these activities is the need to locate relevant information rapidly. In this respect, the World Wide Web has been valuable to knowledge workers.

However, the current WWW has little metadata to describe its information, giving rise to several limitations. In particular, the search process lacks precision because it's based on a search for matching text strings. In the Semantic Web, much—possibly most—of the information will be semantically marked up. At the least, articles will identify their author, date, and subject matter. This identification won't be through a text string, which can be ambiguous, but through a uniform resource identifier (URI) unique to the article.

The current Web's limitation is due to its use of HTML, a markup language that describes format but not semantics. For about the last eight years we have seen the wide-scale use of XML, which provides the foundation for the languages being used in the Semantic Web. XML's prevalence is one factor making the time now ripe for the Semantic Web.

Knowledge management scenario

To give the Semantic Web vision substance, imagine a political scientist, Sally (see figure 1), who wants to research the extent to which British Prime Minister Tony Blair's stance on Zimbabwe has changed over a year and what factors might have caused that change. In the world of the Semantic Web, Sally could search for everything written by Blair on this topic over a specific time period. She could also search for transcripts of his speeches. Information markup wouldn't stop at the article or report level but would also exist at the article section level. So, Sally could also locate articles written by political commentators that contain transcripts of Blair's speeches.

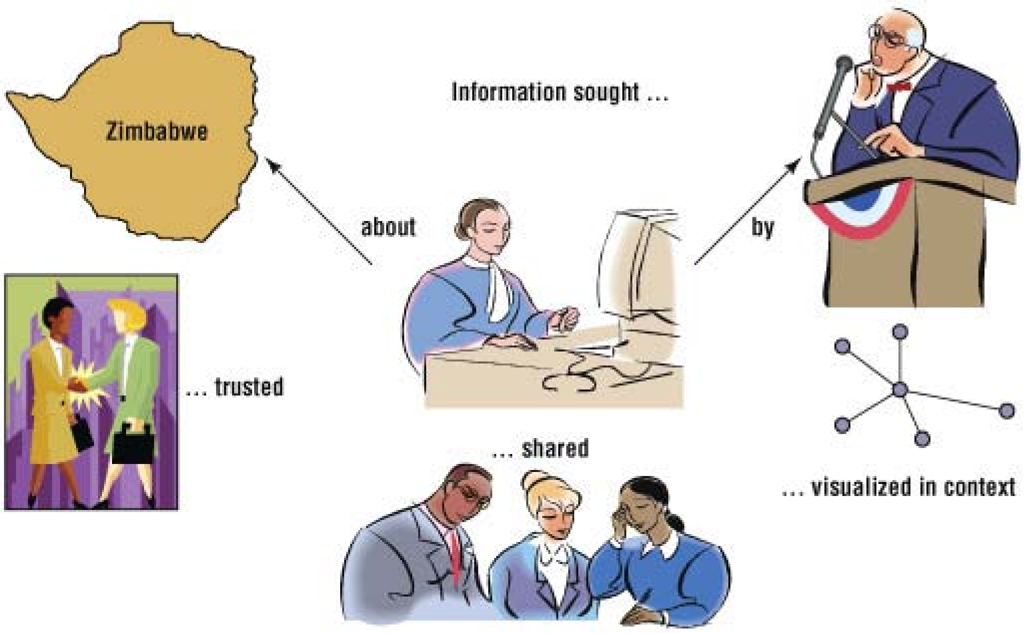

Figure 1. The Semantic Web—a knowledge management scenario.

Having read everything she could find that Blair wrote and said, Sally wants to consider how others have reacted to, and possibly influenced, his stance on Zimbabwe. She could, of course, do this by searching for articles on Zimbabwe by other leading figures that mention Tony Blair. One such figure might be David Cameron, the British parliament's Leader of the Opposition. Again, she can ensure that she locates only those articles written by that particular David Cameron. Moreover, she can ensure that in the located articles in which David Cameron refers to Tony Blair, he is, in fact, talking about Prime Minister Tony Blair and not another Tony Blair. In Sally's case, this latter occurrence is hardly likely. However, Sally could be a physicist, historian, lawyer, accountant, or whatever, and the subject matter could be anything at all.

To enhance the search process, reasoning algorithms will be distributed across the Semantic Web. Knowing that "Tony Blair" and "prime minister" are equivalent, the algorithms will deduce that text written by the "leader of the Labor Party" was written by Tony Blair, because the Labor Party currently forms the British government and hence its leader is the prime minister.

Sally has now searched for articles about Tony Blair and Zimbabwe written by all the key actors on the British political stage. She wants to expand the search to include articles on this topic written by any author. Even the Semantic Web will start returning many more hits than she can deal with. Sally wants to visualize all that has been written on the subject. She wants to understand the relationships between the articles—for example, which articles appeared in the same journals or newspapers. She wants to understand the relationships between the authors—such as who has worked together or published jointly, or are members of the same organizations. She also wants to understand how these relationships have changed over time. Her Semantic Web browser lets her do this. She can visualize these relationships in two or three dimensions and navigate her way through the information space. This lets her not just understand these relationships and how they've influenced particular individuals' views on the topic, but also identify clusters of similar articles and articles that appear unique. This ensures that while she can't read everything, she does read something representative of every viewpoint. Reasoning algorithms will make deductions about these relationships to help her understand the linkages between people and organizations. As Sally continues searching, her system will enhance her defined search strategy with metadata extracted from the documents she finds most interesting. The system will often know better than Sally how to precisely define her interests.

Sally is also keen to know what other workers in her field have found valuable. A collaborative environment accessible via her software toolset lets Sally and her coworkers share interesting articles. It does this by comparing the metadata associated with each article with metadata defining each of her colleagues' current interests. When Sally wants to share an article, she no longer needs to think about which of her colleagues would want to read it. She merely saves the article, or rather its URL, to a collaborative space where the metadata associated with the article will identify a potential readership. She can, of course, also link her own comments on the article. So, Sally can look for articles that her colleagues have found interesting and that are relevant to her current work. The software identifies an article's relevance strictly in terms of its relationship to Sally's interests. Sally's colleagues must decide whether a document has sufficient value for anyone to spend time reading it.

In addition, Sally must be sure that the information she's accessing is accurate. The possibility of hoaxes and simple mistakes on the public WWW is enormous. Sally's browser uses technology such as public-key encryption to ensure that when she accesses a respected newspaper, she isn't the victim of a hoax Web site. Even when she's certain of the provenance of what she's reading, can she really trust it? Can she trust the comments by a political observer unknown to her, perhaps on the other side of the globe?

Outside the WWW world, Sally has built a web of trust. She's learned to trust a particular writer, having read many of that person's books and articles and even having met the writer on several occasions. This writer now speaks highly of a young political scientist who hasn't yet written much. So when Sally sees an article by the newcomer, she's inclined to trust its accuracy. Developers are incorporating this same philosophy into the Semantic Web, so that chains of trust evolve automatically and the Web can even quantify trust. In Sally's domain, individual political scientists ascribe a trust factor to a particular piece of work or to a particular individual's work. Sally can view these trust relationships and even determine to what extent warring camps exist among workers in her field.

Of course, none of this is free. Sally can do all of this because of the wealth of semantically annotated information available to her. Much of this information has come from legacy data, which existed before the Semantic Web's development. Authors will use software tools to help mark up this legacy data. Ideally, such tools would be fully automatic. Where high accuracy is required, they're likely to be semiautomatic, requiring some human intervention.

Sally is interested chiefly in topical information created since the Semantic Web's development. Authors will have created much of this information using annotation tools that encourage the insertion of metadata by the documents' authors. This will be done partly by making it easy for those authors to insert metadata and by suggesting metadata to authors as they create a document. Sally must play her part in this process while she's writing her own articles.

When data is annotated, whether semiautomatically or manually, it's done against a framework, or *ontology*. (Thomas Gruber defines an ontology as "a specification of a conceptualization."[2] An ontology's primary purpose is to share knowledge; so, an ontology is a specification of a shared conceptualization.) Part of this ontology's function is to classify and describe the documents. In this sense it's similar to a taxonomy (which I describe later). However, within the documents, the software will also identify named entities such as politicians, countries, and political parties. Ontologies let you describe relationships between these entities and from this make inferences about them.

Some ontologies will be applicable to all knowledge—for example, those describing a document's properties (author, creation date, and so on). Others will be created for specific topics—for example, to describe the British political constitution. (For a description of the use of ontologies on the Semantic Web, see Elin K. Jacob's article.[3])

Moreover, ontologies and their associated entities must evolve over time. As the practicalities of the British constitution change, so will the ontology describing it. As Sally and her colleagues create new documents, the document creation tool will evolve the ontology and entities by making suggestions about extensions and modifications. In practice, the entities will change more rapidly than the ontology, but the latter will evolve also.

From scenario to technologies

This scenario is a vision of the future from the user's viewpoint. The next step is to understand what technologies we need to support this vision. Much of the technological infrastructure is shared with the original vision of the Semantic Web as annunciated by Berners-Lee and his coauthors. However, there are some new features, which are distinctive to knowledge management.

The original vision was essentially about identifying and using services. The assumption is that those offering the services will be highly motivated to annotate the service information—that is, they'll manually create metadata so that the service description is computer interpretable. Moreover, we're dealing with relatively limited quantities of information.

The knowledge management scenario shatters both these assumptions. Although knowledge workers no doubt believe in the value of annotating their documents, the pressure to create metadata isn't present. In fact, the pressure of time will work in a counter direction. Annotation's benefits accrue to other workers; the knowledge creator only benefits if a community of knowledge workers abides by the same rules. In addition, the volume of information in this scenario is much greater than in the services scenario. So, it's unlikely that manual annotation of information will occur to the extent required to make this scenario work. We need techniques for reducing the load on the knowledge creator.

Returning to the scenario, we can examine Sally's requirements one by one. First, Sally will need a search engine that combines an ontological approach with standard text-based search facilities. This is necessary because for a long time, if not permanently, much of the information on the Web and in intranets won't be semantically annotated, and because a search strategy based on both techniques will likely be valuable.[4] Sally wants her search engine to be able to identify a person, Tony Blair, uniquely and without ambiguity. A unique URI makes this possible.

Moreover, she wants to identify information that has a specific relationship to that person—that is, Tony Blair wrote it. She can do this using the relationship "written by," which is also uniquely defined within a published ontology by a URI.

Indeed, Sally is seeking not just any articles written by Tony Blair (an object) but also specifically about (a relationship) Zimbabwe (another object). A unique URI identifies all objects, concepts, and relationships. Sally also wants to be able to identify information at varying granularity levels. She's looking not only for articles that Blair wrote but also for other documents containing quotes from Blair's writings or speeches. Thus, the metadata must describe not just whole articles but also sections within articles.

Sally then wants to find everything written, by whomever, about Tony Blair's opinion on the Zimbabwe political situation. Again, this depends on the existence of an ontology and documents annotated with metadata using the ontology. However, the volume of information and the complexity of the relationships suggest that manually created annotations won't contain a sufficient richness of information. As I noted earlier, semiautomatic techniques for metadata creation are a key characteristic of the knowledge management vision.

Sally wants to visualize relationships (such as those between articles and between their authors). The ability to do so is particularly urgent for knowledge workers overwhelmed by thousands of documents. Sally also wants to make deductions about these relationships. For example, where a cluster of similar articles exists, she might want to understand what relationships exist between the articles' authors—for instance, one author was another's student, or two authors worked at the same institution at the same time. The ability to make inferences is a key part of the original Semantic Web vision. In particular, the Semantic Web calls for robust reasoning—that is, the capacity to make deductions in the presence of the inconsistencies that will inevitably be found in the global Semantic Web.

An adaptive-search strategy that learns from Sally's feedback and the collaborative knowledge-sharing technology further differentiate the knowledge management scenario from Berners-Lee's, although the required functionality rests on the same infrastructure of ontologies and metadata.

As her searches and collaboration reveal more information sources, Sally will find that different ontologies describe the same, or similar, knowledge domains. These ontologies will contain classes, relationships, and entities that are equivalent but differently expressed. Ontology mediation, which allows translation between ontologies and ontology merging, will let us combine knowledge from the differing ontologies. The original vision of the Semantic Web clearly identified this need. Developers are beginning to introduce techniques for ontology mediation, such as in e-commerce.[5,6]

Sally uses an electronic version of her professional web of trust. Berners-Lee's original article mentioned trust only briefly and didn't mention webs of trust, although he's described this concept elsewhere. (See, for example, a transcript of his talk at the 2002 Japan Prize.)

I've already noted the need for automatic annotation of documents with metadata. In any case, the software tools will use statistical and linguistic analysis and draw on the context of the user's work. At the same time, the user interface will make inserting metadata easy and natural, with minimal disruption to the work in hand.

Automation isn't required just to extract the metadata but also to generate the ontology into which the metadata fits. Sally could, of course, easily generate an ontology to describe political science generally or some particular aspect of her subject. Often, ontology generation won't be particularly onerous. However, ontologies for specialized domains or ontologies that must be maintained and evolved as the domain changes will benefit from some automation. Again, statistical and linguistic techniques can potentially do this at least semiautomatically. Maintaining and evolving ontologies to adapt to changing circumstances will also require semiautomatic techniques.[7]

Visualization techniques can help users see the relationships in an ontology and in the associated metadata. However, no one size fits all. Some users will have more sophisticated requirements and will be able to cope with more sophisticated tools. The interfaces must always combine ease of use with the desired power level.

The Semantic Web today

Progress toward achieving the Semantic Web vision has been along several strands. Commercial software for extracting semantics from unstructured text, chiefly for taxonomic classification, is available for use in corporate intranets and by information service providers. Quite separately, consortia and other organizations are developing predefined semantics for the WWW—for example, to enable e-commerce applications—chiefly at the XML tag level. A great deal has also been done, particularly by the World Wide Web Consortium (W3C), to develop a framework of Semantic Web standards, such as ontology description languages. Researchers have developed tools for creating, populating, and managing ontologies based on these emerging ontology languages, and significant ontologies are being used already.

Extracting semantics

Enterprises already use some of the technology described in this article, in embryonic form, both to provide business information services and in technology for their corporate intranets. Currently, commercial products are limited mostly to the use of taxonomies. A taxonomy is a simple form of ontology that allows only hierarchical relationships. Document classification into taxonomy categories is fairly well developed, although far from 100 percent accurate unless guided by human intervention. Taxonomy learning is at a more rudimentary stage, requiring a semiautomatic approach with human intervention.

Several companies, such as Verity and Entrieva, offer taxonomy-based products. Customers typically receive a predefined taxonomy. It might be an off-the-shelf taxonomy describing a specific topic area, such as some financial specialization, or it might be tailored to the customer's requirements. The product might include a semiautomatic taxonomy-learning tool, and it will certainly offer a taxonomy-editing tool. Documents are classified automatically according to the taxonomy, although the user or a domain expert can overrule the classification. The system uses this feedback to correct the classification algorithm. The system might allow the application of several taxonomies to the same knowledge base to benefit users from different communities, such as finance and general management, who might have different mental frameworks. Some products provide visualization of the taxonomies. Some analyze what knowledge the user is seeking to identify experts and communities of interest. Providers of proprietary business intelligence information also use these products; for example, LexisNexis uses Verity technology.

One route to overcoming taxonomies' hierarchical limitation is through topic maps that describe ontologies. Topic maps originated in 1993 and hence essentially predate current Semantic Web discussions. Indeed, they've evolved separately from the Semantic Web and with different objectives. Their long-term future is unclear. Topic maps might be replaced by the Semantic Web framework, or they might continue to play a role in some information and knowledge management applications. Empolis, for example, incorporates topic maps in its product set.

Building semantics in the Web

Michael Uschold and Michael Gruninger have analyzed how to implement semantics on the WWW and suggest that a continuum of approaches exists.[8,9] At one extreme, the semantics are both informal and implicit, as demonstrated by Web sites that use words such as "price" without any attempt at definition, relying instead on the general consensus as to what the word means. Extracting the semantics is inevitably an uncertain process, although some shopping agents attempt to do so.

The work of the ebXML and RosettaNet consortia represents the next stage. The Organization for the Advancement of Structured Information Standards started the former jointly with the United Nations Center for Trade Facilitation and Electronic Business. (OASIS is a not-for-profit consortium "that drives the development, convergence, and adoption of e-business standards." CEFACT's mission is "to improve the ability of ... organizations ... to exchange products and relevant services effectively.") The ebXML consortium aims to develop standards, including XML standards in areas such as business processes and messaging. RosettaNet is a not-for-profit organization in the electronics and telecommunications industries with particular interest in developing supply chain standards. The essence of both approaches is that consortium members agree on the semantics as part of a standardization process. The semantics are made explicit, although they're informal in the sense that the definitions use natural language.

Moving further along the spectrum, the semantics can be defined formally—that is, using a formal language—but purely for reading by humans. The intention here is to have completely unambiguous semantics that people can refer to in the case of any disputes.

Thus far, the semantics have been made explicit but are still hardwired. Essentially, the lengthy process toward consensus stifles the innovation that has characterized so much activity on the WWW. Moreover, each consortium represents an island of consensus. We could establish dictionaries to translate between these islands, but it would require special-purpose translators.

The final stage, which is the true Semantic Web, seeks to overcome both these problems by making the semantics not only explicit and formal but also machine processable. We define semantics with respect to an ontology. In principle, anyone can publish or use an ontology. Moreover, when two ontologies describe essentially the same domain, we can define equivalences and more complex relationships between them, and translation can occur at runtime. Of course, for convenience and computational efficiency, certain ontologies will dominate certain domains. However, in principle we've restored innovation and spontaneity to the Web.

The Dublin Core Metadata Initiative is a knowledge management activity representing the final stage of Uschold and Gruninger's continuum. This organization is "dedicated to promoting the widespread adoption of interoperable metadata standards and developing specialized metadata vocabularies for describing resources that enable more intelligent information discovery systems." It achieves this by developing metadata standards and frameworks for their interoperation. The Dublin Core has implemented metadata tags such as title, creator, subject, accessRights, and bibliographicCitation using the resource description framework and RDF Schema.

The W3C's work

The W3C has made significant progress toward standardizing the languages necessary for a global Semantic Web. For example, they've finalized RDF. RDF builds on XML, providing a data-modeling framework for knowledge with a structure based on triples of subject, predicate, and object. Building on RDF, RDFS is a simple ontology-modeling language that uses concepts such as classes, subclasses, subproperties, domains, and ranges. RDFS is limited, however. The DARPA Agent Markup Language (DAML) in the US and the Ontology Inference Layer (OIL) in Europe are extensions of RDFS, offering a richer ontology language. Developers have applied their experience with these two languages to create the Web Ontology Language (OWL), which the W3C has also finalized. OWL comprises three increasingly expressive sublanguages: OWL Lite, OWL DL (DL refers to the description logics that OWL DL corresponds to), and OWL Full. The latter two are supersets of their predecessor. Which of these variants can provide all the functionality in the knowledge management scenario, or indeed in the kind of e-commerce scenario that Berners-Lee described, remains an open question.

Ontology tools

Other developers have been working on tools for using these languages—for example, to create and edit ontologies. One of the first—and still the best known—of these tools is from Stanford University's Protégé project. Protégé lets users construct a domain ontology and enter domain knowledge. The project has recently released an OWL editor as a plug-in to the existing software.

For practical applications, we must be able to manage and reason with large ontology and entity databases. This requires an ontology management infrastructure such as KAON (the Karlsruhe ontology management infrastructure), developed at the University of Karlsruhe. KAON includes several tools, such as the OI-Modeler, an ontology editor. Aduna's Sesame is another ontology store. These two tools store the ontology and entities as RDFS and RDF, respectively. KAON includes some proprietary features that go beyond RDFS.

On-To-Knowledge, a European 5th Framework collaborative project, generated the original Sesame prototype. This project also created

Spectacle, a tool for visualizing ontologies, and QuizRDF, a hybrid search engine that combines the ontological and conventional approaches.[10] The Knowledge and Information Management platform extends the ideas in QuizRDF. KIM uses an ontology with classes such as towns, companies, and people. So, relationships between entities can identify, for example, the chief officers of particular companies, letting users identify such entities in documents and make inferences about them.

Practical ontologies

The scenario I described uses ontologies to sustain a global Semantic Web. Much must be resolved before we can achieve this ambitious goal. In the meantime, ontologies are achieving significant benefits in more limited scenarios. In particular, they've been used for some time in the life sciences, where researchers must manipulate large, complex data sets. DARPA maintains a library of ontologies with uses ranging from cancer research to describing baseball, in addition to those describing more fundamental concepts, such as time. Stanford University maintains another library of ontologies, developed with the Protégé editor.

The winners of the 2003 Semantic Web Challenge (see the May/June 2004 *IEEE Intelligent Systems*) are examples of practical Semantic Web applications that are currently possible. The guest editors' introduction to that special issue makes two interesting points.[11] First, manual intervention is still clearly necessary: "although the submissions use many different sources they can incorporate only a minority of these sources automatically." Second, RDFS is sufficient for current applications; OWL's additional expressiveness isn't required. The applications also make limited use of reasoning capabilities (this is related to the relatively simple ontologies used). Whether these characteristics apply merely to the initial applications or will continue to be the case is an open question. The first-place entry is an academic application having some similarities with my scenario.[12] The article describing this application finishes with a series of challenges for the future. Some reflect the problems of harvesting data from the Web; others echo the challenges described next.

Challenges

The challenges to the knowledge management vision of the Semantic Web fall into two main categories: implementing semantic knowledge management systems on the one hand and making them work in an organization on the other.

Implementing semantic knowledge management

Currently, the main obstacle for introducing ontology-based knowledge management applications into commercial environments is the effort needed for ontology modeling and metadata creation. Developing semiautomatic tools for learning ontologies and extracting metadata is a key research area. However, performing typical business tasks can also generate metadata as a side effect. One approach to learning ontologies and creating metadata is to supplement the results of semiautomatic ontology-learning and metadata-generation techniques with information drawn from the context in which the user is working.

The development of semiautomatic techniques for ontology learning will continue for several years. Developers will need to address the issue of scale by designing techniques for increasingly larger knowledge bases. We must also investigate how to use these techniques on the "hidden Web"—that is, how to access knowledge that isn't easily accessible to Web crawlers because it's embedded in databases and only retrievable against specific queries.

Once the ontologies and their related metadata are created, they must evolve. Usage mining, which uses data-mining techniques to detect usage patterns and adapt ontologies accordingly, is one approach. Another approach is to identify new concepts and relationships that are missing in the ontologies, compared to the current text and data sources. Ontology mediation to deal with many overlapping ontologies is another significant research area.

Enhanced searching initially requires reasoning over RDF(S). (RDF(S) denotes RDF and RDFS used together.) The increasing use of OWL variants will make more complex reasoning possible. Reasoning techniques must be scalable and must account for inconsistencies in heterogeneous globally distributed knowledge bases.

The implementation of webs of trust will let us complete the Semantic Web vision. The Semantic Web's current focus is at the ontology and logic layers.[13] However, work in the academic community on trust inference calculi across distributed information systems is ongoing, and work on trust and the Semantic Web is beginning to appear.[14] See, for example, the Semantic Web Trust and Security Resource Guide. This is an area for research both into formal mechanisms for establishing trust and into the human and psychological aspects of how we can use such mechanisms.

Making semantic knowledge management work

Alongside these relatively well-defined technology challenges lies another with technological, but also human and organizational, implications. Knowledge management must be made to work for the good of its users and their organizations.

Organizations have invested a great deal of money in knowledge management resources in the past few years, and a debate as to why those resources often go unused is ongoing. Thomas Stewart argues that knowledge management systems must be targeted to a particular work group's knowledge needs.[15] He further argues that "knowledge management is much more effective if it is not a stand-alone button on somebody's PC but is integrated into a key business process." Charles Seeley has taken up this latter point, arguing that you should look for points in a business process (such as key decision points) where knowledge can be valuably used.[16] His suggestion is to "knowledge enable" business processes, particularly core or strategically important processes. To do this, we must understand where in a business process we can most profitably apply the knowledge management resource and what's needed there. We also need knowledge management software integrated with whatever packages we use to enable normal business processes. Having to move out of a user's typical working environment to "do knowledge management" will act as a disincentive, whether the user is creating or retrieving knowledge. Describing a successful knowledge management implementation, Seeley notes that "the intent was that the work environment and the KM environment ... be the same."[16]

One solution is to design business process software that encourages the user to save to, or retrieve from, the knowledge repository at key decision points. This presupposes that the user follows a well-defined process, supported by specific process-oriented software. Project work proceeding from proposal development and culminating in a final project report typifies such processes. Each stage has its own knowledge requirements. In the initial stage, the user needs information about similar successful bids. At a later stage, the user might seek more technical information to support the design process. This article's scenario describes research work that's inevitably less structured. Even here, though, we can identify processes, breaking them down into components with their own particular needs. Understanding the nature of knowledge work and how the prototypical processes relate to the tools used will require more research. Even unstructured work requires software tools, if only a word processor, and for knowledge management tools to be used they must be integrated with this normal working environment.

In a review of how ontologies can support agents on the Semantic Web, James Hendler makes the related point that "semantic markup should be a by-product of normal computer use."[17] An example he quotes is of a user who, when importing clip art into a document, also imports some associated annotation that can in turn be associated with the document. In general, judicious use of the context in which a document is being created can do much to create annotations for it.

To help make semantic knowledge management effective, researchers have developed a variety of methodologies for introducing ontology-based knowledge management into an organization. In a review of some of these methodologies, one observer commented on their failure to address business value and the social aspects of ontology engineering.[18] Methodologies must account for these factors and keep pace with ontology languages' developing functionality.

The Web's success rests largely on two principles: It satisfies the real needs of users, and its design is sufficiently simple to be scalable. To achieve the challenging vision I've outlined, developers must continue to adhere to these principles. The Semantic Web's final form is an open question. I believe there will be deep semantic interoperability within organizational intranets. This is already the focus of practical implementations, such as the SEKT (Semantically Enabled Knowledge Technologies) project, and across interworking organizations, such as supply chain consortia. In the global Web, semantic interoperability will be more limited. Organizations such as the W3C will provide the framework. As with today's Web, user needs will give it substance.

Acknowledgments

I undertook this work as part of the EU Vision project (IST-2002-38513). (For more information about Vision, see http://km.aifb.uni-karlsruhe.de/fzi/vision.) I thank my colleagues on the project, the project reviewers, and the European Commission, which funded this research. I particularly thank Alex Mädche and Andreas Abecker, who managed the project, and John Davies, my colleague at BT Research.

References

1. T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific American*, May 2001, pp. 34–43.
2. T.R. Gruber, "A Translation Approach to Portable Ontologies," *Knowledge Acquisition*, vol. 5, no. 2, 1993, pp. 199–220.
3. E.K Jacob, "Ontologies and the Semantic Web," *Bull. Am. Soc. Information Science and Technolog*y, vol. 29, no. 4, Apr./May 2003, pp. 19–22.
4. J. Davies and R. Weeks, "QuizRDF: Search Technology for the Semantic Web," *Proc. 37th Hawaii Int'l Conf. System Sciences*, IEEE CS Press, 2004, p. 40112. (pdf)
5. A. Büchner et al., "Semantic Information Mediation among Multiple Product Ontologies," *Proc. 4th World Conf. Integrated Design and Process Technology*, 1999. (pdf)
6. Y. Tzitzikas, N. Spyratos, and P. Constantopoulos, "Mediators over Ontology-Based Information Sources," *Proc. 2nd Int'l Conf. Web Information Systems Eng*. (WISE), IEEE CS Press, 2001, pp. 31–40. (pdf)
7. M. Klein et al., "Finding and Characterizing Changes in Ontologies," *Proc. 21st Int'l Conf. Conceptual Modeling* (ER 2002), LNCS 2503, Springer, 2002, pp. 79–89. (pdf)
8. M. Uschold, "Where Are the Semantics in the Semantic Web?" *AI Magazin*e, vol. 24, no. 3, 2003, pp. 25–36. (pdf)

9. M. Uschold and M. Gruninger, "Creating Semantically Integrated Communities on the World Wide Web," invited talk at Semantic Web Workshop, WWW2002. (pdf)
10. N.J. Davies, D. Fensel, and F. van Harmelen, eds., *Towards the Semantic Web: Ontology-Driven Knowledge Management*, John Wiley & Sons, 2003.
11. M. Klein and U. Visser, "Semantic Web Challenge 2003," *IEEE Intelligent Systems*, vol. 19, no. 3, 2004, pp. 31–33. (pdf)
12. N. Shadbolt et al., "CS AKTive Space," *IEEE Intelligent Systems*, vol. 19, no. 3, 2004, pp. 41–47.
13. P. Warren, "The Next Steps for the WWW—Putting Meaning into the Web," *Computing and Control Eng*., vol. 14, no. 2, 2003, pp. 27–31.
14. M. Richardson, R. Agrawal, and P. Domingo, "Trust Management for the Semantic Web," *Proc. 2nd Int'l Semantic Web Conf.*, 2003, LNCS 2879, Springer, pp. 351–368.
15. T.A. Stewart, "The Case against Knowledge Management," *Business 2.0*, Feb. 2002.
16. C.P. Seeley, "Igniting Knowledge in Your Business Processes," *Knowledge Management Rev*., vol. 5, no. 4, 2002, pp. 12–15.
17. J. Hendler, "Agents and the Semantic Web," *IEEE Intelligent Systems*, vol. 16, no. 2, 2001, pp. 30–37.
18. E.-S. Abou-Zeid, "What Can Ontologists Learn from Knowledge Management?" *J. Computer Information Systems*, Spring 2003, pp. 109–117.

Paul Warren is on the Next-Generation Web team at BT Research, where he's project manager for the Semantically Enabled Knowledge Technologies project. His research interests are the practical applications of technology to knowledge management and technology foresight. He received his MA in theoretical physics from the University of Cambridge and is a member of the IEE. Contact him at BT Research, Adastral Park, Martlesham Heath, Ipswich, Suffolk IP5 3RE, UK; paul.w.warren@bt.com.

Related Links

- DS Online's Distributed Agents Community
- "Ranking Complex Relationships on the Semantic Web"

- "Integrating Concept Mapping and Semantic Web Technologies for Effective Knowledge Management"
- "Process Oriented Knowledge Management: A Service Oriented Approach"

Cite this article:
Paul Warren, "Knowledge Management and the Semantic Web: From Scenario to Technology," *IEEE Intelligent Systems*, vol. 21, no. 1, 2006, pp. 53–59.