Editor: Dieter Fensel
Vrije Universiteit, Amsterdam
dieter@cs.vu.nl

Ontologies and Electronic Commerce

Ontologies provide support in integrating heterogeneous and distributed information sources. This gives them an important role in areas such as knowledge management and electronic commerce. E-commerce is currently facing revolutionary changes: Marketplaces are enabling new kinds of services and interactions between suppliers and vendors. The dynamic generation of supply-chain links, automatic negotiations, market transparencies, coalition forming, and online configurations of products are just a few examples of these changes. However, suppliers and vendors suffer from the same problem—they need to integrate heterogeneous and distributed product descriptions. In most cases, there is no consensus on the products making up a domain, how to describe them, and their proper product catalog structures. Even when consensus is achieved, buyers and sellers might need different views on product data. Even within a buying organization, different users might prefer to see product data differently.

Industry found these needs to be the major bottlenecks in B2B e-commerce. Large initiatives have been set up to define ontologies as a means for mediating e-commerce. Examples of ontologies with large horizontal cover are UN/SPSC (Universal Standard Products and Services Classification code—see www.unspsc.org) and UCEC (www.ucec.org), both of which want to support B2B transactions in all possible areas. (Horizontal standards try to cover all possible product areas. Vertical standards focus on a certain domain and provide much more detailed descriptions.) UN/SPSC defines a concept hierarchy to classify all products, and UCEC defines attributes to describe these products. Ontologies with a large horizontal cover are usually very shallow in respect to a certain product domain. On the other end of the spectrum, we have examples such as RosettaNet (www.rosettanet.org), which provides a vertical ontology that describes products of the hardware and software industries in detail.

Building consensual and reusable product catalogs is nothing more than building an ontology for a certain domain. Consequently, we have seen fruitful cooperation in this area between researchers working on principles, methods, and tools that support ontology development and industry applying these things in e-commerce. Examples include VerticalNet (www.verticalnet.com), which houses vertical marketplaces and ontology departments; Interprice (www.interprice.com), which develops ontology-based information access for customers in B2C transactions; and Content Europe (www.contenteurope.com), which provides advanced ontology-based content management support for B2B e-commerce. In a nutshell, the cooperation of research on ontologies and online commerce are currently strong and successful.[1]

Ontologies are only the first step toward realizing the full power of online e-commerce. Ontologies enable machine-understandable semantics of data, and building this data infrastructure will enable completely new kinds of automated services. Software agents can search for products, form buyer and seller coalitions, negotiate about products, or help automatically configure products and services according to specified user requirements. The combination of machine-processable semantics of data based on ontologies and the development of many specialized reasoning services (also called problem-solving methods[2]) will bring the Web to its full power, enabling what Tim Berners-Lee called a semantic Web.[3]

—Dieter Fensel

References

1. D. Fensel, *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*, Springer-Verlag, Berlin, 2001.
2. D. Fensel, "Problem-Solving Methods: Understanding, Development, Description, and Reuse," *Lecture Notes on Artificial Intelligence (LNAI)*, no. 1791, Springer-Verlag, Berlin, 2000.
3. T. Berners-Lee, *Weaving the Web*, Orion Business Books, London, 1999.

1094-7167/01/$10.00 © 2001 IEEE

Ontologies and Online Commerce

Deborah L. McGuinness, Stanford University

Electronic commerce is exploding—Forrester research (www.forrester.com) predicts global e-commerce will reach 6.8 trillion dollars in 2004. As the market segment grows, it has expanded into broader content areas, which increases the need for thoughtful content organization and browsing support. Ontologies can facilitate organization, browsing, parametric search, and in general, more intelligent access to online information and services. This essay discusses some ontological trends that support the growing domain of online commerce—you can view it as an update to a previous paper in which I identified ontology-enhanced e-commerce application issues and opportunities.[1]

Taxonomies

The online market "discovered" ontologies many years ago. Online yellow pages were organized by a standard industry code scheme to help people navigate. Yahoo took this a step further and made a significant impact with its use of a taxonomy and human tagging to help its users navigate content. What Yahoo introduced is repeated on most content dissemination, search, and commerce sites today—most have five to 15 top-level categories of topics, allowing some kind of drill-down feature into more specific categories and giving some indication of the amount of content in any one area. Most sites expose some kind of topic or class-generalization hierarchy to support browsing. In fact, it would be unusual today if a site did not provide at least three levels of class–subclass organization to help users navigate.

Debates arise over whether taxonomies should contain only strict subclass relationships (for example, if every instance of a more specific class is necessarily an instance of the more general class), if single or multiple parents are allowed in the class hierarchy, and a few other technical issues. However, it is not typically disputed that some kind of class organization is required to support browsing and user expectation settings. It is a common (and accurate) belief that some sort of taxonomy of classes is required for online sites today. Academics such as Dieter Fensel[2] suggest that ontologies provide a silver bullet for e-commerce, and many companies are interested in ontologies. Corporations such as VerticalNet have built significant ontological organizations to support their commerce offerings. However, corporate interest is not restricted to newer technology companies such as VerticalNet, CommerceOne, Cisco, and Yahoo—established companies such as AT&T and Daimler Chrysler are exploring and building ontology expertise. Online commerce will continue to consider ontologies as a necessary component to support at least navigation, user-expectation settings, and parametric searches.

Information Sources

Assuming everyone needs some sort of class taxonomy, we need to find sources of taxonomic information. Fortunately, many taxonomies are available today, and some class organizations that existed prior to the e-commerce revolution are being reused. Two examples are the standard industry classification scheme (the SIC codes used in the Yellow Pages are now called the North American Industry Classification System—see www.ntis.gov/product/naics.htm) and the unified medical language system (UMLS, which is used for medical literature—see www.nlm.nih.gov/research/umls). These are interesting examples, because they are large and long-lived efforts at building large public taxonomies for reuse.

Potentially more interesting is the proliferation of organizations that are building and disseminating freely available class taxonomies to facilitate e-commerce or other online organization. For example, the joint effort between the United Nations Development Program and Dun and Bradstreet to produce the UN/SPSC code (www.unspsc.org) is an effort aimed at producing a taxonomy for classifying both products and services for use throughout the global marketplace. Many B2B sites today are complying (and extending) the UN/SPSC for their own use. Some consortia are being formed, such as RosettaNet (www.rosettanet.org), a self-funded, nonprofit organization that is a consortium of major information technology, electronic components, and semiconductor manufacturing companies working to create and implement industry-wide e-business process standards. It produces controlled vocabularies for process interfaces, dictionaries, product and partner codes, and exchange protocols. Grassroots taxonomy organizations are growing as well. Open Directory (also called DMOZ—www.dmoz.org) is aiming to become *the* user-generated comprehensive dictionary for the Web. DMOZ asks volunteer editors to submit categories and classifications of pages, and at press time, it had over 33,000 editors, 336,000 categories, and 2.3 million sites.

Beyond Monolithic Taxonomies

Today, it is actually becoming less an issue of building one's own class taxonomy but more an effort at identifying what is available for reuse, what portions of existing information are useful for someone's particular needs, how the assumptions of the existing knowledge source fit with a customer's assumptions of reuse, how a customer merges two or more existing knowledge sources, and how a customer fills in the holes that inevitably exist in the available information.

To answer these questions, application designers must understand their content domain and likely content sources, identify how they are likely to use the ontologies,

articulate their needs and assumptions, and attempt to predict their future needs. Needs for ontologies might be simple, such as our use in FindUR (later deployed on AT&T's WorldNet site),[3] in which we used ontologies as a source of information for query expansion. If simple query expansion is all that is required, then simple class taxonomies are adequate.

Even if a simple class taxonomy is all that is required, you might need to combine taxonomies to generate an adequate taxonomy. You might want, for example, to extend the UN/SPSC to have more depth in certain areas, by adding detailed subclasses from another ontology. Possibly more common is the need to use some branches from one ontology and some from another. This forces a user either to merge ontologies, possibly using a tool such as Chimaera[4] or Prompt[5] or using an approach such as the one advocated in DAML,[6] in which users subscribe to many ontologies and choose terms from specific ontologies in their new ontologies. These approaches, in combination (which is supported by the merging tool environments), support building large ontologies from component ontologies.

Most e-commerce sites will not survive by only using simple query expansion, exploiting only class taxonomies—they need some form of structured information to support parametric search. Forrester, for example, claims that "surgical search" is a requirement for future search offerings. In this mode, users expect to be able to present a very precise query for an item—possibly a monitor with a diagonal of at least 19 inches, a resolution of at least 1,024 × 780, a manufacturer of either Sony or Viewsonic, and so on. The ontology must capture class information (for example, the subclasses of monitors) as well as all the parameters that make sense to specify for a class and preferably the range restrictions and types of fillers. This type of search is also called parametric search, and it exists on many online sites today, including the simple consumer search on wine.com or more sophisticated interfaces that approach configurators (such as the one on Dell's site). To support a parametric search, parameters need to be identified on a per class basis. Also, to better support the user, restrictions on the parameter should be specified (for example, price should be in dollars or at least floating-point numbers or integers, manufacturer lists might be stored, common diagonal values might be stored for monitors, and so on). It is a challenging task to find existing ontologies with enough information concerning parameters. Therefore, it is likely that most application developers will need to augment the freely available ontologies.

Most e-commerce sites will not survive by only using simple query expansion, exploiting only class taxonomies—they need some form of structured information to support parametric search.

Markets exist today for controlled vocabularies for populations of ontologies. These controlled vocabularies should contain class, property, and property restriction information. Additionally, an ontology tool market is growing for manipulating ontologies, because customers are looking for tools for ontology evolution. Tools for ontology building, maintenance, validation and verification, merging, and evolution are all becoming increasingly important to support the needs for ontologies in support of online commerce.

Future issues

Ontologies are becoming increasingly important as a component of online commerce offerings. They are useful (and arguably necessary) in supporting at least navigation, browsing, user-expectation setting, and parametric search. Sources of class taxonomies exist, tools for piecing ontologies together are growing, and some sources of parameter information are becoming available. Challenges remain for users in reusing available ontological information, because as standards are still forming, most vocabulary information needs to be augmented, and although some tools exist, most are still on a development path to becoming complete tool suites suitable for mass deployment. These challenges are surmountable and they should diminish over a short time. Efforts such as the DAML program might be one source of many useful tools for these efforts. Commerce itself will likely be another source of ontology tools.

References

1. D.L. McGuinness, "Ontological Issues for Knowledge-Enhanced Search," *Frontiers in Artificial Intelligence and Applications*, IOS-Press, Washington, D.C., 1998; www.research.att.com/~dlm/papers/fois98-abstract.html (current Jan. 2001).
2. D. Fensel, *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*, Springer-Verlag, Berlin, 2001.
3. D.L. McGuinness, "Ontologies for Electronic Commerce," *Proc. Artificial Intelligence for Electronic Commerce Workshop of the AAAI Nat'l Conf*, AAAI Press, Menlo Park, Calif., 1999; http://ksl.stanford.edu/people/dlm/papers/aaai99-abstract.html (current Jan. 2001).
4. D.L. McGuinness et al., "An Environment for Merging and Testing Large Ontologies," *Proc. Seventh Int'l Conf. Principles Knowledge Representation and Reasoning (KR2000)*; www.ksl.stanford.edu/people/dlm/papers/kr00-abstract.html (current Jan. 2001).
5. N. Noy and M. Musen, "PROMPT: An Algorithm and Tool for Automated Ontology Merging and Alignment," *Proc. AAAI-2000*, AAAI Press, Menlo Park, Calif., 2000, pp. 450–455.
6. J. Hendler and D. McGuinness, "The DARPA Agent Markup Language," *IEEE Intelligent Systems*, vol. 15, no. 6, Nov/Dec 2000, pp. 67–73; www.ksl.stanford.edu/people/dlm/papers/ieee-daml01-abstract.html (current Jan. 2001).

Ontology Associations

Ellen Schulten, Heloise Ontology Associates

E-commerce is the hype of our young millennium and with it information technology. In the old days, the business world considered IT as a cumbersome support function—things had to run smoothly, but no one really wanted to deal with it. Outsourcing was often the best option, but if not, the IT department was still rather isolated from the business processes. This has all changed in the new economy—now everybody turns their attention to IT. Traditional companies invest billions and build spin-off companies with the sole purpose of turning their businesses into e-enabled businesses. The stock market now devotes a special place to IT companies, one that is followed more closely than ever. Newspapers publish articles on intelligent agents,

and common business people discuss the advancement of standard ontologies in their industry area as if they had always talked of these topics.

The view from business

From my perspective—which is not that of the scientific community but rather the business world—state-of-the-art information technology never achieved this widespread attention. Members of the IEEE could argue that the same applied to AI research in the eighties but that the necessary patience of the real world to arrive at fully functioning models was absent and led to disillusion and eventually disinterest. Following this analogy, the e-commerce hype will follow a similar path—once the novices out there understand that state-of-the-art technologies are not yet plug-and-play solutions, the interest will gradually fade, and research and business will each go their own way.

However, in my opinion, this will (or should not) be the case. Let's first look at the reasons for this huge interest in e-commerce and state-of-the-art technology from the business world. E-commerce is said to significantly alter business as usual, but what is so new about it? The possibility of conducting transactions online already existed in the seventies. An important contribution to this technology was provided by the initiative of the United Nations to arrive at a common standard for electronic messages passing between partners in international trade. These efforts eventually led to the Electronic Data Interchange For Administration, Commerce, and Transport (EDI/EDIFACT) standard. However, EDI implementations were so costly and labor-intensive that they were confined to only a few industry giants. Now almost every company can replace inefficient and error-prone processes such as calling for the availability of a product, faxing a purchase order, and waiting for a confirmation with a clean and fast Internet connection. Furthermore, whereas an EDI solution connected one vendor with one buyer, the Internet allows the possibility of *N* companies conducting business with *M* companies on a shared Internet platform. This is what is referred to as an *electronic marketplace*.

The e-marketplace

Just imagine the influence of an e-marketplace on traditional business relationships. Existing supply-chain relationships and pricing agreements would be cut loose due to an enormous increase in market transparency—thus heavily impacting marketing and pricing strategies and changing revenue models. (This is already obvious in the B2C area, where pricing agreements on consumer goods such as books and music are under heavy pressure.) Intermediates such as the wholesaler and the distributor

Deborah L. McGuinness is the associate director and senior research scientist of the Knowledge Systems Laboratory at Stanford University. She has built and deployed numerous ontology environments and ontology applications, including some that have been in continuous use for over a decade at AT&T and Lucent. She is coauthor of the current ontology evolution environment from Stanford University and of one of the more widely used description logic systems—CLASSIC from Bell Laboratories. She is also coeditor of the recently released DARPA Agent Markup Language. She is on the advisory board for Ontology.org and Powermarket, the executive council for the AAAI, and the executive steering board for the international organization for description logics, the Ontology Inference Layer effort. Contact her at dlm@ksl.stanford.edu.

Ellen Schulten is an independent consultant in B2B e-commerce. She has a background in change management consultancy at E&Y in the Netherlands and Cap Gemini PBS in the US. She is highly interested in the use of ontologies and ontology tools for e-commerce. She is also active in several e-commerce start-ups and is cofounder of Heloise Ontologies Associates (www.heloisenet.com), a Network of Ontology Experts from leading universities that offers a senior business consultancy on ontologies. She has degrees in philosophy, from the University of Utrecht, and educational sciences, from the Technical University of Enschede, the Netherlands. Contact her at ellen@heloisenet.com.

Wee Keong Ng is an assistant professor of the School of Computer Engineering at the Nanyang Technological University, Singapore. He obtained his MSc and PhD from the University of Michigan, Ann Arbor. He works and publishes in the areas of Web warehousing, information extraction, electronic commerce, and data mining. He has organized, chaired, and held tutorials at international workshops and has served on the program committees of various international conferences. He is a member of the ACM and IEEE Computer Society. Contact him at awkng@ntu.edu.sg; www.ntu.edu.sg/home/awkng.

Ee Peng Lim is an associate professor of the School of Computer Engineering at the Nanyang Technological University, Singapore. His research interests are in Web warehousing, database integration, and digital libraries. He is currently director of the Centre for Advanced Information Systems, a research center focusing on Web computing and database research. He is also the guest editor of the special issue in mobile commerce of the *Journal of Database Management*. His research has been published in a number of international journals including *IEEE Transactions on Knowledge and Data Engineering*, *ACM Transactions on Information Systems*, *Decision Support Systems*, and *Distributed and Parallel Databases*. He is also a member of ACM. Contact him at aseplim@ntu.edu.sg; www.cais.ntu.edu.sg:8000/~aseplim.

Guanghao Yan is a graduate student in the Department of Computer Science, the State University of New York at Stony Brook. He obtained his BSc from Peking University, China and his MSc from the Nanyang Technological University, Singapore. He works in data integration for electronic product descriptions. He is a student member of the ACM and IEEE Computer Society. Contact him at guanghao@cs.sunysb.edu; www.cs.sunysb.edu/~guanghao.

could easily be bypassed, buyers and vendors could share their production and selling forecasts, and therefore delivery cycles would drastically diminish.

Many business people are reluctant to accept these ideas. What about the careful (and convenient) relations and partnerships set up with clients and vendors? I would not want to argue that all existing agreements will dissolve, but in a large part of core business areas, such as strategic buying for production and R&D, the ease of finding the best prices and products over the Internet will at least shake up traditional business relationships.

Simultaneously, new types of relationships will arise—for example, in the area of collaborative planning, forecasting, and replenishment. An interesting example can be found at the Voluntary Interindustry Commerce Standards CPFR committee (see www.cpfr.org). This workgroup originated in 1998 as a subcommittee of VICS and set its mission to create collaborative relationships between buyers and sellers through comanaged processes and shared information. To collaborate, buyer and seller jointly develop a single plan and forecast for demands, promotions, and replenishment strategies. The VICS CPFR communication standards and supporting software are the means through which the partners communicate. Pilot projects of the CPFR committee yielded impressive results—such as an 80 percent increase in business with a trading partner—and significant sales growth occurs concurrently with inventory reductions. Recently, three of the world's largest industry-sponsored marketplaces announced adding CPFR services to their procurement hub.[1]

Another indication that marketplaces will highly impact business relationships is the emerging focus on peer-to-peer technologies, a general term referring to all technologies that will enable two or more actors from different companies to jointly derive some information (for example, a business plan, an order, a customer's request), that they jointly approve, plan, execute, and pay. An interesting question that arises is who will eventually own the e-marketplaces that enable this broad range of business processes, or will there be an owner at all?

Clearly, these potential opportunities and dangers make the business world eager to follow up on new trends and technologies, but that won't be easy. E-commerce trends are generally either described in a marketing brochure (with the right business and the wrong technical terminology) or a technical paper (where it is the other way around). At the same time, business and state-of-the-art technology were never so intertwined—a discrepancy we need to overcome before arriving at efficient e-commercing.

Business as usual, IT as usual?

E-commerce will not significantly alter business as usual if the IT community does not significantly change as well. If IT is to become more the heart of business instead of a support, its content should change. High-level IT jobs in the outside world—new functions such as Content Vice President, e-Director—will demand not only the latest technology support but also a strategic business vision. Technical universities that deliver students without an understanding of business processes will fail to respond to the interesting new idea that IT managers are fully part of the strategic management. The same applies to the research area—most papers discussing application areas in e-commerce hardly pass the introductory course in business terminology. E-commerce is about electronically doing business—that is, putting the business process flows of companies on Web-enabled technology. Any communication between science and businesses will be seriously harmed by the absence of a common ontology.

Let's examine a concrete example. E-commerce is supposed to generate enormous cost reductions. An often-cited example that appeared in many research reports in the first e-commerce days is that "Companies can reduce the costs of a purchase order up to 40 percent."[2] A strategic buyer in a manufacturing company would laugh at this. His job is to find raw materials that the production floor needs, continuously leveraging the best price with the right time frame. The costs of a purchase order mean peanuts to him. Why is this mistake made? The citation was taken out of its (business) context, which referred to the costs of a quantitative, highly utilized, and obviously poorly automated process of buying office supplies. Another more general example is that the exchange of business documents is a major success factor of e-commerce, yet most publications talk about posting a purchase order, sending or receiving a confirmation, and invoicing. The production floor manager will want to integrate his bill of materials, and the buyer will want to send a request for quote. A science that promises business process optimization should know where the pain is felt!

An interesting question that arises is who will eventually own the e-marketplaces that enable this broad range of business processes, or will there be an owner at all?

Of course, this does not imply that intelligent information specialists need know the ins and outs of different industries and processes, but they need a strategic vision. Interdisciplinary links were made among philosophy, psychology, and linguistics—it is now time for some business intelligence.

References

1. *NMM Weekly Market News*, vol. 4, no. 4, 22 Jan. 2000.
2. *AMR Research*, Sept. 1999, JP Morgan, 2000.

Standardization and Integration in Business-to-Business Electronic Commerce

Wee Keong Ng and Ee Peng Lim, Nanyang Technological University
Guanghao Yan, State University of New York at Stony Brook

Electronic commerce encompasses many issues, such as acquiring and storing information, finding and filtering information, securing information, auditing access, cost management and financial instruments, and so on.[1] Among these issues, *finding and filtering information* is of essential importance to a successful B2B electronic system where buyers need online facilities to help them retrieve information

and locate resources that match their expectations and desires. Specifically, buyers would like to find products and services at low costs, using languages and terminologies with which they are familiar.

However, the rich and diverse descriptions that vendors use to describe their products increase the difficulty of locating products and services accurately and efficiently. With the advance of current Web search technology, it is still difficult to manually locate a vendor for a certain product or to compare among different vendors, because most search engines are keyword-based. In fact, one of the biggest challenges for e-commerce today is to create mechanisms to let buyers locate products and services with specific characteristics and to let vendors locate potential buyers with specific traits.[1]

Product information heterogeneity is a critical impediment to efficient business information exchange. There is no uniform description for each product type among vendors. In e-commerce activities involving interactions among different vendors (B2B model) or between one buyer and multiple vendors (consumer-to-business model), a common ontology for the products is critical.

There are two general approaches to resolve the problem of information heterogeneity: standardization and integration. In the standardization approach, a common vocabulary and common protocol are drafted for all parties involved in a business exchange to unanimously support and adopt. This is a common approach seen in industry. In the integration approach, mappings are found between semantic components so that differences can be resolved. We observe that a proliferation of standards could bring us back to square one in the quest to resolve the information heterogeneity problem.

Standardization efforts in B2B e-commerce

In the B2B e-commerce industry, solutions for commerce activities are generally based on standards. As mentioned earlier, there are two prerequisites for efficient e-commerce information exchange: a common vocabulary and a common protocol. A protocol defines the rules of information exchange among parties engaged in an e-commerce activity. Much effort has been expended to provide related standards for these two factors (see www.dnb.com/unspsc, www.rosettanet.org, and www.unspsc.org for examples).[2] Ontology.org is an organization devoted to developing industry-specific XML document type definitions (DTDs) and thus to solving the vocabulary problem. The ICE (Information and Context Exchange) protocol[2] provides a solution for the protocol problem by managing and automating the establishment of syndication relationships, data transfer, and results analysis. In addition, the eCo Framework Project (www.commerce.net/projects/currentprojects/eco) by CommerceNet has also addressed some of the heterogeneity issues. It has created a base set of common terms and mappings among existing terms for e-commerce specifications. The eCo working group considers a list of related specifications among which are the RosettaNet Specification (www. rosettanet.org) and the Common Business Library (CBL).

We observe that a proliferation of standards could bring us back to square one in the quest to resolve the information heterogeneity problem.

Let's examine a typical standardization effort: the RosettaNet specification and the CBL. RosettaNet creates property definitions for various entities in e-commerce, such as property definitions for a certain product and its properties. For example, *modem* is a property (or an attribute) for computers. Once these property definitions are completed, they will be distributed to some standards maintenance organizations that will enumerate possible values for those properties. Then, property definitions as well as their values are distributed to companies in the industry supply chain as standards for business information format, say for product descriptions. Let's take a look at how RosettaNet defines one property of a product. We consider the definition of the *Central Processing Unit* property for a laptop given by the RosettaNet Laptop Technical Specification. There are several fields for the property *Central Processing Unit*: *Property Name*, *Synonym*, *Property Definition*, *Dictionary References*, *Where Used*, *Property Type*, and so on. Moreover, some of these fields contain subfields. For example, *Property Name* has *Abbreviation* and *Acronym* as subfields. All these fields serve as metadata for the product property (attribute).

The CBL by Veo Systems is a set of building blocks with common semantics and syntax to ensure interoperability among XML applications. CBL consists of information models for generic business concepts, including business description primitives such as companies, services, and products; business forms such as catalogs, purchase orders, and invoices; and standard measurements, date and time, locations, and classification codes. CBL consists of an extensible, public set of XML DTDs and modules. These building blocks can be assembled to create complete XML documents representing a business interaction such as a purchase order or an inventory stock query. Where possible, CBL takes advantage of other standards using, for example, relevant ISO standards for dates, currencies, and names. CBL is closely related to RosettaNet, and the property definitions that RosettaNet gives can be referenced by CBL to compose DTDs and modules for various electronic commerce transactions, including product descriptions. To use CBL, an organization starts by creating a CBL document describing its offers and services. Then, it integrates a CBL system with its back-end system by writing custom code that interprets information between the CBL format and the organization's previous format. It's like building a wrapper for back-end systems by using CBL blocks. After that, organizations interact on the basis of CBL semantics and syntax.

The integration approach in B2B e-commerce

As with most standards, it will be some time before e-commerce standards are widely used. It is expected that there will be a multiplicity of standards in the future, given the concurrent efforts among different organizations. Hence, it is conceivable that some form of integration would still be required for the various standards.

Different vendors may differ in the ways they describe their products—they might adopt different sets of attributes or vocabularies to describe the same product. For example, (*year*, *classification*, *singer*, *title*, *company*) may be a *schema*—a set of attributes and their corresponding domains—for

music CDs. We call such a vendor-specific schema a *local product schema*. A *global product schema* is a uniform interface for a product based on which heterogeneous product information can be exchanged correctly and efficiently. The interface functions like a common ontology for vendors of the same product. In general, the problem in any B2B e-commerce system is to derive a global uniform product interface from the different local schemas.

Product schema integration is essentially a process of building mappings among product attributes from different product descriptions. As in other schema integration problems (such as database schema integration), heterogeneity among local product schemas can be classified into two categories, namely *naming conflicts* and *missing attributes*. Naming conflicts include *synonyms*, words similar in meaning but different in spelling, and *homonyms*, words similar in spelling but different in meaning in different contexts. For example, *album* and *title* are synonyms in the local product schemas of music CDs. In addition, some product attributes used by one vendor may not be used by another. This results in missing product attributes. For example, some vendors may use *chassis* as an attribute to describe a PC while others may not.

It has been pointed out that Web data integration has to deal with a large and evolving number of Web sources with little metadata about the characteristics of the sources but a high degree of source autonomy.[3] Specifically, product schema integration has the following characteristics:

- *Limited knowledge of local schemas:* Because product information is proprietary, we may only obtain product schemas without further information about attribute domains or data types from the vendors' Web pages. Thus, conventional schema integration methods built on the availability of attribute domain information are no longer applicable. This presents additional difficulties in understanding the semantics of local product schemas.
- *Large number of local schemas:* The number of different vendors, even for the same product, can be large. In this situation, human intervention is hardly feasible. A low-cost, scalable, and fully automated solution is therefore required.
- *Fast local schema evolution:* Whenever new features are added or old features are removed, the local schema of that product must be updated. For example, newer versions of a multimedia PC product include additional peripherals that extend the local product schema. This gives rise to the problem of dynamic maintenance of the consistency and integrity of an integrated, global schema.

Product schema integration in the context of B2B e-commerce differs from the related problem of schema integration in database systems. Although there are many existing methodologies for schema integration in multidatabase systems, such as the use of knowledge bases, neural networks, or manual normalization before integration, they don't apply to product schema integration in lieu of the characteristics we mentioned. More importantly, the automation of product schema integration is an essential requirement. The large number of local schemas to be integrated and the frequent updates of product schemas make manual schema integration impossible. In short, we need a simple, scalable, and fully automated schema integration technique at the attribute name level.[4,5]

Remaining issues

Product description heterogeneity is an inherent problem in B2B e-commerce due to the autonomy of vendors in describing their product. Although there are concurrent efforts in the industry to standardize product descriptions, the degree of acceptance and the possible multiplicity of standards remains an issue impeding the progress of standardization itself. An alternative and complementary approach is to develop techniques for product schema integration. Although much work has been done in multidatabase schema integration, integration presents a somewhat different problem in the B2B e-commerce context due to special characteristics. With the widespread adoption of XML in the future, some form of integration would still be needed as vendors retain the freedom to define their description vocabulary. ❐

References

1. N. Adam et al., "Strategic Directions in Electronic Commerce and Digital Libraries: Towards a Digital Agora," *ACM Computing Surveys*, vol. 28, no. 4, Dec. 1996, pp. 818–835.
2. N. Webber et al., "The Information and Content Exchange (ICE) Protocol," World Wide Web Consortium Recommendation, Oct. 1998; www.w3.org/TR/NOTE-ice.
3. D. Florescu, A. Levy, and A. Mendelzon, "Database Techniques for the World-Wide Web: A Survey," *ACM Special Interest Group on Management of Data Record (SIGMOD Record'98)*, vol. 27, no. 3, Sept. 1998, pp. 59–74.
4. G. Yan, W.K. Ng, and E.P. Lim, "Incremental Maintenance of Product Schema in Electronic Commerce: A Synonym-Based Approach," *Proc. Second Int'l Conf. Information, Communications & Signal Processing (ICICS'99)*, 1999.
5. G. Yan, W.K. Ng, and E.P. Lim, "Product Schema Integration for Electronic Commerce: A Synonym Comparison Approach," to appear in *IEEE Trans. on Knowledge and Data Eng.*