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 (SIC) scheme in order 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 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 UNSPSC 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 UNSPSC for their own use. Some consortiums 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. Grass roots 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, one may need to combine taxonomies in order to generate an adequate taxonomy. One may want, for example, to extend the UNSPSC 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 branches from another. This forces a user either to merge ontologies, possibly using a tool like Chimaera\(^4\) or PROMPT\(^5\) for merging ontologies or use an approach such as the one advocated in DAML\(^6\) where users just 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 Research, 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 1024 × 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 may be stored, common diagonal values may 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.

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 may be one source of many useful tools for these efforts. Commerce itself will likely be another source of ontology tools.

**REFERENCES**


**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 co-editor of
the recently released DARPA Agent Markup Language. She is on the advisory board for Ontology.org and Powermarket, is on the executive council for the AAAI, and is on the executive steering board for the international organization for description logics, the Ontology Inference Layer effort. Contact her at dlm@ksl.stanford.edu.