Ontologies for Electronic Commerce

Deborah L. McGuinness
Associate Director and Senior Research Scientist Knowledge Systems Laboratory
Stanford University, Stanford, CA 94305
dlm@ksl.stanford.edu

Abstract

Electronic commerce is exploding - Forrester research projects 25 billion dollars in online spending by the year 2000. As the market segment grows, it has expanded into broader content areas. Broader domains increase the need for thoughtful content organization and browsing support. We promote the trend of using ontologies to support more than just search and also to to enhance browsing and more active "smart" notification services. In this paper, we identify some of the issues with respect to existing ontology-enhanced e-commerce applications, report and discuss findings from our own experiences building and using ontologies for web deployments in general and e-commerce specifically, identify some "low-hanging fruit" applications, and discuss some research directions

There are many examples of simple (and sometimes very large) knowledge trees and graphs being used to support web applications. Many early commercial search companies, with Yahoo! as a prominent example, realized the valued of broad and deep organizations of knowledge that could be used to help users navigate and search large spaces. Some, like Lycos, in addition to providing general organization schemes, also provided more specialized categorizations for corporate customers, thus allowing more specialization. Some took the term organization and searching to an art form, like Sageware, producing highly detailed and precise search queries associated with hierarchically organized terms in (typically) more narrow content areas. Some, like Alta Vista, influenced by the statistical results from the information retrieval community, avoided hand crafted organizations of terms and applied statistical methods in efforts to help generate meaningful clusters of related terms. Knowledge management companies as well, such as Intraspect, Dataware, and Correlate, integrate ontologies into their frameworks for organizing and finding documents. While approaches (such as semantic versus statistical) may vary, all have a similar goal of helping the user to find things aided by some kind of grouping of related terms.

We also had this goal in a line of knowledge-enhanced search work. We were motivated by problems in a number of domains, initially health and medicine and later expanding to electronic yellow pages, town calendars, competitive intelligence in advanced technologies, and later to finance and shopping. Our findings published on FindUR(McGuinness 1998) showed that prior to enhancing our sites with simple background knowledge ontologies, naturally occurring queries many times retrieved less than 10% of the relevant retrievals and sometimes none of the highly relevant retrievals. We used our ontologies to support standard query expansion of search terms with their subclasses and instances in the hierarchy and expanded on the Verity Topic Set functionality (Courtot 1997). We achieved substantial improvement where ontologies included query terms, rarely retrieving less than 80% of the highly relevant sites and typically producing lower relevance ranking on the extraneous retrievals. Of course there are limitations to this approach, and any work using query expansion, as the traditional information retrieval community has carefully articulated, but in restricted content domains (or domains where there are not too many senses of terms) with individual pages having short length and potentially non-controlled vocabulary terms, we have been able to consistently make significant improvements. In fact, we observed commercially non-viable retrieval performance without some sort of query enhancement. Further, we claim that e-commerce typically falls into the categorizations where we saw improvements. More academic experimets remain to be done both in order to precisely quantify results and limiting conditions and in order to determine how scaling to broader domains impacts results, but 15 deployments reflected the same trend.

In more recent applications, many web companies have focused on shopping. Yahoo! shopping exploits a very broad and deep ontology. Amazon as well has a very extensive ontology, that while focussed only on books, arguably covers most of the potential content areas on the web since books are written on all subjects. Newer players in the shopping area such as Inktomi's shopping engine have more shallow ontologies published(Inktomi 1999) however they are similarly broad. It is not clear yet if that is a result of a design decision based on beliefs concerning usability or performance or just a result of a goal of short time to market.

Clearly, a number of companies have decided that the expense associated with building and maintaining large ontologies is balanced by competitive advantages. Organizations such as CommerceNet and Ontology.org also support such trends. This could be the opportunity that many researchers seek out in their careers - a place to apply research findings in a fast moving world providing quick deployments across large populations. These applications can then provide the laboratories for our experiments on effectiveness of methods.

In recent applications, the author has extensively evaluated some existing large commercial ontologies and has built a broad and deep e-commerce ontology. One caveat should be noted - time pressures precluded the luxury of undertaking research when existing best practices proved inferior. These conditions however rule the online world and thus most of the emerging applications that will arguably set widespread user expectations will be developed under similar conditions.

We present a few observations and assumptions: (i) Results of online queries concerning shopping are likely to be short documents - typically an item or merchant description. (ii) Many times the pages have a semi-structured form and may be database generated. (iii) The descriptions on one site may not use a well-designed controlled vocabulary. (iv) Broad shopping services must function in heterogeneous environments and thus even if some sites are well controlled, services aggregating information from a number of services will need to understand multiple vocabularies.

We found value in the careful logic of knowledge representation (and description logics in particular) in our solutions. Our work used CLASSIC(Borgida et al 1989) as an embedded knowledge representation engine for identifying inconsistencies, detecting cycles in graphs, finding term usage problems, and identifying subclass relationships. The last few decades of knowledge representation work have provided a great basis for exploiting "Isa" hierarchies and fortunately, this does appear to provide payoffs in the e-commerce domain.

However as library retrieval systems have pointed out for years, related terms (by relationships other than sub-class) are also important. Knowledge representation has identified and researched some other relations such as "part-of" which can also be exploited. However, additional relationships become important as well. For example, "camera accessories" (not a subclass or part of a camera) should be closely associated with cameras. All of the existing deployed web ontologies we reviewed had this kind of notion embedded in their representations. We have begun attempting to articulate the relations of the most use for e-commerce.

Single parent versus multi-parent organizations of terms also appears to be interesting in e-commerce ontologies. It may be non-controversial that naturally arising ontologies would be multi-parented, and in fact, this is born out in most knowledge organizations such as Yahoo!'s. Issues arise however on related issues. For example, one might want to have unique node names and shorter presentation names for the nodes. If one goes down a path of food>cheese>cheese-gift-basket, the short name might be gift-basket, while if one goes down a path of gifts>gift-baskets>cheese-gift-baskets, the short name might be cheese. This can be solved with some extra infrastructure that maintains a short name for each path by which one can get to a node. Another way of solving it is exemplified in Amazon's approach where they have multiple occurrences of what would appear to be the same node. Their nodes sometimes have different names. This however is a challenge to humans or automatic programs that must map appropriate categories to products - the human or program must find all of the appropriate similar nodes without the support of a multi-parented ontology for guidance. It can also be a challenge to an ontology editor who might want tools that support identifying if the node that the editor is about to enter may already exist in another context (and may thus already have an entire substructure under it that would not need to be duplicated in the new location in the ontology).

After reviewing many large ontologies, we believe more strongly that it is difficult for humans to build large ontologies and maintain consistent naming and categorization schemes. We attempted to find adequate commercial tools to help minimize this problem and failed. We ended up using a combination of research description logic tools and ontology tools(Rice et. al. 1996) (with some pre- and post-processing extensions). Our claim however would be that these tools still need some research and development work before they would be commercially viable.

In summary, we find the area of e-commerce to be a researchers dream in terms of encompassing problems where AI has strong results and in terms of being an important and fast moving market. We believe it is a source of motivations that will extend the focus of knowledge representation and ontology research into areas that make our research more well grounded and usable and hopefully more able to have greater impact on the every day person's life.

Acknowledgements: We are indebted to a number of past and current colleagues including our FindUR colleagues - L. Resnick, T. Beattie, H. Manning, S. Solomon, P. Fuoss, D. Kormann, M. Plotnick and the Quintillion team, K. Hanson and the Worldnet team, H. Moore and the Personal Online Services team, L. Croel and the AT&T Labs west team, and R. Maulitz and the P-CHIP team. We are also indebted to the Stanford Knowledge Systems Ontolingua team, particularly J. Rice and R. Fikes, for expertise and ontology support tools and enhancements.

References

A. Borgida, R. Brachman, D. McGuinness, and L. Resnick. CLASSIC: A structural data model for objects. In *Proceedings SIGMOD-89*, 59–67. ACM, 1989.

P. Courtot. *A New Personal and Enterprise Application: SEARCH '97 White Paper, In Verity web site http://www.verity.com/corp/whitepapers/"*.

D. McGuinness. Ontological Issues for Knowledge-Enhanced Search. In *Proceedings of Formal Ontology in Information Systems*. Also in Frontiers in Artificial Intelligence and Applications, IOS-Press, Washington, DC, 1998.

http://support.inktomi.com/Shopping/Public/Docs/ .

J. Rice, A. Farquhar, P. Piernot, and T. Gruber. Using the Web Instead of a Window System. In *Proceedings of CHI*, Vancouver, B.C. Canada, 1996.