The Immediate Prospects for the Application of Ontologies in Digital Libraries

Jody L. DeRidder

Digital Library Center, James D. Hoskins Library, University of Tennessee, Knoxville, Tennessee. USA, <jo@utk.edu>

Jody L. DeRidder received her M.S. in Computer Science from the University of Tennessee in 2002, after developing repositories for the Open Archives Initiative in its alpha phases. As the lead developer for the Digital Library Center of the University of Tennessee Libraries, she has built, customized and altered software to create interoperable digital library systems which provide usability features beyond the norm. Nearing completion of her M.S. in Information Sciences, her research interests have turned to interoperability between systems to support usability, sustainability in digital libraries, and the application and use of ontologies via automated cross-mapping by query engines.

DeRidder, Jody L. **The Immediate Prospects for the Application of Ontologies in Digital Libraries.** *Knowledge Organization, 34(4)*, 227-246. 53 references.

ABSTRACT: The purpose, scope, usage, methodology, cross-mapping and encoding of ontologies is summarized. A snapshot of current research and development includes available tools, ontologies, and query engines, with their applications. Benefits, problems, and costs are discussed, and the feasibility and usefulness of ontologies is weighed with respect to potential and current digital library arenas. The author concludes that ontology application potentially has a huge impact within knowledge management, enterprise integration, e-commerce, and possibly education. Outside of heavily funded domains, feasibility depends on assessment of various evolving factors, including the current tools and systems, level of adoption in the field, time and expertise available, and cost barriers.

1. Introduction: defining ontology

Each of us has a slightly different way of looking at the world. Across cultures and research areas, these differences become palpable. What is clearly understood within a community may be unknown elsewhere and technically specific terminology needs to be translated, as if to a different language, for the general user. For applications to be able to serve us in search and retrieval across all these variations, human knowledge needs to be made comprehensible to computer programs. Building an ontology requires capturing concepts (including implicit ones), the relationships between them, and any constraints on those relationships (de Bruijn 2003, 35). In technical terms, an ontology represents a "language" of concepts, relations, instances and axioms (de Bruijn and Polleres 2004), which enable computer applications to logically reason out solutions or adapt queries. Stanford University offers a sample ontology application which suggests wine selections for your choice of food includes encoding examples and explanations (Hsu 2003). To illustrate an ontology description of an object, a graphic example of an ontology application to an audio tape of a performance of a single concerto (in the ABC ontology) is shown in figure 1 (Hunter, 2001).

1.1 Points to consider

For ontologies to be useful and feasible in digital libraries, several requirements must be met. First, there must be evidence that they are helpful to users. Usefulness must outweigh the cost and effort of creation and maintenance. Here we must consider further the identification of our user audiences, and the purpose and scope of what we wish to accomplish. Secondly, what is the state of the art? What parts of this territory have been mapped out, and what are still murky waters? Is there, or will there soon be, broad support for the use of cross-mapped ontologies? If the road is clear and support is avail-

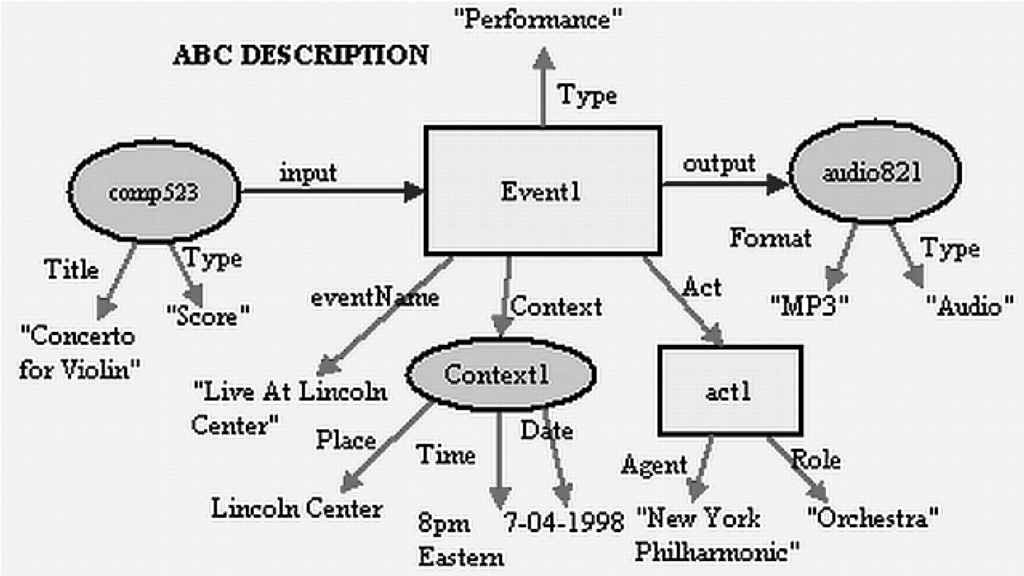

Figure 1.

able, it behooves us to make our digital libraries accessible via ontology mapping, to increase accessibility, interoperability, and to leverage the work in the broader arena to meet our constituents' needs. If it will be years before the path is paved, standards will likely change rapidly over that time. Those with the funding and the capability can lead the way, contributing to the development of standards and interoperability. If funding and capabilities are limited, it is wiser to wait till the paths are well-laid, and the process is easier. Thirdly, we need user-friendly tools and methodology. What are the steps? What personnel and tools are needed? As the field is still clearly in the beginning stages, an overview of current research and development is provided for further investigation. Finally, we must seriously consider the costs. What level of funds, personnel, and expertise are available?

1.2 Benefits

As systems grow in decentralized manners, semantic heterogeneity is inevitable; how do we provide functional search and retrieval across distributed digital libraries? Searching by keyword retrieves irrelevant information when a term has multiple meanings; and information is missed when multiple terms have the same meaning. In addition, concepts that may not be represented by the terminology in the document or metadata are not available to searchers. Information retrieval is a negotiation process, and as digital content multiplies, users need assistance in wading through the results of their searches. A comparison of precision and recall between full text searching, latent semantic indexing, and ontology-based retrieval (with manual assignment of concepts to query) finds ontologies capable of providing far better retrieval efficiency (Paralič and Kostial 2003).

Digital libraries routinely provide their services without human assistance; thus it is essential that their metadata be suitable for computation, supporting inference. The reference interview is not available; therefore, computer applications need to be able to reason about their contents to reformulate queries, deduce relations between works, and customize services to the task and user. This is only possible via ontologies (Weinstein and Birmingham 1998).

Imagine a user entering a query, and the computer application offers different meanings for the entered terms; the user selects the intended meaning, or chooses one of the related terms offered. The query engine transforms the query into a language that matches the terminology used in describing the data sources. In addition, it locates material related to your query, based on logical deduction and inference, offering these results on the side. In this manner, relevance and pertinence are improved, and browsing is enabled. With ontologies, we enable computer applications to perform intelligent searching instead of keyword matching, query answering instead of information retrieval, and to provide customized views of materials. A standardized vocabulary referring to natural language semantics enables automatic and human agents to share information and interoperate functionally (Fensel et al.2003c).

1.3 Depth and breadth

There are many different ways to classify ontologies; two of the most useful reflect the depth and the breadth of the ontology. In the **depth** dimension, the specificity of the ontology determines its "weight." Lightweight ontologies are little more than taxonomies, and include only concepts and their properties, relationships between those concepts, and controlled vocabularies. Heavyweight ontologies also include axioms and constraints that increase the capability of a computer application to logically reason with the data given. Dublin Core might be considered an extremely light-weight ontology, whereas Cyc (created using the Knowledge Interchange Format, a proposed standard) may be the most extensive top-level ontology currently in existence (de Bruijn 2003, 6-9). (Two limited open-source versions of this encyclopedic ontology are available: *OpenCyc* and *Research Cyc*.) In the **breadth** dimension, there are general (top-level, or global) ontologies, domain ontologies (specific to a particular area) and application ontologies, which describe concepts depending on the task as well as the domain (some refer to application ontologies as another form of domain).

1.4. Cross-mapping issues

In order to provide searching via natural vocabulary, a mapping is needed from the natural language of each user group to the entries in each metadata vocabulary. This is known as an "entry vocabulary index" or EVI. In addition, to search across databases, it is necessary to have mappings between each possible pair of system vocabularies, or ontologies. Mapping between ontologies must be done by people competent in both domains; the current status is that human assistance in mapping will likely be necessary for some time to come, for high quality mappings (Bockting 2005).

Problems in cross-mappings can be of several types. Data objects of the same name may describe different real-world elements; concepts may be ascribed to different levels of the metadata structures (an attribute in one ontology may be a class in another); conceptual approaches may preclude a functional correspondence; descriptions of a single real-world element may vary considerably and conflict with one another; and one of the ontologies may have incorrect information (Adam, Atluri, and Adiwijaya 2000). A concept in one ontology may not exist in another, or may have an entirely different meaning. For example, in the Harmony Project, members of the closely-related domains of digital libraries and cultural heritage and museum communities sought to merge the digital library ABC Ontology (Lagoze and Hunter 2001), with the CIDOC (International Committee for Documentation of the International Council of Museums) Conceptual Reference Model (ICOM/CIDOC and CIDOC CRM 2005). They uncovered cultural biases particularly in terms of the nature of change; while both ontologies were concerned with change over time, one modeled the change of objects, while the other modeled changes in the context and meaning for those objects (Doerr, Hunter, and Lagoze 2003). A comprehensive overview of the problem areas of mapping, including variation of expressiveness and the differing modeling paradigms or styles, is discussed by Klein, and diagrammed in figure 2 (Klein, 2001).

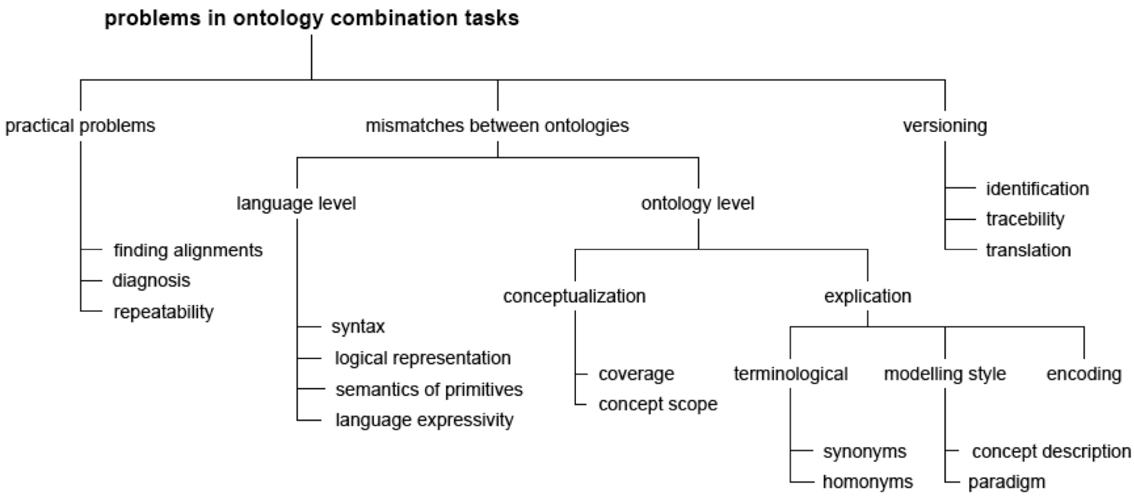

Figure 2.

An IMLS-funded effort (National Leadership Grant No. 178), based on prior research partially supported by a DARPA (Defense Advanced Research Projects Agency) contract, explored the feasibility of cross-mapping vocabularies of numeric data sets and text files (Buckland et al. 2007). It was discovered that the vocabularies for topical categorization vary greatly, requiring interpretive mappings between systems, and that specification of geographical area and time period are problematic. Both names of places and of time periods are culturally based, unstable, and ambiguous. The use of geospatial coordinates is suggested as the only effective method of relating locations to search terms, which means that both gazetteers and map visualizations become critical to implement search retrieval in a user-friendly manner. A similar application needs to be developed for time periods, and this issue is being addressed in a subsequent IMLS-funded study by the Electronic Cultural Atlas Initiative (Electronic Cultural Atlas Initiative 2006). Among other objectives, the intent is to contextualize objects in library and museum collections by using or adapting existing and emerging standards and protocols. This initiative is described further in (Petras et al. 2006).

Ontologies must be expected to evolve over time as knowledge and understanding grow, and terminology changes. Their mappings to other ontologies must also evolve, and this evolution may require change in other ontologies to which they are mapped (de Bruijn and Polleres 2004, 11). Thus the initial effort to develop ontologies is insufficient; they must not only be maintained but also versioned over time, and compatibility with other ontologies considered with each evolution. Cross-mappings are rare, expensive, time-consuming, and difficult to maintain. With 135 semantic types and 54 relationships, the Unified Medical Language System Metathesaurus is a notable example (Smith et al. 2004).

1.5. A bird's eye view

It is insufficient to consider ontology mapping as a singular or only a local problem. Many differing ontologies already exist with overlapping domains of knowledge and application (de Bruijn 2003). And there are at least three basic conceptual approaches to interoperability: a global ontology to which all local ontologies are mapped, a peer-to-peer system (where mappings exist between local ontologies where needed), and a combination of the two. A central, heavyweight global ontology is clearly preferable for computer applications, as one-to-one mappings of all involved ontologies does not scale. However, obtaining global agreement on controlled terms and relationships is infeasible, so a layering approach based on generality is more likely to succeed, with mapping between domains and higher level ontologies as needed (Meersman 1999). A single general light-weight ontology to be shared by multiple domains was explored by (Stuckenschmidt and van Harmelen 2005). After developing their framework, the authors stated that the shared ontology can only be developed if all sources of information are known, and the conceptualization of each source is accessible; they concluded this was only feasible for a single domain (Stuckenschmidt and van Harmelen 2005, 249). De Bruijn and Polleres add that a limitation to this approach would be the likely lack of agreement on the interpretation of the concepts in the shared ontology by all the authors of local ontologies (de Bruijn and Polleres 2004).

Another possible middle ground between the peer-to-peer approach and the central core ontology method, would be to implement layers or a hierarchical application (de Bruijn and Polleres 2004). One way to envision this is to compare a scientific discipline with a group of islands, where each area of research is an island, and each island has a further breakdown of specificity into "dialects." If a single island had 3 dialects, each dialect would be a Level 1 ontology, probably the most specific in terminology. A shared ontology for the entire island would be a Level 2 ontology. Islands (or domains) could map to one another as needed. A shared ontology for the group of islands would be a Level 3 ontology, the most general so far. Other sets of islands could have similar structure, and again, the hierarchy could continue as needed, but with a distributed, organically growing base rather than a single top-down application. This may be the only feasible solution, as it reflects the grassroots approach and grows as needed.

2. State of the art

Currently, the semantic search engine *Swoogle* states that there are at least 10,000 ontologies in use on the WWW, and provides a list of ontology repositories, semantic web search engines and crawlers. The 2005 version of *Swoogle* indexed 337,182 documents, while the 2006 version currently lists their number of documents at 2,030,039 (*Swoogle* 2006), a major increase. This cursory comparison indicates a growing interest in the implementation of ontologies.

2.1 Within domains

Ontologies seem to have already found a home in instructional technology, as an outgrowth of KOS (Knowledge Organization Systems). The primary difference is that ontologies apply logic to the relations (Binding and Tudhope 2004). Other differences are that existing KOS lack conceptual abstractions, semantic coverage, consistency, and automatable processing (Soergel et al. 2004). Ontologies are important to education because concepts and the relationships between them "provide a powerful, and perhaps the only, level of granularity with which to support effective access and learning" (Smith et al., 2004, 2). A portal already exists for sharing tools, projects, research and information for ontology use in education (Dicheva et al. 2006), and a commercial success in the education arena is *Xyleme*, which depends upon the existing heterogenous XML structure in documents for pattern-matching, mapping, encoding, and creating "views" for abstract query response (Aquilera et al. 2000).

The Alexandria Digital Earth Protoype (ADEPT), currently in use for teaching geography courses at the University of California, employs an ontology to link the current lecture material to a graph showing its relation to other concepts, and also links to examples from the digital library. All three views are presented at the same time, to give students the context and examples they need to make sense of what the teacher is trying to communicate. In addition, the ontology supports a Virtual Learning Environment that lets the teacher create, use, and re-use learning materials in different fields of science and in various learning environments (Smith et al. 2004).

Yet here the content of the digital library itself is limited to examples, primarily images and graphs. For digital libraries containing complex materials, there exists the need for two levels of access: discovery of resources, and discovery within the resources, the latter of which requires the creation of descriptions of semantic and internal structural organization through resource decomposition. The GREEN digital library project explored the problems and possibilities in this area, using term extraction algorithms, performing text analysis, and extending a combination of metadata schemes (LOM for learning objects and MatML for materials). This group noted the need for a convergence of metadata schemes and robust mechanisms for navigating a complex associational web of resources (Shreve and Zeng 2003). Clearly, the ability to locate specific content, regardless of its location within materials, would be extremely useful for isolating information and minimizing the time spent sifting through search results. As the quantity of materials online explodes, findability becomes critical.

An example of ontology use in enterprise integration would be the Unified Medical Language System (UMLS), which provides services for computer applications across a multitude of health-industry areas. The UMLS Metathesaurus is a compendium and synthesis of more than 100 different thesauri, classifications and code sets for health care, billing, statistics, medical literature, research and resources, and requires constant updating and renovation. The Metathesaurus preserves the many views present in the source vocabularies, as each may be useful for different tasks. Hence, it must be customized to be effective in any one application (U.S. National Library of Medicine, March 2006a). UMLS includes a Semantic Network to "provide a consistent categorization of all concepts represented in the UMLS Metathesaurus and to provide a set of useful relationships between these concepts" (U.S. National Library of Medicine March 2006b). In addition, the SPECIALIST Lexicon provides a general English vocabulary that includes biomedical terms, for Natural Language Processing (NLP), to improve searchability for the general user (U.S. National Library of Medicine, March 2006c).

E-Commerce potential is clearly indicated in the level to which ontologies have already proven their value in critical government defense, finance, and manufacturing. An example in the business arena is Australia's InfoMaster. In the United States, Ontology Works, founded in 1998 by former members of the intelligence community, currently serves the critical needs of such clients as the U.S. Department of Defense, the U.S. Department of Justice, Science Applications International Corporation, Boeing, Northrop Grumman, and the Sierra Nevada Corporation. Ontology Works is a highly successful commercial venture, and claims to have the most sophisticated ontology-driven database on the market (Ontology Works, 2005). Another commercial success is Ontobroker, a deductive, object-oriented database system, now available via Ontoprise.

MOMIS (Mediator environment for Multiple Information Sources) has been used to model a tourism information provider system. In the MOMIS Integration Methodology, local source schemata are extracted. If the source material is unstructured, text is extracted, analyzed, and an XML schema is generated. Then a meaning for each element of the source schema is chosen from a lexical database of English, WordNet (prompts for choices are given to a human; the choice is manual). A common thesaurus, a global schema, and sets of mappings to local schemata are generated. Finally, a meaning is assigned (semi-automatically) to each element of the global schema. The query manager then rewrites the incoming global query as an equivalent set of queries to match the local source schemata; local sources are queried with these, and the resulting responses are fused and reconciled into a final response (Bergamaschi et al. 2005).

Exploration has been made into non-textual content as well. Annotation of historical images with a domain-specific ontology enables users to retrieve images for which they inadequate historical knowledge and keywords (Soo et al. 2002). An Amsterdam research group has developed a Visual Ontology Using MPEG-7 and WordNet, which supports descriptions of colors and shapes of objects, to support automatic annotation (Hollink et al. 2005). By extracting and analyzing visual features, mapping clusters of sequences and patterns to ontological concepts, another experiment has demonstrated the feasibility of semi-automated ontology annotation of domain-specific videos (Bertini, et al. 2005). In a fourth model, audio tapes of sports broadcasts were annotated (Khan, McLeod, and Hovy 2003), though the text analyzed was extracted from the closed captions that came with the audio objects. In this project, only three relations were modeled (isA, Instance-Of, and Part-Of), and an automatic query expansion mechanism was built using WordNet as a generic ontology, though they found it too incomplete to functionally model the domain.

According to (Ontology Works 2005), the leading research groups in ontologies are IFOMIS (The Institute for Formal Ontology and Medical Information Science), ECOR (European Centre for Ontological Research), LOA (Laboratory for Applied Ontology), and NCOR (National Center for Ontological Research). Based on the number of recent ontological projects, Stanford University's Knowledge Systems, Artificial Intelligence Laboratory and the Sirma Group's OntoText Semantic Technology Lab should perhaps be added to this list.

2.2. Across domains

One of the primary purposes of cross-mapping is to allow searching of heterogenous resources from a single interface. The Digital Government Research Center Energy Data Collection project used an overarching ontology (SENSUS) to provide searching across over 50,000 database tables, manually defining the domain model with 500 concept nodes, then mapping them with intentionally vague semantic meaning to the possible 70,000 nodes of the larger ontology. While much of the model building was automated, it was far from simple to create a coherent domain model out of the variation of metadata and domain terms within the databases. The end product cannot support automated inference, but does enable browsing and non-expert searching with familiar terms (Hovy, 2003).

OntoMedia, an opensource effort, builds on the CIDOC Conceptual Reference Model and the IFLA-NET (International Federation of Library Associations and Institutions) FRBR model (Functional Requirements for Bibliographic Records) to facilitate the annotation of semantic content of multimedia. It provides the user with a graphical user interface with metadata indexing and search capabilities, for organizing multimedia collections, though the ontology is presented as a general, high-level ontology for reuse across domains (Lawrence et al. 2005).

Semantic Interoperability of Metadata and Information in unLike Environments (SIMILE) is a joint project of MIT Libraries and MIT Computer Science and Artificial Intelligence Laboratory, which leverages and extends DSpace. The intent is to enhance general interoperability across distributed information stores of varying types, and to provide useful end-user services for mining that material (Leuf 2006, 223-4). In an early prototype of the project, VRA Core (Visual Resources Association Data Standards Committee) and IMS LOM (Learning Object Metadata) were translated into RDF schemas with enrichment obtained from Wikipedia and the prototype OCLC Library of Congress Name Authority Service. Then the datasets were transformed from XML to RDF/ XML using XSLT. While the developers were able to automate linkage of RDF datasets using string similarity techniques, the approach was error prone and results had to be manually reviewed. In addition, the enrichment techniques could be automated as well, but again, required human intervention to verify the validity of the data produced (Butler et al. 2004).

3. Fundamentals

3.1. Methodology

A recent analysis of the state of ontology engineering bemoans a lack of guidance, unified methodology, cost benefit analysis tools, and selection support to choose engineering approaches (Simperl and Tempich 2006). Real-world applications require comprehension of the scope and progression of the project, customizable workflows, user-friendly tools, and automation of the majority of the tasks. While several ontology management tools are relatively mature, many necessary ontology engineering activities are not yet adequately supported by technology, and critical aspects, such as automation of ontology creation, application, and mapping are still being researched. The basic model for implementation of an ontology (without consideration of ontology mapping: see figure 3) includes a feasibility study, domain analysis, conceptualization, encoding, maintenance and use (Simperl and Tempich 2006).

One unusual investigation tested the hypothesis that the more indexing is geared toward the user task, the better the results. Kabel, Hoog, Wielinga and Anjewierden (Kabel et al. 2004) compared the efficiency, effectiveness, precision of use, and quality of results when users were given access to keywords versus a domain index versus an instructional index, for creating lesson plans. The domain index was content-based, with specific terminology. The instructional index provided classification of objects by use in instructional material, and hence was task-oriented (an application ontology). An example of this would be a "behavioral description" with "specific" scope, and the instructional role of "illustration." Their hypothesis was generally correct. The domain index provided more efficient, effective

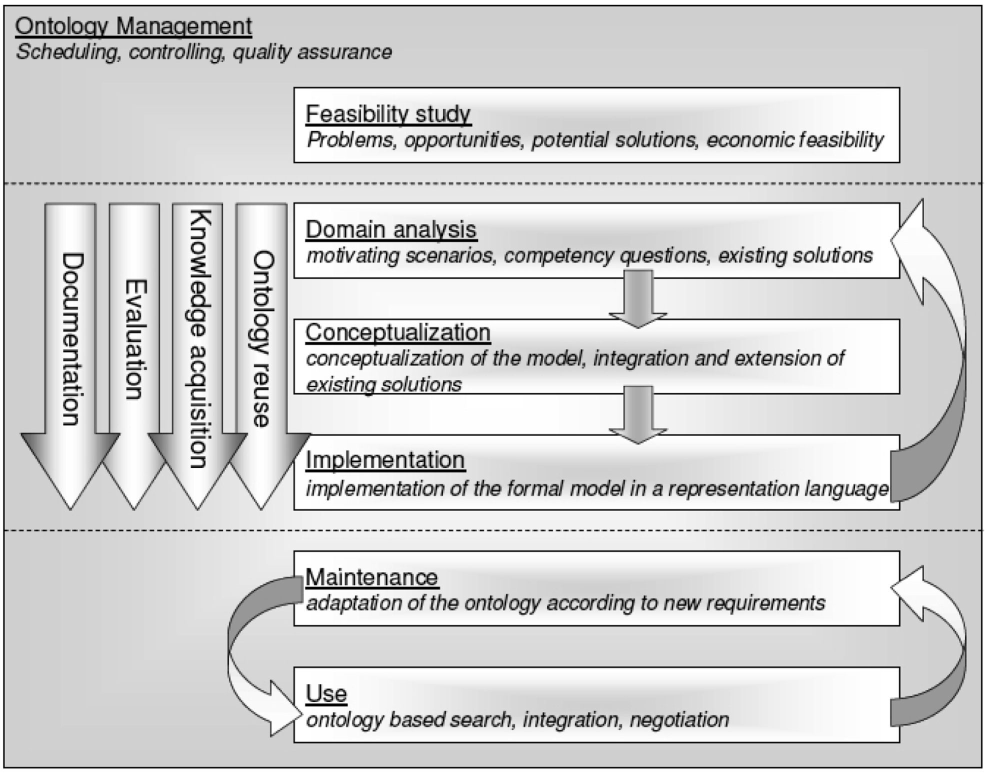

Figure 3. Ontology Engineering Activities

search and retrieval than the keyword search, and the instructional index provided better precision than the use of keywords and domain indexing. Hence, it appears that we need to clearly understand the needs of our users, in order to choose the type of ontology that will actually provide the specificity they need for the task at hand.

SchholOnto (Shum et al. 2000), for example, is an effort to develop an ontology for discourse about research, rather than for the research itself, which is an

3.2. Purpose and scope

Before choosing, adapting, or creating an ontology, the purpose and the user audience must be determined. If the domain is clearly delineated and there is no desire for interoperability or cross-mapping to outside ontologies, the scope and direction are simplified. If, however, the desired outcome is more diverse and interoperable, the choices made in this assessment will be both critical and complex.

interesting twist. Designed to provide an ontology for scholars to interpret, discuss, analyze and debate about existing literature, ScholOnto (developed using OCML (Operational Conceptual Modeling Language) overlays existing metadata and does not attempt to directly describe the content of the research. Instead, the ontology provides a structure to clarify the intellectual lineage of ideas, their impact, scholarly perspectives on those ideas, inconsistencies in approaches or claims, and convergences of different streams of research (Shum et al. 2000, 3). Here, the comments about the literature become the objects for retrieval and for building new structures to define the usefulness of the object. This is a social networking function, an interactive community-created layer over the research itself. This could be an invaluable way to add context and clarity to understanding and exploration of a domain. Thus, the application of ontologies to digital libraries might not be in querying the documents themselves, but in building relationships and connections and social context around the documents.

3.3. Conceptualization

If ontologies exist that can be adapted to the purpose at hand, tools are needed to perform such adaptation. If an appropriate ontology does not yet exist, tools are needed for modeling and constructing the ontology. Selecting or creating an ontology involves a fundamental tradeoff between the degree of complexity and generality versus the degree of efficiency of interpretation and reasoning within the language (Weinstein and Birmingham 1998, 35). Maximum consideration must be given to the desired services. The following findings are intended to provide a starting point for further exploration.

One possibility is that of creating ontologies out of existing metadata schemes or thesauri, adapting and adding as needed. The more complex and structurally coherent the metadata scheme, the more feasible this may be. One effort under development is an adaptation of the AGROVOC Thesaurus, developed and maintained by the Food and Agriculture Organization of the United Nations (Soergel et al. 2004). An older effort to transform MARC (MAchine Readable Cataloging) uncovered difficulties in the varying dimensions and multiple levels of granularity containing partial descriptions, which is a requisite feature of bibliographic data (Weinstein and Birmingham 1998). Another possibility is creating an ontology from scratch, using existing models to pave the way. OCML (Operational Conceptual Modeling Language) supports the construction of ontologies and problem solving methods, and is supported by a large library of reusable models (via the WebOnto editor). Currently in use by several projects, OCML is available free of charge for non commercial use.

Building on previous work is a third option, and the one which offers the greatest variety of tools at present. Many of these are domain-specific.

The ABC Metadata Model Constructor fundamental classes for digital libraries were determined by analyzing commonalities between Dublin Core, INDECS (Interoperability of Data in e-Commerce Systems), MPEG-7 (Multimedia Content Description Interface), CIDOC (International Committee for Documentation of the International Council of Museums) Conceptual Reference Model and the IFLANET (International Federation of Library Associations and Institutions) FRBR model (Functional Requirements for Bibliographic Records). These classes form building blocks for developing either application or domain-specific ontologies, with event-aware views for modeling different manifestations of a relationship (Hunter 2001). This tool provides graphical user interfaces and is free to download, but it is still an experimental prototype (Leuf 2006, 217-8), without support, and assumes users understand Java, RDF, and basic ontology and metadata principles.

WebOnto is a freely available Java applet coupled with a customized web server (LispWeb), which provides browsing, visualization and editing of knowledge models via the web. WebOnto is currently being used with ScholOnto (discussed above) and PlanetOnto, for search, retrieval, news feeds, alerts, and presentations of laboratory-related information.

The Kraft project outlines steps to building shared ontologies: ontology scoping, domain analysis, ontology formulation, and top-level ontology (Jones et al. 1998). However, their methodology lacks comprehensive evaluation of ontologies and is not applicable to global domains (Stuckenschmidt 2005, 68).

The Protégé opensource Ontology Editor provides two main ways of modeling ontologies, and can export in various formats including OWL. Used extensively in clinical medicine and the biomedical sciences, Protégé covers the full range of development processes (Leuf 2006, 209-210).

KAON (KArlsruhe ONtology) offers a stable opensource, comprehensive tool suite for ontology creation, management, and a framework for building applications; it was designed for business applications requiring scalability and efficient reasoning capabilities (Leuf 2006, 213).

Chimæra is a system for creating and maintaining distributed web ontologies, as well as for merging ontologies and providing multidimensional diagnoses to identify problems (Leuf 2006, 210-211). Chimæra can load and export files in OWL, and is available opensource.

There are many possible variations in the ability of software to combine and relate ontologies; Klein provides a comparison of several different approaches (see table 1): SKC (Scalable Knowledge Composition), Chimæra, PROMPT, SHOE (Simple HTML Ontology Extensions), OntoMorph, metamodel, OKBC (Open Knowledge Base Connectivity) and layering. Of these, OntoMorph addresses the majority of the stated problems in combining ontologies. However, Klein states that "mismatches in expressiveness between languages is not solvable" and more comprehensive schemes need to be developed for interoperability of ontologies (Klein 2001).

3.4 Encoding

For computer applications to be able to use ontologies, they must be encoded in machine-readable languages: in particular, all implicit relations between concepts must be explicitly encoded. To enable interoperability between ontologies and query engines, we need to agree on standards for these encodings. As in any other area, there is some disagreement on what is the most useful path. OntologyWorks used the draft ISO (International Organization for Standardization) standard, SCL (Simple Common Logic), which has been superceded by the Common Logic Standard, currently under development (ISO 2006). Since OntologyWorks does not seek interoperability with the broader public (it is a commercial effort), their focus was on what was most efficient and effective for their needs. However, if this standard is adopted by the ISO, it will likely compete with OWL for wider ontology development. CyCorp developed its own language, CycL, for their powerful Cyc system; however, their opensource components (OpenCyc and ResearchCyc) provide translators to certain other languages, and the ability to export selectively in OWL (CyCorp 2002). Schematron, "a language for making assertions about patterns found in xml documents," is based on the tree pattern uncovered in the marked-up document. It allows you to determine which variant of a language you are working with, as well as to verify that it conforms to a particular schema (Leuf 2006, 218). Schematron was published as a draft ISO standard in 2004.

Issues		SKC	Chim.	PROMPT	SHOE	OntoM.	Metamodel	OKBC	Layering
Language level mismatches	Syntax					M	M	M	M
	Representation					M	M	M	M
	Semantics					M	M		M
	Expressivity								
Ontology level mismatches	Paradigm					M			
	Concept description					M			
	Coverage of model								
	Scope of concepts	M	U	U	M	M			
	Synonyms	M	U	U	M	M			
	Homonyms	M				U			
	Encoding	M			M	M			
Practical problems	Finding alignments	U	U	U					
	Diagnosis of results		A	A		A			
	Repeatability	A		A		A			
Ontology versioning	Identification				M				
	Change tracking				M				
	Translation								

Table1. Table of problems and approaches for combined use of ontologies

Legend A: Solves problem automatically
U: Solutions suggested to user
M: Provides mechanism for specifying solution (Klein 2001)

A proposed RDF Thesaurus Specification provides "conceptual relationships for encoding thesauri, classification systems and organized metadata", as well as a proposal for encoding a core set of thesaurus relationships (Cross et al. 2003). The two standards that have been adopted by the World Wide Web Consortium are the Resource Description Framework (RDF) and the Web Ontology Language (OWL). RDF is a simple notation for representing relationships between and among objects. RDF uses URIs (Uniform Resource Identifiers) for identification, and describes resources in terms of three parts: subject, predicate (the type of property about the subject), and object (the value of the property about the subject) (World Wide Web Consortium, 2004b). OWL, the Web Ontology Language, was developed for defining and instantiating web ontologies so that computers can logically interpret information. An extension of RDF, OWL has 3 increasingly complex sublanguages:

– OWL Lite is the simplest, and most closely related to thesauri.
– OWL DL is based on description logics, which enable computer applications to reason logically and make inferences.
– OWL Full is provides maximum expression with no computational guarantees.

OWL Full will probably never have wide usage due to its lack of tractability and lack of logic support; practical applications will likely use some subset of OWL DL, as it can provide both power and functionality. (de Bruijn 2003, 74).

3.5 Tools

Much of the research in the cross-mapping arena is focused on identifying and seeking solutions to the problems, rather than developing tools. However, XeOML offers an extensible markup language for mapping ontologies against one another, two at a time. Simple mappings are one-to-one relations, and complex mappings may involve more than one element or element type in either or both languages (Pazienza 2004).

MetaNet is a metadata term thesaurus created by the Harmony project to provide additional semantic knowledge that does not exist in XML-encoded metadata descriptions. Since many entities and relationships occur across all domains, it is possible to generate a simplified set of semantic relationships between metadata terms in domain schemas to the preferred terms in the ABC ontology, and then (based on this relationship), generate semantic relationships (cross-domain) between each of those original metadata terms, outputting the results in RDF (Hunter, 2001). In addition, Harmony offers the ABC Metadata Model Constructor for use with their ABC ontology, an RDF visualization tool for complex metadata (RDFViz), and a simple RDF query language (Rudolf "Squish").(Brickley et al., 2002b)

The SIMILE project (Semantic Interoperability of Metadata and Information in unLike Environments) assessed existing tools in 2003, including RDF editors (IsaViz and RDFAuthor), schema editors (Protégé-2000, KAON OI-Modeller, and Ontolingua), ontology visualization software (OntoRama and Ontosaurus), application profile editors (SCART: The MEG Registry Client), metadata instance editors (Haystack, Standardized Hyper Adaptable Metadata Editor, and Simple Instance Creator), XForms for combining XML and forms, and thesaurus construction software (WebChoir vocabulary tools, Thesaurus Builder, MultiTes, and Term Tree) (Gilbert and Butler 2003). They determined that the existing tools only assist users in formally capturing existing models, rather than helping them to model their own schema. In addition, they found no formal approach for creating RDF models, so they proceeded to fill the gaps. Some of the tools they created include: a faceted browser for RDF browsing via standard web browsers (Longwell), an interactive graphical RDF visualization browser (Welkin), a tool for converting existing syntaxes into RDF (RDFizers), a tool that summarizes the structure of an XML dataset (Gadget), and a generic ontology for rendering RDF in a human-friendly manner (Fresnel, still in development) (Mazzocchi, Garland and Lee 2005).

University of Maryland's Mindswap Lab has developed an open-source OWL-DL reasoner, Pellet, for which commercial-level support is available. The InfoSleuth project is working to develop a commercial query server that dynamically adapts to the available information sources and services, fusing related information from heterogenous resources and abstracting results to the level appropriate to the user needs (Telcordia Technologies 2005). Query engines can currently be classified coarsely by whether they use a centralized ontology to which all others are mapped, or whether they support individual mappings between ontologies. TSIMMIS, InfoMaster, MOMIS, and Xyleme (an industrial solution) are based on a framework in which a single central schema is mapped to local schemas (Pazienza et al., 2004). The Bremen University Semantic Translator for Enhanced Retrieval (BUSTER) is a middleware of this same type, designed to access and integrate multiple ontologies which are based on a common vocabulary. The general top-level ontology it uses is based on simple Dublin Core with some added refinements. (Visser and Schuster 2002). Thus the user must commit to the basic generalized vocabulary that is used to define concepts in all the source ontologies, and is not presented with a specific domain view (Stuckenschmidt 2005, 199-207).

In contrast, the OBSERVER (Ontology Based System Enhanced With Relationships for Vocabulary hEterogeneity Resolution) system requires the user to select his terms from one of the ontologies it supports; the source material that ontology covers is then queried (Figure 4). If the results are not satisfactory, the user query is rewritten into the ontologies of other information sources in order to query other holdings (Mena et al. 2000).

OBSERVER uses synonyms, hypernyms, hyponyms, overlap, disjointedness and coverage to map between ontologies, storing these relations in a central repository to use for translating queries Stuckenschmidt, 2005, 192-198). In this manner, heterogenous databases and ontologies are managed without the need for a single global ontology (Mena et al. 2000). MAFRA (the Ontology MApping FRAmework) also is based on distributed mediation systems rather than a centralized one (Pazienza et al. 2004).

3.6 Costs

While ontologies offer benefits in terms of interoperability, browsing and searching, reuse, and structuring knowledge in a domain, the costs must be considered. Costs include construction, learning, cross-mapping, and maintenance and continual development of both the ontologies and the software (Menzies 1997). Information about cost is difficult to obtain, as most efforts are prototypes or commercial developments. Tim Berners-Lee, a major proponent of the Semantic Web, downplays the total cost, and fails to consider methodologies, depth of ontology, or even level of usability in his online assessment (Berners-Lee 2005). In a later article with others, however, this stance is modified somewhat by implying that

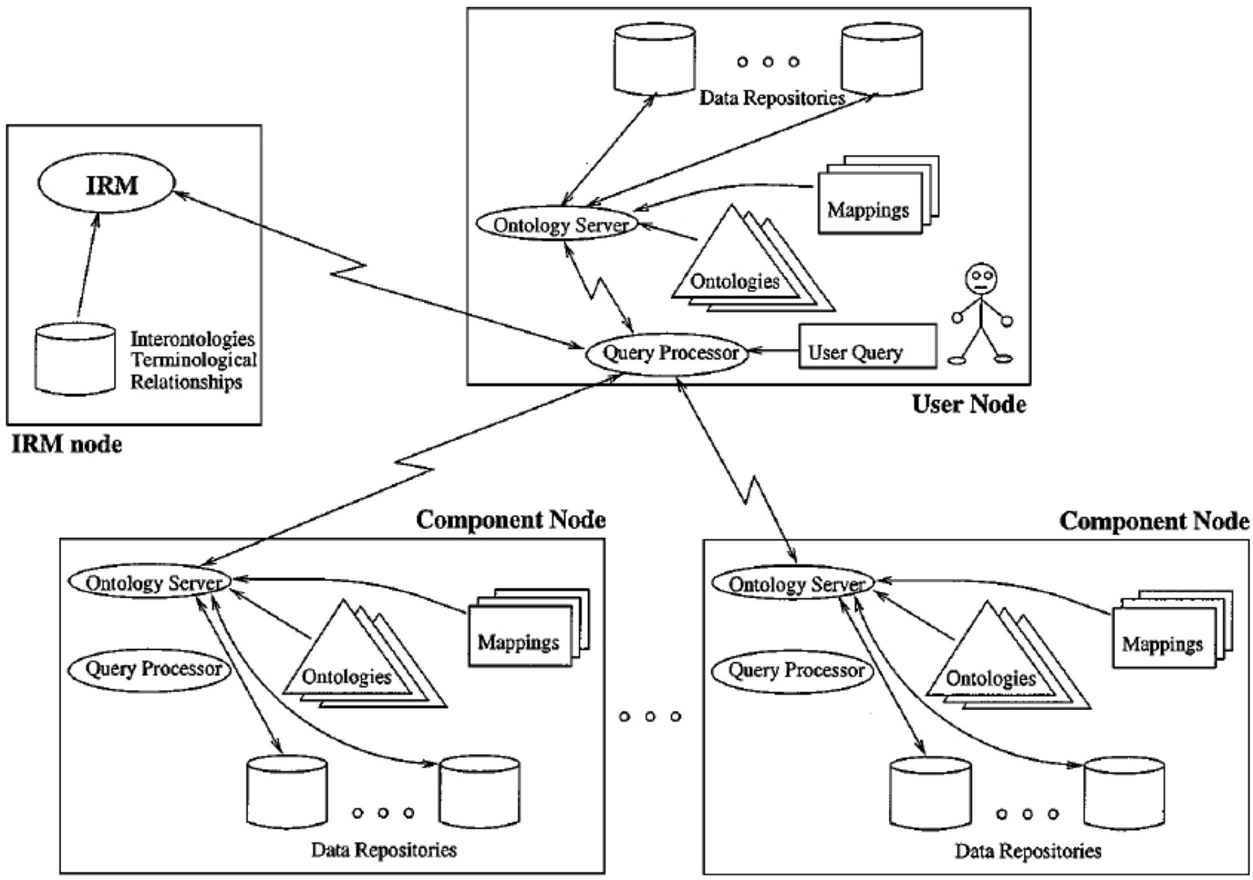

Figure 4

general web applications may only need lightweight ontologies; and recognition that in certain commercial applications, the use of powerful heavyweight ontologies will easily recoup the cost (Shadbolt et al. 2006). Recently a cost estimation approach has been developed (ONTOCOM; a detailed description is available in (Bontas and Mochol 2006) and an example of its application to a particular ontology (DILIGENT) is described (Bontas and Tempich 2005), though the actual results of the many formulas upon the various cost drivers are not included in this publication. These cost drivers include:

Product factors: complexity of the domain analysis, conceptualization, implementation, instantiation, evaluation, integration, reusability, and documentation (Institut für Informatik 2006):

- Personnel factors: ontologist/domain expert capability & experience, language and tool experience, and personnel continuity
- Project factors: tool support, multi-site development, and required development schedule
- Reuse/maintenance factors: ontology understandability, domain/expert unfamiliarity, and complexity of evaluation, modifications, and translations

Development of an ontology requires a shared conceptualization by domain experts, users and designers (de Bruijn 2003, 5); this is not only difficult, but requires such a high initial investment, it will only be supportable where there is commercial interest (Stuckenschmidt and van Harmelen 2005, 249). While the initial cost of ontology implementation is frightening, one IBM researcher predicts the long term maintenance of an ontology to be 80% of the cost (Welty 2005). In a recent survey of 34 ontology engineering projects, half of which were commercial, all participants emphasized the resource-intensive nature of domain analysis and the lack of low barrier methods and tools (Simperl and Tempich 2006). The implications are that there must be a clear and pressing need for the benefits of ontological indexing and retrieval, sufficient to provide extensive funding or the dedicated volunteer labor of known and trusted professionals. From the limited survey of the landscape performed for this report, it appears that funding is currently available in medical fields, environmental research, national defense, and business applications. The educational field may contain sufficient volunteer experts, university support, and grant-funded development to make ontology development feasible for instructional materials.

To be able to effectively apply an ontology, much less change it, one must learn it, another time-consuming task. Apart from domain knowledge, the person encoding the document must have a level of understanding approaching that of a skilled knowledge engineer (Marshall and Shipman 2003). To expect the average citizen to have or develop the necessary knowledge and skill to coherently apply a domain ontology to a document is infeasible (Marshall 2004). If the users will not apply the ontologies, then the application of metadata to resources must be performed by the institution or service. Hence the users only bear the cost if they pay for the service, either directly or indirectly; this implies that ontologies may indeed only be feasible, in the long term, for applications in commercial services.

The only other solution to this cost would be the automation of application of ontologies to resources. The development of this functionality depends heavily on research and tools developed by the artificial intelligence community. Some of the techniques developed include a noun phrasing technique for concept extraction and concept association based on context, frequency and co-occurrence of terms (Chen 1999). However, precise meanings for every relation are necessary for automatic classification (Weinstein and Birmingham 1998). A 2003 assessment stated that there are a number of issues to be resolved before natural language can be understood by computers; and the majority of information present on the web is in natural language (Fensel 2003a). However, for technical fields with more structured terminology, a text-mining system for scientific literature, Textpresso, shows considerable promise for assisting in automatic ontology annotation. While the machine cannot replace the human expert, it can increase efficiency greatly (Müller et al. 2004). Further investigation into current developments in this area is warranted.

For the ontology to be widely usable and interoperable, cross-mapping to other ontologies and domains is necessary, requiring the involvement of multiple domain experts (Adam, Atluri, and Adiwijaya 2000). And ontologies (and their supporting software) must be expected to change (de Bruijn 2003, 35), as knowledge and terminology are continually evolving. It is quite possible that this aspect may restrict the usability of ontologies to specified domains. Cross-mapping is only likely if there is sufficient need and funding to offset the expense, and then it is not likely to be maintained over time without continued funding and demand.

4. Conclusions

The decision about if, when, and how one should apply ontologies to one's digital library is a complex one. There are many aspects to consider, and several of those aspects are moving targets. Any assessment or survey, such as this one, can only be a snapshot of an evolving landscape, and as such, is useful primarily in helping one get his bearings for the moment. Further research and feasibility studies are necessary components for any digital library considering the application of ontologies.

Some purposes of ontologies may be particularly useful. Dieter Fensel predicted in 2003 that three areas in which ontology application potentially has a huge impact are knowledge management, enterprise integration, and e-commerce (Fensel 2003b). Already this prediction seems to be proving true. If one's digital library falls into these domains, the usefulness may outweigh the cost: funding created by demand for a service may well be sufficient to overcome other obstacles. Usefulness in educational realms seems quite promising, but the return on investment has yet to be proven (Milam 2005).

Outside of heavily funded domains, feasibility is yet to be determined. If the target audience for the digital library is the general public, at no cost to the user, then it is not likely that the application of ontologies is currently monetarily feasible. Ontologies incur tremendous expenditures of resources in their creation or adoption, application, cross-mapping, maintenance, and possibly software development. Tools exist to assist in modifying existing ontologies, but they are not simple, and require extensive domain knowledge and understanding of the concepts and relations required for the ontology to be functional. Tools to apply ontologies to existing resources are still under development. Cross-mapping ontologies for use beyond a single domain is a new territory; if the source ontologies have the same basis, query engines appear to have good results, but that's a rather telling caveat. Otherwise, it seems that only general mappings are feasible, supporting general queries with limited precision. To some extent, mappings can be automated, but must still be reviewed by a human.

Systems to support ontology use (query engines and semantic web browsers) are becoming available, but their usefulness is limited by the ontologies and their mappings. And the cost of maintenance and continual evolution of an ontology is yet unmeasured. On the other hand, a general ontology language has been adopted by W3C, tools and systems continue to evolve, and new ontologies appear every year. If funding exists, and an acceptable ontology exists in OWL for a domain covered by a particular digital library, it would be reasonable to assess the existing tools for application and delivery, and possibly move forward in implementation. As ontology and tool development lowers the technical and cost barriers, general digital libraries should certainly become involved: this is perhaps in the very near future.

References

ACS. *World Ranking Thesaurus Software*. Active Classification Solutions. http://www.termtree.com.au/, viewed 1 June 2007.

Adam, Nabil R., Atluri, Vijayalakshmi and Adiwijaya, Igg. 2000. SI [system integration] in digital libraries. *Communications of the ACM*, 46: 6.

Aquilera, Vincent, Cluet Sophie, Veltri, Pierangelo, Vodislav, Dan, and Wattez, Fanny. 2000. Querying XML documents in Xyleme. In *Proceedings of the ACM SIGIR Workshop on XML and Information Retrieval*, 28 July 2000. http://www.haifa.il.ibm.com/sigir00-xml/final-papers/xyleme/XylemeQuery/XylemeQuery.html.

Becker, Peter, Green, Steve, and Roberts, Nataliya. *Ontorama. knowledge, visualization and ordering laboratory*. http://www.kvocentral.org/software/ontorama.html, viewed 1 June 2007.

Beckett, Dave, Steer, Damian, Heery, Rachel, and Johnston, Pete. 2003. *MEG registry project*. UKOLN Metadata for Education Group. http://www.ukoln.ac.uk/metadata/education/regproj/, last updated 12 September 2003.

Bergamaschi, Sonia, Beneventano, Domenico, Guerra, Francesco, and Vincini, Maurizio. 2005. Building a tourism information provider with the MOMIS system. *Journal of information technology and tourism* 7: 3-4. http://tourism.wu-wien.ac.at/Jitt/JITT_7_34_Bergamaschi_et_al.pdf.

Berners-Lee, Tim. 2005. Putting the Web Back in Semantic Web. In *4th International Semantic Web Conference*, November, 2005, slide 17. http://www.w3.org/2005/Talks/1110-iswc-tbli/.

Bertini, M., Del Bimbo, A., and Torniai, C. 2005. Automatic video annotation using ontologies extended with visual information. In *ACM Multimedia Conference 2005*. ACM, November 2005.

Binding, Ceri and Tudhope, Douglas. 2004. KOS at your Service: programmatic access to knowledge organisation systems. *Journal of digital information* 4: 4. http://jodi.tamu.edu/Articles/v04/i04/Binding.

Bockting, Sander. 2005. A semantic translation service using ontologies. In University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science *3rd Twente Student Conference on IT*, Enschede, 20 June 2005. http://bockting.student.utwente.nl/documents/semantic_translation_service_using_ontologies.pdf.

Bontas, Elena Paslaru., and Mochol, Malgorzata. 2006. *Ontology engineering cost estimation with ONTOCOM*. Technical Report TR-B-06-01, Freie Universität Berlin, Germany, 7 February 2006. http://ontocom.ag-nbi.de/docs/tr-b-06-01.pdf.

Bontas, Elena Paslaru, and Tempich, Christoph. 2005. *How much does it cost? Applying ONTOCOM to DILIGENT*. Technical report TR-B-05-20, Freie Universität Berlin, Germany, 27 October 2005. http://ontocom.ag-nbi.de/docs/tr-b-05-20.pdf.

Brickley, Dan, Miller, Libby., Hunter, Jane and Lagoze, Carl, Principal Investigators. 2002a About Harmony. A joint project of the Distributed Systems Technology Center (Australia), the Institute for Learning and Research Technology (UK), and Cornell Digital Library Research Group (USA). http://metadata.net/harmony/index.html, viewed 29 May 2007.

Brickley, Dan, Miller, Libby, Hunter, Jane and Lagoze, Carl, Principal Investigators. 2002b. Harmony: Results. A joint project of the Distributed Systems Technology Center (Australia), the Institute for Learning and Research Technology (UK), and Cornell Digital Library Research Group (USA). http://metadata.net/harmony/Results.htm, viewed 30 May 2007.

Buckland, Michael, Chen, Aitao, Gey, Fredric C., and Larson, Ray R. 2006. Search across different media: numeric data sets and text files. In *Information technology and libraries* 25: 182. http://metadata.sims.berkeley.edu/searchacross.pdf.

Butler, Mark. H., Gilbert, John, Seaborne, Andy, and Smathers, Kevin. 2004. *Data conversion, extraction and record linkage using XML and RDF tools in Project SIMILE*. Hewlett-Packard Company Technical Report HPL-2004-147, 31 August 2004. http://metadata.sims.berkeley.edu/searchacross.pdf.

CES 2003. *MetaNet: An Overview*. A project of the European Commission Community Research Information Society Technologies, published by the University of Edinburgh, Research Centre of the School of Education, Centre for Educational Sociology. http://www.epros.ed.ac.uk/metanet/, last modified 3 November 2003.

Chalupsky, Hans. *OntoMorph: a translation system for symbolic knowledge*. University of Southern California Information Sciences Institute. http://www.isi.edu/~hans/ontomorph/presentation/ontomorph.html, viewed 30 May 2007.

Chen, Hsinchun. 1999. Semantic research for digital libraries. *D-Lib Magazine* 5: 10. http://www.dlib.org/dlib/october99/chen/10chen.html.

Cross, Phil, Brickley, Dan, and Koch, Traugott. 2003. RDF thesaurus specification (draft). Institute for Learning and Research Technology Technical Report Number 1011, 21 July 2003. http://www.ilrt.bris.ac.uk/publications/researchreport/rr1011/report_html.

CyCorp. 2007. *What is Cyc?* CyCorp, Inc. http://www.cyc.com/cyc/technology/whatiscyc, viewed 29 May 2007.

CyCorp. 2007. *Opencyc.org: OpenCyc license information*. CyCorp, Inc. http://www.opencyc.org/license, viewed 29 May 2007.

CyCorp. 2007. *ResearchCyc*. CyCorp, Inc. http://research.cyc.com/, viewed 29 May 2007.

CyCorp. 2002a *The Syntax of CycL*. CyCorp, last updated 28 March 2002. http://www.cyc.com/cycdoc/ref/cycl-syntax.html

CyCorp. 2002b. *Frequently Asked Questions about OpenCyc, Version 07b*. CyCorp, last updated 20 September 2002. http://www.opencyc.org/faq/opencyc_faq

Davies, John, Fensel, Dieter, and Van Harmelen, Frank, editors. 2003. *Towards the semantic web: ontology-driven knowledge management.* West Sussex: John Wiley & Sons.

de Bruijn, Jos. 2003. Using ontologies: enabling knowledge sharing and reuse on the semantic web. *Digital Enterprise Research Institute* Technical Report DERI-2003-10-29, October 2003. http://www.deri.at/fileadmin/documents/DERI-TR-2003-10-29.pdf.

de Bruijn, Jos and Polleres, Axel. 2004. Towards an ontology mapping specification language for the semantic web. *Digital Enterprise Research Institute* Technical Report DERI-2004-06-30, June 2004. http://www.deri.at/fileadmin/documents/DERI-TR-2004-06-30.pdf.

Delugach, Harry, Editor. 2007. *Common logic standard*. ISO Final Draft International Standard 24707. http://cl.tamu.edu/, viewed 30 May 2007.

DGRC. The EDC Project. Digital Government Research Center. http://www.isi.edu/dgrc/dgrc-research.html, viewed 29 May 2007.

Dicheva, Darina, Sosnovsky, Sergey, Gavrilova, Tatiana, and Brusilovsky, Peter. 2006. Ontologies for Education Portal. A collaborative project of Winston Salem State University, University of Pittsburgh, and Saint-Petersburg State Polytechnic University. http://iiscs.wssu.edu/o4e/viewhome.do?tm=O4E.xtm, viewed 1 April 2007.

Doerr, Martin, Hunter, Jane, and Lagoze, Carl. 2003. Towards a core ontology for information integration. *Journal of digital information* 4:1, Article 169, 9. http://jodi.ecs.soton.ac.uk/Articles/v04/i01/Doerr/.

Domingue, John. WebOnto. *The Open University Knowledge Media Institute*. http://kmi.open.ac.uk/projects/webonto/, viewed 30 May 2007.

Domingue, John, and Motta, Enrico. PlanetOnto. *The Open University Knowledge Media Institute*. http://kmi.open.ac.uk/projects/planetonto/, viewed 30 May 2007.

Dublin Core Metadata Initiative. 2004. *Dublin Core Metadata Element Set, Version 1.1: reference description*. Dublin Core Metadata Initiative, 20 December 2004. http://dublincore.org/documents/dces/, viewed 29 May 2007.

Electronic Cultural Atlas Initiative. 2006. *Support for the learner: what, where, when, and who*. University of California, Berkeley. http://ecai.org/imls2004/, viewed 29 May 2007.

Faculty of Engineering at Modena. 2004. The Mediator envirOnment for Multiple Information Sources (MOMIS) Project. University of Modena e Reggio Emilia, 2 October 2004. http://www.dbgroup.unimo.it/Momis/, viewed 15 May 2007.

Fensel, Dieter. 2003a. From a presentation for the *Next Web Generation* Seminar at the University of Innsbruck, Summer, 2003. In de Bruijn, J. Using Ontologies: Enabling Knowledge Sharing and Reuse on the Semantic Web. *Digital Enterprise Research Institute* Technical Report DERI- 2003-10-29, October 2003. http://www.deri.at/fileadmin/documents/DERI-TR-2003-10-29.pdf.

Fensel, Dieter. 2003b *Ontologies: a silver bullet for knowledge management and electronic commerce*. In de Bruijn, J. *Using ontologies: enabling knowledge sharing and reuse on the semantic web*. Berlin: Springer-Verlag. *Digital Enterprise Research Institute* Technical Report DERI-2003-10-29, October 2003. http://www.deri.at/fileadmin/documents/DERI-TR-2003-10-29.pdf.

Fensel, Dieter, Hendler, Jim, Lieberman, Henry, and Wahlster, Wolfgang. 2003c Introduction. In *Spinning the semantic web*. (Cambridge: MIT Press) http://w5.cs.uni-sb.de/teaching/ws03/internetagenten/Introduction.pdf, viewed 15 May 2007.

Food and Agriculture Organization of the United Nations. 2007. Agriculture Information Management Standards: AGROVOC Thesaurus. http://www.fao.org/aims/ag_intro.htm, last updated 22 May 2007.

FZI WIM and AIFB LS3. 2007. KAON Tool Suite. http://kaon.semanticweb.org/, last updated 10 May 2005.

Genesreth, Michael R. 2004. *Knowledge interchange format*: draft proposed American National Standard (dbANS), NCITS.T2/98-004. Stanford Logic Group, Stanford University. http://logic.stanford.edu/kif/dpans.html, viewed 14 April 2007.

Gilbert, John, and Butler, Mark H. 2003. *Review of existing tools for working with schemas, metadata, and thesauri*. Hewlett-Packard Company Technical Report HPL-2003-218, 6 November 2003. http://www.hpl.hp.com/techreports/2003/HPL-2003-218.pdf.

Gray, Peter, Gray, Alex, Fiddian, Nick, Shave, Michael, and Bench-Capon, Trevor, Principal Investigators. 2000. KRAFT: Knowledge Reuse & Fusion/Transformation. A joint project of The University of Aberdeen Computer Science Department, The Cardiff University School of Computer Science, and The University of Liverpool Computer Science Department. http://www.csd.abdn.ac.uk/~apreece/Research/KRAFT/, viewed 30 May 2007.

Hjørland, Birger. 2007. Knowledge organization systems. *Core concepts in library and information science (LIS)*, 11 February 2007. http://www.db.dk/bh/lifeboat_ko/CONCEPTS/knowledge_organization_systems.htm, viewed 15 May 2007.

Hovy, Eduard. 2003. Using an ontology to simplify data access. *Communications of the ACM* 46.

Hollink, Laura, Worring, Marcel, and Schreiber, A. Th. (Guus). 2005. Building a visual ontology for video retrieval. In *ACM Multimedia Conference 2005*. ACM, November 2005. http://www.cs.vu.nl/%7Eguus/papers/Hollink05b.pdf , viewed 10 November 2007. Ontology available at http://

appling.kent.edu/nsdlgreen/default.htm, viewed 29 May 2007.

Hsu, Eric I. 2003. *Wine agent 1.0: how does it work*? Stanford University Knowledge Systems Artificial Intelligence Laboratory, last updated 8 April 2003. http://www.ksl.stanford.edu/projects/wine/explanation.html

Hunter, Jane. 2001. MetaNet; a metadata term thesaurus to enable semantic interoperability between metadata domains. *Journal of Digital Information* 1:8, No. 42, 8 February 2001. http://jodi.tamu.edu/Articles/v01/i08/Hunter/.

Hüsemann, Bodo. 2006. *OntoMedia*. University of Muenster. http://www.ontomedia.de/, viewed 29 May 2007.

ICOM/CIDOC Document Standards Group and CIDOC CRM Special Interest Group. 2005. *Definition of the CIDOC Conceptual Reference Model, Version 4.2*. edited by Crofts, Nick, Doerr, Martin, Gill, Tony, Stead, Stephen, and Stiff, Matthew, June 2005. http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.2.pdf.

IFLA. 1998. *Functional requirements for bibliographic records.* Final Report. *International Federation of Library Associations and Institutions*, Cataloguing Section, FRBR Review Group. http://www.ifla.org/VII/s13/frbr/frbr.htm, viewed 29 May 2007.

InfoMaster. 2006. InfoMaster: the power to make better decisions. Efekt Pty Ltd. http://www.infomaster.com.au/, viewed 29 May 2007.

IMS. 2007. Learning resource meta-data specification. *IMS Global Learning Consortium, Inc*. http://www.imsproject.org/metadata/, viewed 29 May 2007.

Information Sciences Institute, University of Southern California. *Large resources: ontologies (SENSUS) and lexicons*. Information Sciences Institute, University of Southern California. http://www.isi.edu/natural-language/projects/ONTOLOGIES.html, viewed 29 May 2007.

Institut für Informatik. 2006a. Ontology engineering cost estimation with ONTOCOM. *Institut für Informatik*, Networked Information Systems, Freie Universität Berlin. http://ontocom.ag-nbi.de/index.html, viewed 1 June 2007.

Institut für Informatik. 2006b. ONTOCOM cost drivers. *Institut für Informatik*, Networked Information Systems, Freie Universität Berlin. http://ontocom.ag-nbi.de/ontocom.html, viewed 1 June 2007.

International Standards Organization. 2006. *Information technology – common logic (CL): a framework for a family of logic-based languages*. (ISO/IEC JTC 1/SC 32 N 1498), 31 December 2006. http://cl.tamu.edu/docs/cl/24707-31-Dec-2006.pdf.

International Standards Organization. 2004. *MPEG-7 overview*. (ISO/IEC JTC1/SC29/WG11 N 6828) Coding of Moving Pictures and Audio. http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm, viewed 29 May 2007.

International Standards Organization, International Electrotechnical Commission. 2004. *Document schema definition languages (DSDL) – part 3: rule-based validation – schematron*. (ISO/IEC FDIS 19757-3). http://www.schematron.com/iso/dsdl-3-fdis.pdf.

Jelliffe, Rick. *Schematron: a language for making assertions about patterns found in XML documents*. http://www.schematron.com/overview.html, viewed 30 May 2007.

Jones, D.M, Bench-Capon, T.J.M., and Visser, P.R.S. 1998. Methodologies for ontology development. In José Cuena, ed., *IT&KNOWS information technology and knowledge systems: Proceedings of the XV. IFIP World Computer Congress, 31 Aug.-4 Sept. 1998, Vienna, Austria and Budapest, Hungary*. Vienna: Austrian Computer Society/International Federation for Information Processing, pp. 62-75.

Kabel, S., de Hoog, R., Wielinga, B.J., Anjewierden, A. 2004. The added value of task and ontology-based markup for information retrieval. *Journal of the American Society for Information Science and Technology* 55:348-62.

Kahn, L., McLeod, D., and Hovy, E. 2004. Retrieval effectiveness of an ontology-based model for information selection. *The international journal on very large data bases* 13: 71- 85.

Karger, David R., Principal Investigator. *Haystack Project*. Massachusetts Institute of Technology Computer Science and Artificial Intelligence Laboratory (MIT CSAIL). http://haystack.lcs.mit.edu/, viewed 1 June 2007.

Kent State University. Green's Functions Digital Library. A collaborative effort of Kent State University, National Institute of Standards and Testing (NIST), and Massachusetts Institute of Technology (MIT). http://appling.kent.edu/nsdlgreen/default.htm, viewed 29 May 2007.

Klein, Michael. 2001. Combining and relating ontologies: an analysis of problems and solutions. In *International Joint Conferences on Artificial Intelligence Workshop on Ontologies and Information*

Sharing. http://www.informatik.uni-bremen.de/agki/www/buster/IJCAIwp/Finals/klein.pdf.

Knowledge Media Institute and the Open University. 2004. Scholarly ontologies project: Summary. http://kmi.open.ac.uk/projects/scholonto/summary.html, viewed 30 May 2007.

Knowledge Media Institute. 2000. *OCML: operational conceptual modelling language*. http://kmi.open.ac.uk/projects/ocml/ viewed 14 March 2007.

KSL. 2005a Chimæra. Stanford University Computer Science Department, Knowledge Systems Artificial Intelligence Laboratory. http://www-ksl.stanford.edu/software/chimaera/, viewed 30 May 2007.

KSL. 2005b Ontolingua. Stanford University Computer Science Department, Knowledge Systems Artificial Intelligence Laboratory. http://www-ksl.stanford.edu/software/ontolingua/, viewed 1 June 2007.

Lagoze, Carl and Hunter, Jane. 2001. The ABC ontology and model. *Journal of digital information* 2:2, Art. 77. http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Lagoze/.

Lawrence, Faith, Tuffield, Mischa M., Jewell, Mike O., Prügel-Bennett, Adam, Millard, David E., Nixon, Mark S., Schraefel, Monica, and Shadbolt, Nigel R. 2005. OntoMedia creating an ontology for marking up the contents of heterogenous media. *In Proceedings Ontology Patterns for the Semantic Web ISWC-05 Workshop*. http://eprints.ecs.soton.ac.uk/11153/01/onto_workshop.pdf.

Leuf, Bo. 2006. *The semantic web: crafting infrastructure for agency*. West Sussex: John Wiley & Sons.

Lin, David and Hunter, Jane. 2001. *ABC Metadata Model Constructor: The ABC Ontology*. A result of the DSTC (Australia), JISC (UK), and NSF (US) funded Harmony Project. Developed under the direction of Carl Lagoze. http://metadata.net/harmony/constructor/ABC_Constructor.htm, viewed 30 May 2007.

Library of Congress. 2007. *MARC standards*. Library of Congress, Network Development and MARC Standards Office. http://www.loc.gov/marc/, last updated 24 January 2007.

LSDIS. 2005. *ADEPT: Alexandria Digital Earth Prototype* (1999-2004). Large Scale Distributed Information Systems, University of Georgia, Computer Science Department. http://lsdis.cs.uga.edu/projects/past/ADEPT/, viewed 29 May 2007.

Marshall, Catherine C. 2004. *Taking a stand on the semantic web*. http://www.csdl.tamu.edu/~marshall/mc-semantic-web.html, viewed 15 May 2007.

Marshall, Catherine C. and Shipman, Frank M. 2003. Which semantic web? In *HyperText '03 Conference*, Nottingham, UK, 26-30 August, 2003. Copyright 2003 ACM. http://www.csdl.tamu.edu/~marshall/ht03-sw-4.pdf, viewed 15 May 2007.

MatML Working Group. 2004. *MatML: XML for materials property data*. http://www.matml.org/schema.htm, viewed 29 May 2007.

Mazzocchi, Stephano, Garland, Stephen, and Lee, Ryan. 2005. SIMILE: practical metadata for the semantic web. In O'Reilly XML.com 2005. *XML From the Inside Out*, 26 January 2005. http://www.xml.com/pub/a/2005/01/26/simile.html

Meersman, Robert. 1999. Semantic ontology tools in is design. In Zbigniew W. Raś and Andrzej Skowron, eds., *Foundations of intelligent systems: 11th International Symposium, ISMIS'99, Warsaw, Poland, June 8-11, 1999: proceedings*. Berlin and New York: Springer, 30-45.

Mena, Eduardo, Illarramendi, Arantza, Kashyap, Vipul, and Sheth, Amit P. 2000. OBSERVER: an approach for query processing in global information systems based on interoperation across preexisting ontologies. *Distributed and Parallel Databases* 8: 223-71.

Menzies, Tim. 1999. Cost benefits of ontologies. *Intelligence* 10, no. 3: 26-32.

Milam, John. 2005. Ontologies in higher education. In HigherEd.org. http://highered.org/docs/milam-ontology.pdf, viewed 1 April 2007.

Mindswap Lab. 2006. *Pellet: an OWL DL reasoner*. Developed at the University of Maryland's Mindswap Lab, commercially supported by Clark & Parisial, LLC. http://pellet.owldl.com/, last updated 4 November 2006.

Miller, George A. *WordNet: a lexical database for the English language*. Cognitive Science Laboratory, Princeton University. http://wordnet.princeton.edu/, viewed 29 May 2007.

MIT 2007a *Simile: semantic interoperability of metadata in unlike environments*. A joint project of Massachusetts Institute of Technology Libraries and Massachusetts Institute of Technology Computer Science and Artificial Intelligence Laboratory. http://simile.mit.edu/, viewed 29 May 2007.

MIT and Hewlett Packard. 2007b *DSpace: Welcome to DSpace*. A joint project of Massachusetts Institute of Technology Libraries and Hewlett Packard. http://www.dspace.org/, viewed 29 May 2007.

Motta, Enrico. *OCML: operational conceptual modelling language*. The Open University Knowledge Media Insitute. http://kmi.open.ac.uk/projects/ocml/, viewed 30 May 2007.

Müller, Hans-Michael, Kenny, Eimear E., and Sternberg, Paul W. 2004. Textpresso: an ontology-based information retrieval and extraction system for biological literature. *Public library of science: biology* 2: 11. http://www.pubmedcentral.nih.gov/articleren-der.fcgi?tool=pubmed&pubmedid=15383839.

Multisystems. 2007. *Thesaurus construction and publishing solutions*. Multisystems. http://www.multites.com/, viewed 1 June 2007.

Noy, Natasha. *Prompt*. Stanford Medical Informatics, Stanford University. http://protege.stanford.edu/plugins/prompt/prompt.html, viewed 30 May 2007.

OCLC. 2007. Learn more about LC Name Authority Service. *Online Computer Library Center Programs and Research, ResearchWorks*. http://www.oclc.org/research/researchworks/authority/default.htm, viewed 29 May 2007.

OKBC Working Group. 1995. *Open Knowledge Base Connectivity Home Page*. A joint project of Cy-Corp, Information Sciences Institute, Stanford Knowledge Systems Laboratory, Science Applications International Corporation (SAIC), SRI International, and Teknowledge; Richard Fikes, working group chair. http://www.ai.sri.com/~okbc/, viewed 30 May 2007.

Ontology Works, Inc. 2005. *Ontology Works knowledge server*. http://www.ontologyworks.com/ks.php, viewed 8 March 2007.

Ontoprise. 2007. *Know how to use know-how*. Ontoprise GmbH. http://www.ontoprise.de/content/, viewed 29 May 2007.

Palmer, Matthias, Naeve, Ambjörn, Enoksson, Fredrik, Nilsson, Mikael, Eriksson, Henrik, Danils, Jan, and Stark, Jöran. 2007. *SHAME: standardized hyper adaptible metadata editor*. http://kmr.nada.kth.se/shame/wiki/Overview/Main, last updated 5 December 2006.

Paralič, Jan and Kostial, Ivan. 2003. Ontology-based Information Retrieval. In *Proceedings of the 14th International Conference on Information and Intelligent Systems (IIS2003)*, Varadin, Croatia. ISBN 953-6071-22-3, 23-28.

Parallel Understanding Systems Group. *SHOE: simple html ontology extensions*. Department of Computer Science, University of Maryland at College Park. http://www.cs.umd.edu/projects/plus/SHOE/index.html, viewed 30 May 2007.

Pazienza, MariaTeresa., Stellato, Armando, Vindigni, Michele, Zanzotto, Fabio Massimo. 2004. XeOML: an XML-based extensible ontology mapping language. Paper presented at the *3rd International Semantic Web Conference (ISWC2004)* in Hiroshima, Japan, November 2004. http://ai-nlp.info.uniroma2.it/stellato/publications/2004_ISWC-04_XeOML%20An%20XML-based%20extensible%20Ontology%20Mapping%20Language.pdf, viewed 6 February 2007.

Petras, Vivien, Larson, Ray, and Buckland, Michael. 2006. Time period directories: a metadata infrastructure for placing events in temporal and geographic context. In *Opening information horizons: Joint Conference on Digital Libraries*, Chapel Hill, NC, 11-15 June 2006. http://metadata.sims.berkeley.edu/tpdJCDL06.pdf, viewed 24 February 2007.

Pietriga, Emmanuel. 2007. *IsaViz: a visual authoring tool for RDF*. World Wide Web Consortium, RDF Developer. http://www.w3.org/2001/11/IsaViz/, May 2007.

Riva, Alberto. *LispWeb*. Common Lisp Web Server. http://snpper.chip.org/lispweb, viewed 30 May 2007.

Russ, Tom and Patil, Ramesh. 2006. *Loom ontosaurus*. University of Southern California Information Sciences Institute. http://www.isi.edu/isd/ontosaurus.html, last updated 5 December 2006.

Rys, Michael. 1998. *The Stanford-IBM Manager of Multiple Information Sources (TSIMMIS)*. Stanford University, last updated 4 April 1998. http://www-db.stanford.edu/tsimmis/, viewed 3 February 2007.

Shadbolt, Nigel, Hall, Wendy, and Berners-Lee, Tim. 2006. The semantic web revisited. *IEEE Intelligent Systems* 21: 96-101. http://eprints.ecs.soton.ac.uk/12614/01/Semantic_Web_Revisted.pdf.

Shreve, Gregory M. and Zeng, Marcia Lei. 2003. Integrating resource metadata and domain markup in an NSDL collection. In *Proceedings of the International DCMI Metadata Conference and Workshop*, Seattle, WA, 28 September - 2 October, 2003. http://www.siderean.com/dc2003/604_paper62.pdf, viewed 16 March 2007.

Shum, Simon Buckingham, Motta, Enrico and Dominigue, John. 2000. ScholOnto: an ontology-based digital library server for research documents and discourse. *International Journal on Digital Libraries* 3: 3. http://kmi.open.ac.uk/projects/scholonto/docs/ScholOnto-IJoDL-2000.pdf, viewed 4 February 2007.

SID. 2001. Interoperable Database Group, Research Group of Distributed Information Systems (SID). http://sid.cps.unizar.es/OBSERVER/, updated 12 April 2001.

Silva, Nuno, Varady, Zoltan, Westerhausen, Frank, Fodor, Oliver, Silva, PedroVieira, and Maio, Paulo. 2006. *MAFRA toolkit*. http://mafra-toolkit.sourceforge.net/, last updated 2 January 2006.

Simperl, Elena Paslaru Bontas, and Tempich, Christoph. 2006. Ontology engineering: a reality check. In *5th International Conference on Ontologies, Databases, and Applications of Semantics*. http://ontocom.ag-nbi.de/docs/odbase2006.pdf, viewed 16 March 2007.

Sirin, Evren. *Simple Instance Creator (SIC)*. Maryland Information and Network Dynamics Lab Semantic Web Agents Project (MINDSWAP). http://www.mindswap.org/~evren/SIC/, viewed 1 June 2007.

Smith, Terence R., Zeng, Marcia L., and the ADEPT Project Team. 2004. Building semantic tools for concept-based learning spaces: knowledge bases of strongly-structured models for scientific concepts in advanced digital libraries. *Journal of digital information* 4:4, Art. 263, 28 January 2004. http://jodi.tamu.edu/Articles/v04/i04/Smith/.

Soergel, Dagobert, Lauser, Boris, Liang, Anita, Fiesseha, Frehiwot, Keizer, Johannes, and Katz, Stephen. 2004. Reengineering thesauri for new applications: the AGROVOC Example. In *Journal of digital information* 4:3, No. 257. http://jodi.tamu.edu/Articles/v04/i04/Soergel.

Soo, Von-Wun, Lee, Chen-Yu and Yeh, Jaw Jium. Using sharable ontology to retrieve historical images. In *International Conference on Digital Libraries, Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries*, ACM Press, July 2002.

Stanford Medical Informatics. 2007. *Protégé*. Stanford University School of Medicine, Stanford Medical Informatics. http://protege.stanford.edu/, last updated 25 May 2007.

Steer, Damian. 2003a. *RDF author*. http://rdfweb.org/people/damian/RDFAuthor/, last modified 2 August 2003.

Steer, Damian. 2003b. *The Meg Registry Client Software (SCART)*. UKOLN Metadata for Education Group. http://www.ukoln.ac.uk/metadata/education/regproj/scart/, viewed 1 June 2007.

Stuckenschmidt, Heiner, and van Harmelen, Frank. 2005. *Information Sharing on the Semantic Web*. Berlin: Springer.

Swoogle. 2006. *Swoogle manual: FAQs*. University of Maryland, Baltimore County Ebiquity Research Group. http://swoogle.umbc.edu/index.php?option=com_swoogle_manual&manual=faq, viewed 16 March 2007.

Thesaurus Builder. 2007. *Thesaurus Builder thesaurus management software*. Thesaurus Builder. http://www.thesaurusbuilder.com/, viewed 1 June 2007.

TECUP Consortium. 2001. *INDECS: interoperability of data in e-commerce systems*. TECUP Consortium, lead by Universität Göttingen / Niedersächsische Staats- und Universitätsbibliothek (UNIGOE) as Coordinator. http://gdz.sub.uni-goettingen.de/tecup/indecs.htm, viewed 30 May 2007.

Telcordia Technologies. 2005. *The InfoSleuth agent system*. Applied Research Greenhouse, Telcordia Technologies. http://www.argreenhouse.com/InfoSleuth/index.shtml, viewed 11 April 2007.

U.S. National Library of Medicine. 2006a *UMLS metathesaurus fact sheet*. National Institutes of Health, U.S. Department of Health and Human Services, 28 March 2006. http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html, viewed 14 March 2007.

U.S. National Library of Medicine. 2006b. *Unified medical language system*. National Institutes of Health, U.S. Department of Health and Human Services, 28 March 2006. http://www.nlm.nih.gov/research/umls/about_umls.html, viewed 14 March 2007.

U.S. National Library of Medicine. 2006c. *SPECIALIST lexicon fact sheet*. National Institutes of Health, U.S. Department of Health and Human Services, 28 March 2006. http://www.nlm.nih.gov/pubs/factsheets/umlslex.html, viewed 14 March 2007.

U.S. National Library of Medicine. 2005. *UMLS knowledge source server (UMLSKS)*, Version 5.0. National Institutes of Health, U.S. Department of Health and Human Services, 30 August 2005. http://umlsks.nlm.nih.gov/, viewed 14 March 2007.

Visser, Ubbo and Hübner, Sebastian. 2003. *Bremen University Semantic Translator for Enhanced Retrieval (BUSTER)*. http://www.informatik.uni-bremen.de/agki/www/buster/new/application.html, last modified 25 May 2003.

Visser, Ubbo and Schuster, Gerhard. 2002. Finding and integration of information: a practical solution for the semantic web. In *Proceedings of ECAI 02, Workshop on Ontologies and Semantic Interoperability,* pp. 73-78. http://www.informatik.uni-bremen.de/agki/www/buster/papers/ECAI02WS.pdf viewed 14 March 2007.

VRA. 2007. *VRA core*. Visual Resources Association, The International Association of Image Media Professionals. http://www.vraweb.org/projects/vracore4/, viewed 29 May 2007.

WebChoir. 2006. *Project vocabulary tools*. WebChoir, Inc. http://www.webchoir.com/products/pvt.html, viewed 1 June 2007.

Weinstein, PeterC. and Birmingham, William P. 1998. Creating ontological metadata for digital library content and services. *International journal on digital libraries* 2: 20-37. http://deepblue.lib.umich.edu/handle/2027.42/42334.

Welty, Chris. 2004. Ontology maintenance support: text, tools, and theories. Presentation at the 7th International Protégé Conference, Bethesda MD. http://protege.stanford.edu/conference/2004/slides/2.1_Welty_Ontology_Maintenance_Support_v3.pdf, viewed 16 March 2007.

Wiederhold, Gio, Jannink, Jan, Prasenjit, Mitra, Decker, Stefan, and Vasan, Pichai S. *Scalable knowledge composition (SKC)*. Stanford University InfoLab. http://infolab.stanford.edu/SKC/, viewed 30 May 2007.

World Wide Web Consortium. 2007. *Extensible markup language (XML)*. W3C Architecture Domain. http://www.w3.org/XML/, viewed 29 May 2007.

World Wide Web Consortium. 2006. *XForms 10* (Second Edition): W3C Recommendation 14 March 2006. http://www.w3.org/TR/xforms/, viewed 26 March 2007.

World Wide Web Consortium. (2004a) *OWL web ontology language guide*. W3C Recommendation, February 2004. http://www.w3.org/TR/owl-guide/, viewed 26 March 2007.

World Wide Web Consortium. (2004b) *RDF primer: W3C recommendation* 10 February 2004. http://www.w3.org/TR/REC-rdf-syntax/, viewed 21 March 2007.

World Wide Web Consortium. 2004c. *RDF/XML syntax specification (revised)*. W3C Recommendation 10 February 2004. http://www.w3.org/TR/rdf-syntax-grammar/, viewed 29 May 2007.

World Wide Web Consortium. 2000. *Resource description framework (RDF) schema specification* 1.0. http://www.w3.org/TR/2000/CR-rdf-schema-20000327/, viewed 29 May 2007.

World Wide Web Consortium. 1999. *XSL transformations (XSLT)*, Version 1.0. W3C Recommendation 16 November 1999. http://www.w3.org/TR/xslt, viewed 29 May 2007.

XYLEME. 2006. *Xyleme: harness the power of XML*. XYLEME, 2006. http://www.xyleme.com/, viewed 29 May 2007.