Semantic (Web) Technology In Action: Ontology Driven Information Systems for Search, Integration and Analysis

Amit Sheth
Semagix and LSDIS Lab,
University of Georgia,
Athens, GA
amit@cs.uga.edu

Cartic Ramakrishnan
LSDIS Lab,
University of Georgia
Athens, GA
cartic@cs.uga.edu

Abstract

Semantics is seen as the key ingredient in the next phase of the Web infrastructure as well as the next generation of information systems applications. In this context, we review some reservations expressed about the viability of the Semantic Web. We respond to these by identifying a Semantic Technology that supports the key capabilities also needed to realize the Semantic Web vision, namely representing, acquiring and utilizing knowledge. Given that scalability is a key challenge, we briefly review our observations from developing three classes of real world applications and corresponding technology components: search/browsing, integration and analytics. We distinguish these proven technologies from some parts of the Semantic Web approach and offer subjective remarks, which we hope will foster further debate.

1 Introduction

Semantics is arguably the single most important ingredient in propelling the Web to its next phase, and is closely supported by Web services and Web processes that provide standards based interoperability of applications. Semantics is considered to be the best framework to deal with the heterogeneity, massive scale and dynamic nature of resources on the Web. Issues pertaining to semantics have been addressed in other fields like linguistics, knowledge representation and AI. The promise of semantics and challenges in developing semantic techniques are not new to researchers in the database and information system field either. For instance, semantics has been studied or applied in the context of data modeling, query and transaction processing etc. Recently, a group of both database and non-database researchers came together at the *Amicalola State Park* for an intensive look at the relationship between database research and the Semantic Web. During this collaboration, they identified three pages worth of opportunities to further database research while addressing the challenges in realizing the Semantic Web [16]. A follow on workshop also presented opportunity to present research at the intersection of database and the Semantic Web [`http://swdb.semanticweb.org`]. Nevertheless, many researchers in the database community continue to express significant reservations toward the Semantic Web. Table 1 shows some examples of criticisms or skeptical remarks about Semantic Web technology (taken from actual NSF proposal reviews and conference panel remarks).

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

"As a constituent technology, ontology work of this sort is defensible. As the basis for programmatic research and implementation, it is a speculative and immature technology of uncertain promise."
"Users will be able to use programs that can understand semantics of the data to help them answer complex questions .... This sort of hyperbole is characteristic of much of the genre of semantic web conjectures, papers, and proposals thus far. It is reminiscent of the AI hype of a decade ago and practical systems based on these ideas are no more in evidence now than they were then."
"Such research is fashionable at the moment, due in part to support from defense agencies, in part because the Web offers the first distributed environment that makes even the dream seem tractable."
"It (proposed research in Semantic Web) pre-supposes the availability of semantic information extracted from the base documents–an unsolved problem of many years, ..."
"Google has shown that huge improvements in search technology can be made without understanding semantics. Perhaps after a certain point, semantics are needed for further improvements, but a better argument is needed."

Table 1: Some Reservation among DB researchers about the Semantic Web

These reservations and skepticisms likely stem from a variety of reasons. First, this may be a product of the lofty goals of the Semantic Web as depicted in [3]. Specifically, database researchers may have reservations stemming from the overwhelming role of description logics in the W3C's Semantic Web Activity and related standards. The vision of the Semantic Web proposed in several articles may seem, to many readers, like a proposed solution to the long standing AI problems. Lastly, some skepticism stems from legitimate concern about the scalability of the three requisite core capabilities of the Semantic Web: (a) ontology creation and maintenance of large ontologies, (b) semantic annotation, and (c) inference mechanisms or other computing approaches involving large, realistic ontologies, metadata and heterogeneous data sets. Despite these reservations, some of them well justified, we believe semantic technology is beginning to mature and will play a significant role in the development of future information systems. We believe that database research will greatly benefit by playing critical roles in the development of both Semantic Technology and the Semantic Web. In addition, we also feel that the database community is very well equipped to play their part in realizing this vision. We aim here to:

- Identify some prevalent myths about the Semantic Web
- Identify instances of Semantic (Web) Technology in action and how the database community can make invaluable contributions to the same.

For the purpose of this article, as well as for tagging real and existing versus more futuristic and speculative alternatives, we distinguish between Semantic Technology and Semantic Web technology. By Semantic Technology [11] (a term that predates the "Semantic Web"), we imply application of techniques that support and exploit semantics of information (as opposed to syntax and structure/schematic issues [Sheth 99] [1]) to enhance existing information systems. In contrast, the Semantic Web technology (more specifically its vision) is best defined as *"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."* [3]. Currently in more practical terms, Semantic Web technology also implies the use of standards such as RDF/RDFS, and for some, OWL. It is however important to note that while description logic is a center piece for many Semantic Web researchers, it is not a necessary component for *all* Semantic Web applications. For the Semantic Technology as the term is used here, complex query processing, involving both metadata and ontology takes center stage, and is where database technology continues to play a critical role. For our purpose, semi-formal ontologies are those that do not claim formal semantics and/or are populated with partial or incomplete knowledge. For example, in such an ontology, all schema level constraints may not be observed in the knowledgebase that is instantiated *wrt* the ontology schema. This becomes especially relevant when ontology is populated by many persons or by extracting and integrating knowledge from multiple sources.

[1] For a commercial use of term "Semantic Technology" see [11].

2 Examples of Semantic (Web) Technology Applications and Observations

We summarize a few applications developed using a commercial technology to provide insights into what Semantic (Web) Technology can do today. Based on the increasing complexity and the deeper role of semantics, we divide the applications into three types[2]:

Semantic search and contextual browsing: In *Taalee* (now *Semagix*) Semantic Search Engine [18], the ontology consisted of general interest areas with five major categories and over 16 subcategories such as News, Entertainment and Sports. Blended Semantic Browsing and Querying (BSBQ) provided domain specific search (search based on these domain specific attributes) and contextual browsing. The application involved crawling/extracting audio, video and text content from well over 200 sources (e.g. CNN website). This application was commercially deployed for a Web-audio company called *Voquette*. An interesting related application not developed by Semagix is reported in [7].

Semantic integration: In *Equity Analyst's Workbench* [15] audio, video and text content from tens of sites and NewML feeds aggregated from 90+ international sources (such as news agencies of various countries) were constantly classified into a small taxonomy and domain specific metadata was automatically extracted. The equity market ontology used by this application consists of over one million facts (entity and relationship instances). An illustrative example of a complex semantic query involving metadata and ontology this application supported is: Show analyst reports (from many sources in various formats) that are competitors of Intel Corporation. In an application involving *Repertoire Management* for a multinational Entertainment conglomerate, its ontology with relatively simple schema is used to extract over 14.5 million instances. The application provided integrated access to heterogeneous content in the companys extensive media holding while addressing semantic heterogeneity (e.g., in naming of artist, tracks, etc.)

Analytics and Knowledge Discovery: In *Passenger Threat Assessment* application for homeland security [1] and *Semagix's Antimoney Laundering* solution [14], the knowledge base is populated from many public, licensed and proprietary knowledge sources using the CIRAS ontology. The resulting knowledge base has over one million instances. Periodic or continuous metadata extraction from tens of heterogeneous sources (150 files formats, HTML, XML, dynamic Web sites, databases, etc.) is also performed. When the appropriate computing infrastructure is used, the system is scalable to hundreds of sources, or about a million documents per day per server. A somewhat related non-Semagix business intelligence [9] application has demonstrated scalability by extracting metadata (albeit somewhat limited types of metadata with a significantly smaller ontology) from a billion pages [12].

Our experience in building the above real-world applications has led us to some empirical observations:

- Applications validate the importance of ontology in the current semantic approaches. An ontology represent a part of the domain or the real-world for which it represents and captures a shared knowledge around which the semantic application revolves. It is the "ontological commitment" reflecting agreement among the experts defining the ontology and its uses, that is the basis for "semantic normalization" necessary for semantic integration.
- Ontology population is critical. Among the ontologies developed by Semagix or using its technology, median size of ontology is over 1 million facts. This level of capture of knowledge makes the system very powerful. Since it is obvious that this is the sort of scale Semantic Web applications are going to be dealing with, means of populating ontologies with instance data need to be automated.
- Two of the most fundamental "semantic" techniques are named entity recognition and semantic ambiguity resolution (also closely tied to data quality problem). Without good solutions to these none of the applications listed above will be of any practical use. Both require highly multidisciplinary approaches, borrowing for NLP techniques, statistical and IR techniques and possibly machine learning techniques.

[2] At least the applications underlined are known to have been developed by commercial technology/product or deployed.

- Semi-formal ontologies that may be based on limited expressive power are most practical and useful. Formal or semi-formal ontologies represented in very expressive languages (compared to moderately expressive ones) have in practice, yielded little value in real-world applications. One reason for this may be difficulty capturing knowledge that uses the more expressive constructs of a representation language. This difficulty is especially apparent when trying to populate an ontology that uses a very expressive language to model a domain. Hence the additional effort in modeling these constructs for a particular domain is often not justifiable in terms of the gain in performance. Also there is a widely-accepted trade-off between expressive power of such representation languages and computational complexity associated with inference mechanisms for such languages. Practical applications often end up using languages that lie closer to less expressive languages in the "expressiveness vs. computational complexity continuum". This resonates with so-called Hendler's hypothesis ("little semantics goes a long way").
- Large scale metadata extraction and semantic annotation is possible. Storage and manipulation of metadata for millions to hundreds of millions of content items challenges us to apply and improve the performance and scalability of known database techniques in the presence of new, complex, structures.
- Support for heterogeneous data is key - it is too hard to deploy separate products within a single enterprise to deal with structured and unstructured data/content management. New applications involve extensive types of heterogeneity in format, media and access/delivery mechanisms (e.g., news feed in NewsML news, Web posted article in HTML or served up dynamically through database query and XSLT transformation, analyst report in PDF or WORD, subscription service with API-based access to Lexis/Nexis, etc). Database researchers have long studied this issue of integrating heterogeneous data.
- Semantic query processing with the ability to query both ontology and metadata to retrieve heterogeneous content is highly valuable. Consider the query "Give me all articles on the competitors of Intel", where ontology gives information on competitors, supports semantics (with the understanding that "Palm" is a company and that "Palm" and "Palm, Inc." are the same in this case), and metadata identifies the company an article refers to, regardless of format of the article. Analytical applications could require sub-second response time for tens of concurrent complex queries over large metadata base and ontology, and can benefit from further database research. High performance and highly scalable query processing that deal with more complex representations compared to database schemas and with more explicit role of relationships, is important. Database researcher can also contribute to the strategies of dealing with large RDF stores.
- A vast majority of the Semantic (Web) applications that have been developed or envisioned rely on three crucial capabilities namely ontology creation, semantic annotation and querying/inferencing. Enterprise scale application share many requirements in these three respects with pan Web applications. All these capabilities must scale to millions of documents and concepts (rather than hundreds to thousands).

3 Discussion

Ontologies come in bewildering variety; Figure 1 represents just three of the dimensions of variation. To keep a focus on real world applications and for the sake of brevity, we restrict the scope to task specific and domain specific ontologies. As observed recently by Gruber [6], currently the ontologies that are semi-formal have demonstrated very high practical value. We believe that ontology development effort for semi-formal ontologies can be significantly smaller compared to that required for developing formal ontologies or ontologies with more expressive representations. This is especially evident when considering ontology instance population efforts. Semi-formal ontologies have provided good examples of both value and utility. E.g., GO (`http://www.geneontology.org/`), is arguably a nomenclature from the perspective of representation and lacks formal and richer representation necessary to qualify as a formal ontology (discussed in more detail later).

Research in database and information systems have and will continue to play a critical role with respect to the scalability. We review the crucial scalability aspect of the three capabilities next.

3.1 Availability of large and "useful" ontologies

Although an ontology schema may resemble at a representational level a database schema, and instances may reflect database tuples, the fundamental difference is that ontology is supposed to capture some aspect of real-world or domain semantics, as well as represent ontological commitment forming the basis of semantic normalization. Methods for creating ontologies can be grouped into the following types:

- Social processes where a group of users go through a process of suggestions and iteratively revise versions of ontologies to capture domain semantics.
- Automatically (or semi-automatically) extract the ontology schema (i.e., ontology learning) from content of various kinds. Despite a recent spate of interest, this approach relates to the knowledge acquisition bottleneck faced in the AI research of eighties, and we have little practical experience to rely upon [5].
- Automatic or semi-automatic (with human curation) population of the knowledge base with respect to human design ontology schema. We can report on practical experience with a scalable approach of this type since several ontologies with over million instances have been create with a total of 4 to 8 weeks of effort (e.g., knowledge extraction in *SemagixFreedom* [15]).

3.2 Semantic Metadata Extraction or Semantic Annotation of massive content

Annotating heterogeneous content with semantics provided by relevant ontology (or ontologies) is a key challenge for the Semantic Web. Recently there have been commercial results providing detailed semantic annotations of heterogeneous content (structured, semi-structured, and unstructured with different formats) [15, 8], as well as research reporting annotation of over a billion Web pages [12]. As observed in the efforts on automatic semantic annotation, two resources necessary for realizing the semantic web are: (a) large scale availability of domain specific ontologies; and (b) scalable techniques to annotate content with high quality metadata descriptions based on the terms, concepts or relationships provided by these ontologies. We believe main area of challenge here is to support increasing number of practical and scalable techniques for semantic disambiguation.

3.3 Inference Mechanisms that Scale

Inference mechanisms that can deal with the massive number of assertions that would be encountered by Semantic Web applications are required. The claimed power behind many of the proposed applications of Semantic Web technology is the ability to infer knowledge that is not explicitly expressed. Needless to say, this feature has attracted attention from the AI community since they have been dealing with issues relating to inference mechanisms for quite some time. Inference mechanisms are applicable only in the context of formal ontologies. One of the most common knowledge representation languages has been Description Logic [10] on which DAML, one of the earliest Semantic Web languages is based. It was in fact one of the less expressive members of the DL family. The reason for limiting the expressive power of such a knowledge representation formalism was very clear when the decidability and complexity issues were considered. Although several optimized methods of inference were introduced later [2], inference mechanisms were still overshadowed by the performance advantage of traditional database systems. This has lead to reluctance among many database researchers to accept the Semantic Web vision as viable. Description Logics may form a part of some Semantic Web solutions of the future. We are however convinced it is not the only knowledge representation formalism that will go on and make the Semantic Web vision a reality. It is possible to do some sort of rudimentary inference using RDFS (using *subClassOf* and *subPropertyOf*). However, using RDF/RDFS does not force one to use *only* inference mechanisms of some sort in applications. Since RDFS has a graph model associated with it there is the possibility to use other techniques to answer complex queries [1].

3.4 Why semi-formal, less expressive ontologies?

Ontologies serve several purposes, including: having an agreement between humans, having a common representation for knowledge, having machines (software) get common interpretation of something that humans have agreed to, and forming the basis for defining metadata or semantic annotation. Tom Gruber, who many would credit with bringing the term to vogue in contemporary knowledge representation and information systems research, identifies three types of ontologies-informal, semiformal and formal [6]. He stresses the value of semi-formal ontology in meeting several challenges; especially that of information integration. Some researchers in the Semantic Web community would argue for only formal ontologies (and discount the value of semi-formal ontologies). We do not doubt that formal ontologies have a potential role in Semantic Web research. However, database researchers should particularly realize the value of and exploit semi-formal ontologies. Figure 1 stresses that there is a very large body of work that can and needs to be done using semi-formal ontologies.

GO ontology, which is more a nomenclature and taxonomy, than a formal ontology, is highly successful and extensively used. Although GO is technically a nomenclature rather than an ontology [13] [3] it has been successfully used to annotate large volumes of data and consequently support interoperability and integration from heterogeneous data sets. This shows that highly expressive formal ontologies are not required for all Semantic Web applications. It also shows that real world applications often can be developed with very little semantics (Hendler's hypothesis: "little semantics goes a long way"). There is a very good reason as to why semi-formal ontologies are both more abundant and more useful than formal ontologies. The answer lies in the ease with which semi-formal ontologies can be built to a scale that is useful in real-world applications.

Our objective in touting the value of semi-formal ontologies is to prevent research in the Semantic web field from leading straight into the very problems that AI found itself in. We hope to do this by reducing the prevalent emphasis on formal ontologies and pure deductive inference mechanisms. The reader should note however that we do not completely discount the value of the same. We liken the current research direction in the Semantic Web community to, attempting to construct a new building using the flawed building blocks that lead to the downfall of previous building attempts. Our reasoning behind this is that most motivating examples described in this field pay little or no attention to the fundamental (read hard) problems of entity/relationship identification and ambiguity resolution. Database researchers working on schema integration are only too familiar with the problems relating to ambiguity resolution. According to [17], most scenarios described for potential applications of the Semantic Web trivialize these fundamentally hard problems while emphasizing the trivial problems. Our views coincide with those expressed in [13, 4]. In [4] the Semantic Web community is urged to not waste their efforts on "fixing the plumbing" (referring to infrastructure issues) and to focus their efforts on the more fundamental problems.

4 Semagix Freedom: An example of state of the art Semantic Technology

Let us briefly describe a state of the art commercial technology and product that is built upon the key perspectives we presented above. *Semagix Freedom* exploits task and domain ontologies that are populated with relevant facts in all key functions: automatic classification of content, ontology-driven metadata extraction, and support for complex query processing involving metadata and ontology for all three types of semantic applications identified in Section 2. It provides tools that enable automation in every step in the content chain - specifically ontology design, content aggregation, knowledge aggregation and creation, metadata extraction, content tagging and querying of content and knowledge. Scalability, supported by a high degree of automation and high performance based on main memory based query processing has been of critical importance in building this commercial technology and product. Figure 2 below shows the architecture of *Semagix Freedom*.

[3]"For data annotation, in principle not a full fledged ontology as described above is required but only a controlled vocabulary since the main purpose is to provide constant and unique reference points." [13]

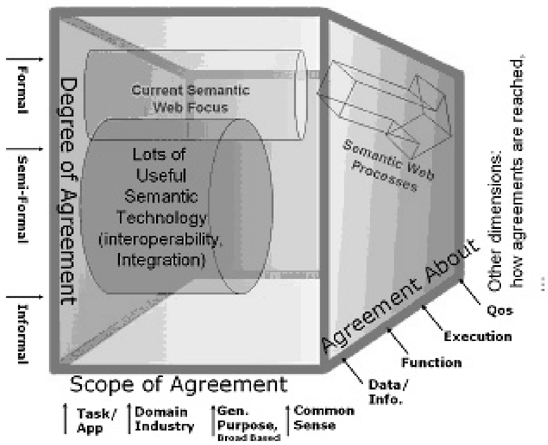

Figure 1: Dimensions along which ontologies vary

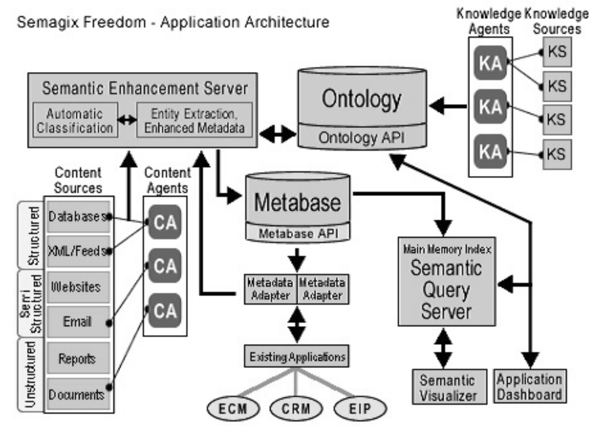

Figure 2: Semagix Freedom Architecture

Freedom provides a modeling tool to design the ontology schema based on application requirements. The domain specific information architecture is dynamically updated to reflect changes in the environment, and is easy to configure and maintain. The Freedom ontology is populated with knowledge–any factual, real-world information about a domain in the form of entities, relationships, attributes and constraints. The ontology is automatically maintained by Knowledge Agents Figure 2, top right). These are software agents created without programming that traverse trusted knowledge sources that may be heterogeneous, but either semi-structured or structured (i..e, concept extraction from plain text to populate ontology is currently not supported but may be supported in future). Knowledge Agents exploit structure to extract useful entities and relationships for populating the ontology automatically. Once created, they can be scheduled to automatically keept the ontology up-to-date with respect to changes in the knowledge sources. Semantic ambiguity resolution (between entity instances) is one of the most important capabilities associated with this activity, as well as with the metadata extraction. Ontology can be exported in RDF/RDFS barring some constraints that cannot be presented in RDF/RDFS.

Freedom also aggregates structured, semi-structured and unstructured content from any source and format. Two forms of content processing are supported: automatic classification and automatic metadata extraction. Automatic classification utilizes a classifier committee based on statistical, learning, and knowledgebase classifiers. Metadata extraction involves named entity identification and semantic disambiguation to extract syntactic and contextually relevant semantic metadata (Figure 2, left). Custom meta-tags, driven by business requirements, can be defined at a schema level. Much like Knowledge Agents, Content Agents are software agents created without programming using an extensive toolkit. Incoming content is further "enhanced" by passing it through the Semantic Enhancement Server [8]. The Semantic Enhancement Server can identify relevant document features such as currencies, dates, etc., perform entity disambiguation, tag the metadata with relevant knowledge (i.e, the instances within the ontology) and produce a semantically annotated content (that references relevant nodes in the ontology) or a tagged output of metadata. Automatic classification aid metadata extraction and enhancement by providing context needed to apply the relevant portion of a large ontology.

The Metabase stores both semantic and syntactic metadata related to content. It stores content in a relational database as well as a main-memory checkpoint. At any point in time, a snapshot of the Metabase (index) resides in main memory (RAM), so that retrieval of assets is accelerated using the Semantic Query Server. This index is both incremental (to keep up with new metadata acquisition) and distributed (i.e., layered over multiple processors, to scale with number of contents and size of the Metabase). The Semantic Query Server is a main memory-based front-end query server. The Semantic Enhancement and Query Servers enable semantic applications (or agents) to query Metabase and ontology using http and Java-based APIs, returning results in XML with published DTDs. This ability, with the context provided by ontology and ambiguity resolution, form

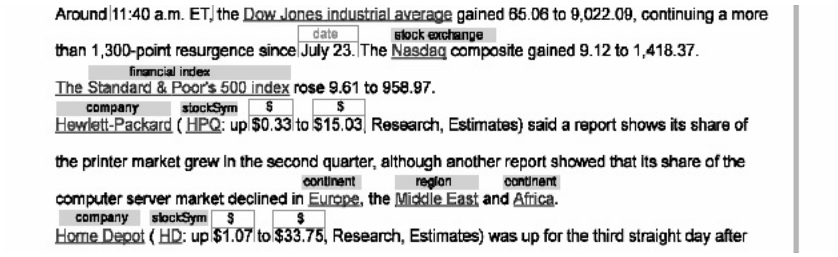

Figure 3: An example of Automatic Semantic Metadata Extraction/Annotation

the basis for contextual, complex, and high performance query processing, providing highly relevant content to the semantic applications. We end the review of Freedom by summarizing some of its characteristics:

- Typical size of an ontology schema for a domain or task ontology: 10's of (entity) classes, 10's of relationships, few hundred property types
- Average size of ontology population (number of instances): over a million of named entities
- Number of instances that can be extracted and stored in a day (before human curation, if needed): up to a million per server per day
- Number of text documents that can be processed for automatic metadata extraction per server per day: hundreds of thousands to a million
- Performance for search engine type keyword queries: well over 10 million queries per hour with approx. 10ms per query for 64 concurrent users
- Query processing requirement observed in an analytical application: approx. 20 complex queries (involving both Ontology and Metabase) to display a page with analysis, taking a total of 1/3 second for computation (roughly equivalent to 50+ queries over an RDB with response time over 50 seconds).

5 Conclusion

Formal ontologies in description logic based representation; supported by deductive inference mechanisms may not be the primary (and certainly not the only) means of addressing major challenges in realizing the Semantic Web vision. The database community should realize that the Semantic Web vision is not one of solving the AI problem, or OWL with subsumption based inference mechanisms. Instead, it can make critical contribution to the Semantic Web by drawing upon its past work and further research on topics such as supporting processing of heterogeneous data/content, semantic ambiguity resolution, complex query processing involving metadata and knowledge represented in semi-formal ontology, and ability to scale with large amount of structured and semi-structured information. In supporting this view point, we provided an overview of one instance of the commercial technology that has been used to develop a broad variety of real world semantic applications. We also provided high level information on scalability requirements observed in supporting these applications. Alternative strategies to realize the vision of the Semantic Web will need to show they will need to scale and perform at least as well as what todays commercial technologies (such as the one briefly discussed in this article) do, and probably well beyond that.

Acknowledgements: Comments to a draft from Dean Allemang, Kemafor Anyanwu, Vipul Kashyap, Harumi Kuno and Kevin Wilkinson are gratefully acknowledged. We also received editing assistance and comments from members of the LSDIS lab including Chris Halaschek, Christopher Thomas, Meenakshi Nagarajan and Kunal Verma.

References

[1] A. Sheth et al. Semantic association identification and knowledge discovery for national security applications. *Journal of Database Management*, 2004 (to appear).

[2] F. Baader and U. Sattler. An Overview of Tableau Algorithms for Description Logics. *Studia Logica*, 69:5-40, 2001.

[3] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, May 2001.

[4] M. Brodie. The Long and Winding Road To Industrial Strength Semantic Web Services, 2003. Keynote Talk, *I*SWC 2003, http://iswc2003.semanticweb.org/brodie.pdf.

[5] Gómez-Pérez and Manzano-Macho. A survey of ontology learning methods and techniques, 2003. http://ontoweb.aifb.uni-karlsruhe.de/Members/ruben/Deliverable%201.5.

[6] T. Gruber. It is what it does: The pragmatics of ontology. Invited talk at *Sharing the Knowledge-International CIDOC CRM Symposium*, Mar. 2003. http://tomgruber.org/writing/cidoc-ontology.htm.

[7] R. Guha, R. McCool, and E. Miller. Semantic search. In *The Twelfth International World Wide Web Conference*, Budapest, Hungary, May 2003.

[8] B. Hammond, A. Sheth, and K. Kochut. Semantic Enhancement Engine: A Modular Document Enhancement Platform for Semantic Applications over Heterogeneous Content. In *Real World Semantic Web Applications*, pages 29-49. IOS Press, Dec. 2002.

[9] IBM. WebFountain. http://www-1.ibm.com/mediumbusiness/venture_development/emerging/wf.html.

[10] D. Nardi and R. J. Brachman. An introduction to description logics. *The Description Logic Handbook*, pages 5-44, 2002.

[11] I. Polikoff and D. Allemang. Semantic technology. *TopQuadrant Technology Briefing v1.1*, Sept. 2003. http://www.topquadrant.com/documents/TQ03_Semantic_Technology_Briefing.PDF.

[12] S. Dill et al. SemTag and SemSeeker: Bootstrapping the Semantic Web via automated semantic annotation. In *Proceedings of the 12th International WWW Conference (WWW 2003)*, Budapest, Hungary, May 2003.

[13] S. Schulze-Kremer. Ontologies for molecular biology and bioinformatics. *In Silico Biology*, 2(17), 2002.

[14] Semagix, Inc. Anti-money laundering. http://www.semagix.com/solutions_ciras.html.

[15] A. Sheth, C. Bertram, D. Avant, B. Hammond, K. Kochut, and Y. Warke. Semantic content management for enterprises and the web. *IEEE Internet Computing*, pages 80-87, July/August 2002.

[16] A. Sheth and R. Meersman. Amicalola Report: Database and Information Systems Research Challenges and Opportunities in Semantic Web and Enterprises. *ACM SIGMOD Record*, 31(4):98-106, Dec. 2002. http://lsdis.cs.uga.edu/SemNSF/.

[17] C. Shirky. The Semantic Web, Syllogism, and Worldview. *Networks, Economics, and Culture*, Nov. 2003.

[18] J. Townley. The streaming search engine that reads your mind. *Streaming Media World*, Aug. 2000. http://smw.internet.com/gen/reviews/searchassociation/index.html.