2005

From Semantic Search & Integration to Analytics

Amit P. Sheth
Wright State University - Main Campus, amit@sc.edu

From Semantic Search & Integration to Analytics

Amit Sheth[1,2]

[1]LSDIS lab, University of Georgia, 415 Graduate Studies Research Center,
Athens, GA 30602-7404
amit@cs.uga.edu
[2]Semagix Inc., 297 Prince Avenue,
Athens, GA 30601
amit.sheth@semagix.com

Abstract. Semantics is seen as the key ingredient in the next phase of the Web infrastructure as well as the next generation of enterprise content management. Ontology is the centerpiece of the most prevalent semantic technologies and provides the basis of representing, acquiring, and utilizing knowledge. With the availability of several commercial products and many research tools, specifications and increasing adoption of Semantic Web standards such as RDF for metadata and OWL for ontology representation, ontology-driven techniques and systems have already enabled a new generation of industry strength semantic applications. In particular, Semagix's Freedom has powered applications in leading verticals such as, financial services, government & intelligence, pharmaceuticals, and media & entertainment. In this paper, we portray some of the requirements of high-end enterprise applications requiring search to integration, and more advanced analytical capabilities, discuss the enterprise scale capabilities expected of a semantic technology, and how Semagix has put an ontology-driven approach to use.

1 Introduction

Semantics is arguably the single most important ingredient in propelling the Web to its next phase. Semantics is considered to be the best framework to deal with the heterogeneity, massive scale, and dynamic nature of the resources on the Web and within enterprises. Issues pertaining to semantics have been addressed in many fields like linguistics, knowledge representation, artificial intelligence, information systems and database management. A Semantic Technology involves application of techniques that support and exploit semantics of information (as opposed to syntax and structure/schematic issues [11]) to enhance existing information systems [8]. Recently, a particular view of the web has evolved, coined the Semantic Web. The Semantic Web is best defined as ``an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation" [1]. In more practical terms, at least currently, Semantic Web technology implies adoption and use of standards such as RDF/RDFS for metadata representation, and OWL for ontology representation.

Dagstuhl Seminar Proceedings 04391
Semantic Interoperability and Integration
http://drops.dagstuhl.de/opus/volltexte/2005/46

2 Examples of Semantic Applications

There a number of ways to classify applications built using ontologies and the Semantic (Web) Technologies, (e.g., see [8] and [2]). For the purpose of our discussion, we take a simpler classification of search, integration and analytics. This classification helps us study commercial semantic applications in terms of increasing complexity and deeper role of semantics:

- Semantic search and contextual browsing:
  - In Taalee (now Semagix) Semantic Search Engine [16], the ontology consisted of general interest areas with several major categories (News, Sports, Business, Entertainment, etc.) and over 16 subcategories (Baseball, Basketball, etc in Sports). Blended Semantic Browsing and Querying (BSBQ) provided domain specific search (search based on relevant, domain specific attributes) and contextual browsing. The application involved crawling/extracting audio, video and text content from well over 250 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 [5].

Fig. 1. Equity Analyst Workbench demonstration semantic integration of heterogeneous, multimodal content

- Semantic integration:
  - In Equity Analyst Workbench (Shown in Fig. 1) [10], A/V and text content from tens of sites and NewsML feeds aggregated from 90+ international sources (such as News agencies of various countries) were continuously classified into a small taxonomy, and domain specific metadata was automatically extracted (after one time effort to semi-automatically create a source-specific extractor agent). 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 populated with over 14.5 million instances (e.g., semantically disambiguated names of artists, track names, etc). The application provided integrated access to heterogeneous content in the company's extensive media holding while addressing semantic heterogeneity.

- Analytics and Knowledge Discovery:
  - In the Passenger Threat Assessment application for national/homeland security [14] and Semagix's Anti-money Laundering solution (see Fig. 2) [9], the ontology is populated from many public, licensed and proprietary knowledge sources, and is kept up-to-date with changes in knowledge source on a daily basis. The resulting ontology has over one million instances. Periodic or continuous metadata extraction from tens of heterogeneous sources (150 files formats, HTML, XML feeds, dynamic Web sites, relational databases, etc) is also performed [6]. 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 [7] application has demonstrated scalability by extracting metadata (albeit somewhat limited types of metadata with a significantly smaller ontology) on the Web scale from well over 2.5 billion pages [3].

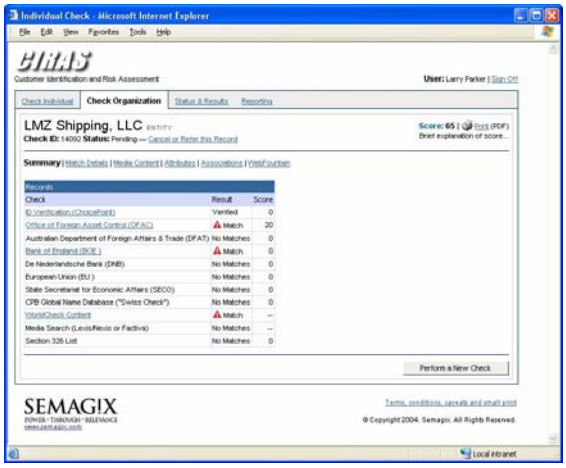

Fig. 2. CIRAS customer risk-assessment tool

3 Observations from Semantic Applications

Based on our experience in building the above real-world applications, we now review some empirical observations. We have found that applications validate the importance of ontology in the current semantic approaches. Ontology captures shared knowledge by representing a part of the domain or the real-world 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 the "semantic normalization" necessary for semantic integration. Our observations break down as follows:

- **Ontology Depth, Expressiveness**
  - Many real-world ontologies may be described as semi-formal ontologies (as opposed to formal ontologies). Semi-formal ontologies are ontologies that may be populated with partial or incomplete knowledge, may contain occasional inconsistencies, or occasionally violate constraints (e.g. all schema level constraints may not be observed in the knowledgebase that instantiates the ontology schema). Such situations are unavoidable when the ontology is populated by many persons or by extracting and integrating knowledge from multiple sources (also see [4]). A good analogy is "dirty data" which is usually a fact of life in most enterprise databases.
  - Formal or semi-formal ontologies represented in very expressive languages (compared to moderately expressive ones) have, in practice, yielded little value in some real-world applications. One reason for this may be that it is often very difficult to capture the knowledge that uses the more expressive constructs of a representation language. This difficulty is especially apparent when trying to populate an ontology using 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 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 the so-called Hendler's hypothesis ("little semantics goes a long way"). On the other hand, we have seen applications, especially in scientific domains such as biology, where more expressive languages are needed, and even OWL is not adequate (for this paper, we will not discuss these applications).
  - As we go from less demanding search/browsing/personalization to more demanding integration/portal applications to even more demanding analytical/business intelligence/knowledge discovery applications, there is a greater need for deeper (domain and task specific) semantic metadata. Also needed is a processing and application logic shift from entities/concepts to relationships. Query processing requirements become increasingly demanding for analytical applications. For example, a typical analytical application involved approx. 20 complex queries (over both ontology and metadata) to display a page with analysis but required sub-second response

time for computation (this roughly equivalent to 50+ queries over a relational database with response time over 50 seconds).

- **Ontology Scope**
  - Currently, most paying customers are interested in developing Enterprise applications in the sense that the scope of the agreement (ontological commitment) may be enterprise-wide, even though the data/content involved in the application may involve a combination of proprietary data within enterprise, subscribed/syndicated content and open source (Web) content. Even two customers within the same industry for seemingly similar applications have different views, resulting in similar but still not the same ontology. For example, industry-sector-analyst classification and instance data can vary between two brokerage houses. While broad industry wide ontologies and knowledge bases typically involve strong social processes involving years of committee efforts, typical ontologies for Enterprise applications are narrow, domain or task/application ontologies (e.g., initial effort may focus on ontology for anti-money laundering, rather than entire financial services domain) that require strong tools that IT professionals and domain experts can use to design the ontology schema and populate the ontology from a few high quality knowledge sources. Sometimes ontologies may be bootstrapped using existing database schemas and industry wide metadata standards but involve substantial modeling efforts.
- **Ontology Size and Knowledge/Metadata Extraction**
  - Ontology population is critical. Among the ontologies developed by Semagix or using its technology, a median size of ontology is over 1 million instances/facts (and about 20% have exceeded 10 million instances). This level of knowledge makes the system very powerful. Furthermore, in many cases, it is necessary to keep these ontologies current or updated with facts and knowledge on a daily or more frequent basis. Both the scale and freshness requirements dictate that populating ontologies with instance data needs to be automated.
  - Large scale metadata extraction and semantic annotation is possible. IBM Web fountain [7] related technology [3] demonstrates the ability to annotate on a Web scale (i.e., over 2.5 billion pages), while Semagix Freedom related technology [6] demonstrates capabilities that work for a few million documents per day per server. However, the general trade-off of depth versus scale applies. For example, more scalable metadata extraction is typically done for extracting simpler types of metadata (i.e., for the metadata at the lower part of metadata pyramid), while deeper or more semantic extraction is typically attempted on Enterprise scale (rather than Web scale). Storage and manipulation of metadata for millions to hundreds of millions of content items requires the best applications of known database techniques with the challenge of improving upon them for performance and scale in presence of more complex structures.

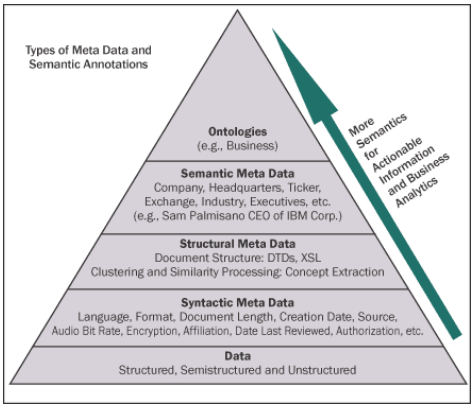

Fig. 3. Metadata Pyramid (From Syntax to Semantics), from [11]

- **Semantic Operations**
  - A vast majority of the Semantic (Web) applications that have been developed or envisioned rely on three crucial capabilities: ontology creation, semantic annotation and querying/inferencing. Enterprise-scale applications share many requirements in these three respects with pan Web applications. All these capabilities must scale to many millions of documents and concepts (rather than hundreds to thousands) for current applications, and applications requiring billions of documents and concepts have also been discussed (esp. in intelligence and government space) but not yet deployed.
  - Two of the most fundamental "semantic" techniques are "named entity identification", and "semantic ambiguity resolution". Without good solutions to these, none of the applications listed will be of any practical use. For example, a tool for annotation is of little value if it does not support ambiguity resolution. Both require highly multidisciplinary approaches, borrowing for NLP/lexical analysis, statistical and IR techniques and possibly machine learning techniques. A high degree of automation is possible in meeting many real-world semantic disambiguation requirements, although pathological cases will always exist and complete automation is unlikely. In a recent ontology population effort using Semagix tools, 97% of ambiguities were resolved automatically [15].
  - Support for heterogeneous content is key – it is too hard to deploy separate products within a single enterprise to deal with structured, semi-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 RSS, 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, enterprise's own relational databases and content management systems such as Documentum or Notes, e-mails, etc).

Database researchers have long studied the issue of integrating heterogeneous data, and many of the techniques to deal with semantic heterogeneity come in handy, especially at the schema levels, but a broader array of techniques are required (including statistical, lexical/NLP, and machine learning) to deal with instance level heterogeneity. And while the enterprise no longer wishes to be divided between separate worlds of structured and unstructured data management, the middle ground of semi-structured data (XML-based data and RDF based metadata) is growing at an explosive rate.

- 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 to which an article refers, regardless of format of the article. Analytical applications could require sub-second response time for tens of concurrent complex queries over a large metadata base and ontology, and can benefit from further database research. High performance and highly scalable query processing techniques that deal with more complex representations compared to database schemas and with more explicit roles of relationships, is important. Database researchers can also contribute to the strategies of dealing with large RDF stores.

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. Freedom exploits populated ontologies as part of its comprehensive ontology-driven process for supporting all three types of semantic applications identified in Section 2. Among the key capabilities supported in this process include automatic classification of content, semantic metadata extraction, support for complex query processing involving metadata and ontology, business rule specification and other techniques for analytical processing, Furthermore, it deals with a broad variety content formats and types spanning structured to unstructured data, and the corresponding challenges of dealing with syntactic, structural and semantic heterogeneity. It provides tools and a comprehensive set of APIs which enable automation in every step in the semantic application building process - 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. Fig. 4 shows the architecture of Semagix Freedom.

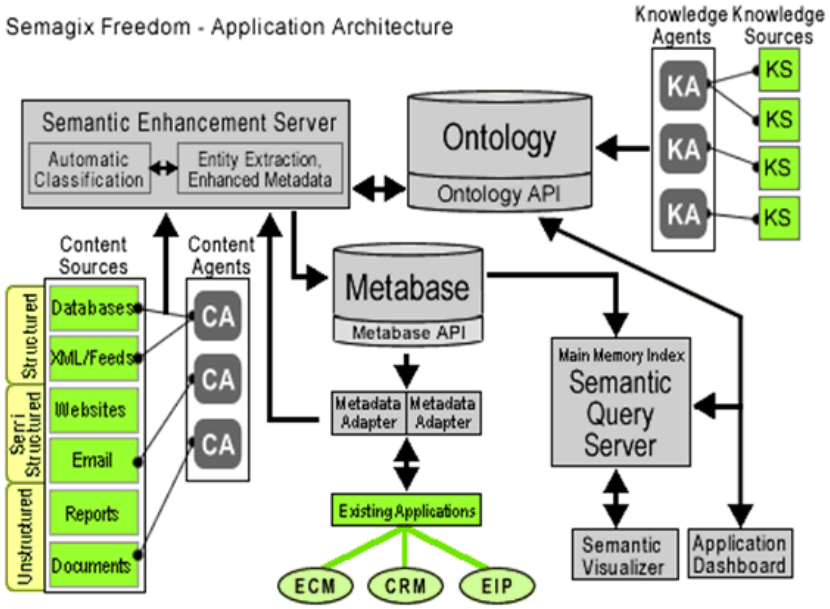

Fig. 4. Semagix Freedom Architecture

4.1 Semagix Freedom System Architecture

The centerpiece of the Freedom architecture is the domain/application ontology. This domain specific information architecture is dynamically updated to reflect changes in the environment, and it is easy to configure and maintain. The Freedom ontology is populated with knowledge, which is any factual, real–world information about a domain in the form of entities, relationships, attributes and certain constraints. The ontology is automatically maintained by Knowledge Agents (Fig. 4, 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 keep the ontology up-to-date with respect to changes in the knowledge sources. Semantic ambiguity resolution (is the entity instance the same or related to an existing entity instance? Is this the "John Doe" Board Member the same as the "John Doe" CEO in the ontology?) is one of the most important capabilities associated with this activity, as well as with the metadata

extraction. The 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 (Fig. 4, 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 [6]. 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 aids 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 into 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 provide semantic applications (or agents) ability to query the 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 the basis for contextual, complex, and high performance query processing, providing highly relevant content to the semantic applications.

4.2 Freedom in Action: Application Creation

The development of an ontology-driven semantic application in Semagix Freedom can be divided into distinct stages (Fig. 5 presents a simplified lifecycle).

The first stage of the lifecycle is the creation of a schema that serves as the definitional component of the ontology. Typical ontology schemas usually involve tens of classes and relationship types for a given application or domain (although some may be larger, depending on application scope, representation language, etc.). Examples of such applications/domains include anti-money laundering, terrorism, pharmaceutical drug discovery, Glycan structure, etc. The second task in the lifecycle is the population of the ontology at the instance level. Instances of these classes and relationships between these instances, i.e. knowledge, can be considered to be the assertional component of the ontology.

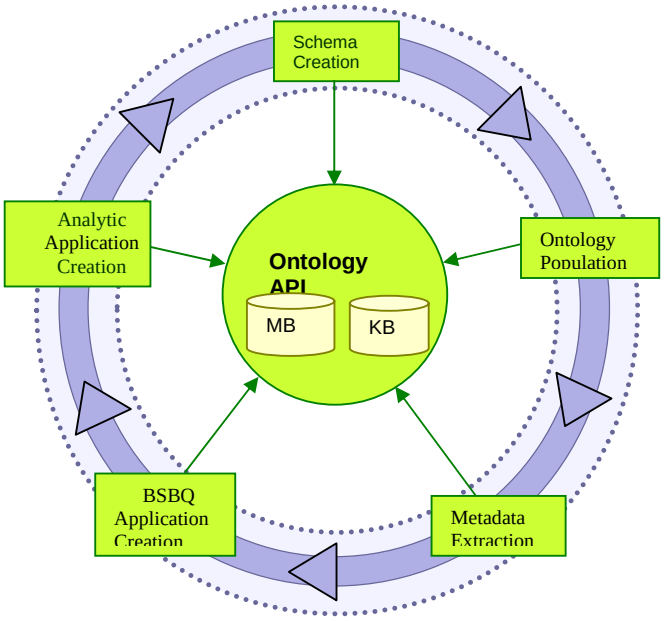

Fig. 5. Semantic Application Lifecycle

The next phase of the lifecycle involves the semantic annotation of heterogeneous (unstructured, semi-structured, and structured) content from a variety of sources. The process of attaching semantic annotation to a document or other piece of content is referred to as metadata extraction. Semantic applications are created by exploiting metadata and ontology with associated knowledgebase. A typical ontology-based system provides APIs to query the metadata and knowledge, and builds the application logic and GUI front end. A relatively simple example is an end-user query interface for semantic search and/or contextual browsing. One powerful, yet intuitive, interface to such a system involves a blend of semantic browsing and querying, also known as Blended Semantic Browsing and Querying (BSBQ). Using this type of interface, a user can seamlessly follow his train of thought to cross-navigate between related knowledge and content. A more advanced alternative for semantic application development could involve the creation of high-end analytical tools used for the creation of complex queries. Next, we provide small samples of various interfaces that correspond to the above; most details are skipped for brevity.

4.2.1 Schema Creation

The development of an ontology-driven application typically starts with the creation of an ontology schema. This schema contains the definition of the various classes, attributes, and relationships that encapsulate the business objects that model a

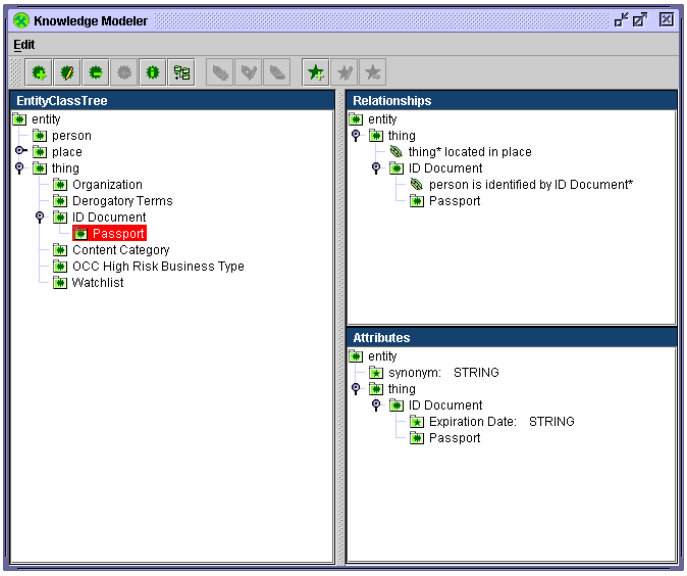

Fig. 6. Knowledge Modeler

particular domain. This model is usually devised with the help of a "domain expert" who has a deep understanding of the real-world objects and concepts in the domain. The "Knowledge Modeler" component of the Semagix Freedom toolkit (Fig. 6) is an example of an interface that allows the user to create an ontology schema.

In the left pane of this tool, a user may define the ontology schema by creating a hierarchical structure of classes (similar to a directory structure). The addition of a new class as a child of an existing class indicates the "is a" relationship. After classes have been created, pairs of classes can be selected and relationships created between them (for example, "drug has side-effect symptom"). Properties of the new relationship such as cardinality may be specified using this interface. The available relationships for a selected class (including its inherited relationships) are shown in the top right panel of the Knowledge Modeler. The user may also select an individual relationship (for example, "person identified by SSN") or add an attribute definition. Attribute definitions are displayed in the bottom right portion of the interface. In addition to the tree-based view provided by the Knowledge Modeler, the ontology schema can also be viewed as a directed graph.

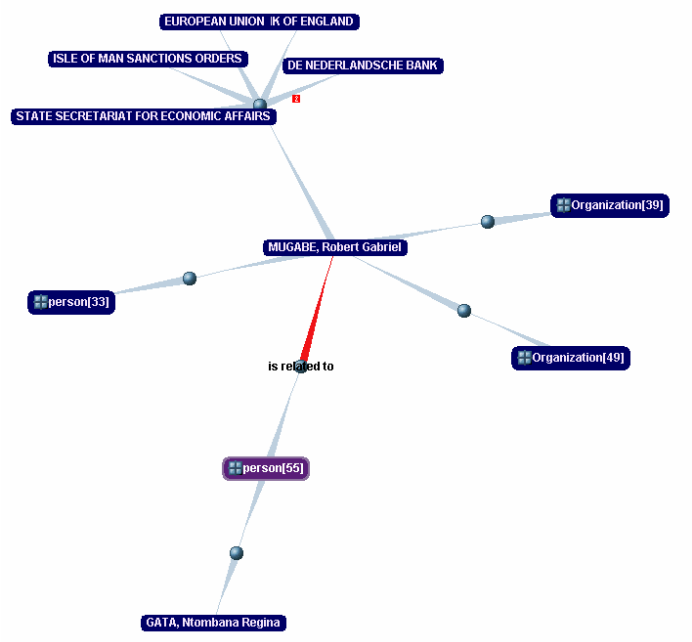

Fig. 7. Graph-based Ontology Instance View

The schema definition for the ontology-based system not only includes the definition of classes, relationships, and attributes (the knowledge model), but also defines a set of document categories with a collection of metadata attributes for each category (the metabase model). The "Metabase Modeler" component of the Freedom toolkit provides an interface for defining this part of the system.

4.2.2 Ontology Population

Once the structure for the ontology has been defined, it can then be populated with instances of classes, attributes, and relationships. The collection of these instances is also referred to as a knowledgebase. A number of problems are inherent in creating real-world applications that depend on a fairly large ontology. One such problem (from a user-interface perspective) is how to enable a user to effectively manage such a large data set. One basic, yet essential tool is an interface for viewing and editing the classification, relationships, and attributes for a single entity. The "Entity Viewer" component of the Freedom toolkit is an example of such an interface. While this tool is effective for detailed information about a single entity, it is not well suited for giving the user a good overall picture of the contents of the knowledgebase. For this purpose, a graph-based view of the ontology is available (shown in Fig. 7).

Using this tool, the user can begin browsing the knowledgebase starting from a particular entity (the "focus" entity). Initially, only entities that are directly related to the focus entity are displayed. The user may then explore the graph in a particular

direction by clicking on one of these related entities and choosing the "expand" option. In this way, the user decides what portion of the knowledgebase is relevant and obtains a better understanding of its contents by traversing the relationships between related entities.

Although many tools for viewing directed graphs have been created thus far, most become unusable or unintelligible when applied to real-world information. For example, it is a common occurrence to have many entities, perhaps thousands, related to a single entity via the same relationship (consider the relationship "ticker symbol traded on stock exchange", for example). To handle this scenario, the concept of a synthetic "collection node" was introduced. On our example, the collection node would "contain" the thousands of ticker symbol entities related to a single exchange entity. A single collection node would be related to the "NASDAQ" entity node (for example) via the "traded on" relationship. If the user wishes to see a particular member or members of the collection, those entities can be "released" from the collection by allowing the user to select these. The released entities would then be connected to the collection node with the synthetic "contains" relationship (which is not a part of the ontology schema to begin with).

4.2.3 Metadata Extraction

The next step in the development of an ontology-driven, semantic application often involves enhancing unstructured content (documents) with semantically relevant metadata. In Freedom, content extractor agents along with software modules called "experts" are used to perform this enhancement. The content agent first retrieves the textual contents of the document from a given source. If the category (domain) of the document is not known a priori, it may be automatically determined using a classifier committee technique. Given the domain of the document, the expert then attempts to find entities that are explicitly mentioned in the text of the document. Fig. 8 illustrates the detection of entities and other phrases within a piece of unstructured text. Once this set of entities is determined, a new set of inferred entities can be derived. For example, it may be inferred that a document belonging to the "business" category containing the text "MSFT" is actually about the entity "Microsoft". The expert then adds both the explicit and implicitly detected entities to a document metadata container, thus performing metadata extraction.

Fig. 8. Entity and Phrase Detection and Extraction

4.2.4 BSBQ Application Creation

After creating a body of semantically annotated documents (a metabase) as well as a set of inter-related ontology instances (a knowledgebase), it is now possible to create an application that will make use of both. Typical Internet users are familiar with two techniques of Web "travel" – browsing and querying. Browsing, via hyperlinks, allows users to navigate between documents that refer to each other; while searching (via Google, for example) "teleports" a user to an individual document. When creating a semantic application, we can combine these two techniques into one to create an intuitive, yet powerful query tool. This hybrid technique is referred to as Blended Semantic Browsing and Querying, or BSBQ. An example BSBQ application is displayed in Fig. 9.

Using the toolbar at the top of the application, the user may search for the name of a specific entity in the knowledgebase. If a matching entity is found, it is displayed in the right side of the application as a directed graph. Users may view related knowledge by expanding the graph from a selected node. Each time an entity node in the graph is selected, the attribute details are displayed in a table on the top left side of the screen. In addition to attributes, all semantically relevant documents (as produced by the content extractor agents) are displayed in the list at the bottom left of the application.

Double-clicking a document in the list displays its content and relevant semantic metadata in a popup window (shown on the right of Fig. 9). From this window, the

user may select any one of the metadata items and "refocus" the graph onto that item, seamlessly following his train of thought to cross-navigate between content and knowledge.

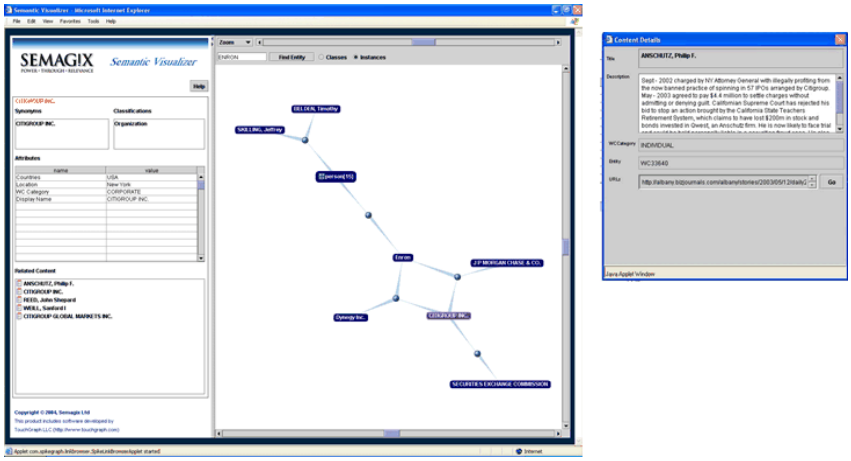

Fig. 9. The Semantic Visualizer BSBQ application

4.2.5 Analytical Tools

In addition to general-purpose BSBQ applications, analytical applications can be designed to provide advanced users an interface for performing ontology-specific computation as well as formulating complex queries. The CIRAS (Customer Identification and Risk Assessment Solution) Anti-Money Laundering application developed by Semagix is an example of one such application that adds an additional layer of business logic and computation to an existing semantic framework. The CIRAS analytical tools are built upon a knowledgebase containing interrelated "people" and "organization watchlist" entities possessing attributes like "address", "date of birth", etc. This application also uses a metabase of hundreds of thousands of documents containing information about individuals involved in various types of illegal activities. The CIRAS application then applies a series of rule-based heuristics over Freedom's Ontology API to compute a "risk score" for a given individual or organization. For example, an individual who "works for" an organization that "appears on" a government-maintained watchlist would receive a higher risk score.

5 Summary

In this paper we have demonstrated the central role of ontology in contemporary Semantic (Web) Technologies and their real-world applications. As we go to more demanding applications from search to integration and analytics, we also observed the role of expressiveness/depth, scope, population and the associated capabilities for metadata extraction, complex query processing involving both ontology and metadata,

rule processing, visual interfaces, etc. We also showed how a variety of semantic applications are created using a state of the art commercial product, Semagix Freedom.

Acknowledgements: This article is based in part on [13] and [12]. I also appreciate Matthew Perry's help in preparing this article.

References

1. Berners-Lee, T., Hendler, J., and Lassila, O.: 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.
2. Brown, R.: Applications of Ontologies, http://www.ontologyengineering.com/ApplicationsOfOntologies.htm
3. Dill, S., et. al.: SemTag and SemSeeker: Bootstrapping the Semantic Web via automated semantic annotation. Proceedings of the 12th International WWW Conference (WWW 2003), Budapest, Hungary, May 2003.
4. Gruber, T.: It Is What It Does: The Pragmatics of Ontology, invited talk at Sharing the Knowledge- International CIDOC CRM Symposium, March 26-27, Washington, DC, http://tomgruber.org/writing/cidoc-ontology.htm
5. Guha, R., McCool, R., and Miller, E.: Semantic Search, The Twelfth International World Wide Web Conference, Budapest Hungary, May 2003
6. Hammond, B., Sheth, A., and Kochut, K.: Semantic Enhancement Engine: A Modular Document Enhancement Platform for Semantic Applications over Heterogeneous Content, in Real World Semantic Web Applications, V. Kashyap and L. Shklar, Eds., IOS Press, December 2002, pp. 29-49
7. IBM WebFountain, http://www-1.ibm.com/mediumbusiness/venture_development/emerging/wf.html
8. Polikoff, I., and Allemang, D.: "Semantic Technology," TopQuadrant Technology Briefing v1.1, September 2003. http://www.topquadrant.com/documents/TQ03_Semantic_Technology_Briefing.PDF
9. Semagix-CIRAS Anti-Money Laundering, Semagix, Inc. http://www.semagix.com/solutions_ciras.html
10. Sheth, A., Bertram, C., Avant, D., Hammond, B., Kochut, K., and Warke, Y.: Managing Semantic Content for the Web, IEEE Internet Computing, July/August 2002, pp. 80-87.
11. Sheth, A.: Semantic Meta Data For Enterprise Information Integration, DM Review, July 2003. http://dmreview.com/article_sub.cfm?articleId=6962
12. Sheth, A., and Avant, D.: Semantic Visualization: Interfaces for exploring and exploiting ontology, knowledgebase, heterogeneous content and complex relationships, NASA Virtual Iron Bird Workshop, March 30-April 1, 2004. http://ic.arc.nasa.gov/vib/day2/index.php
13. Sheth, A., and Ramakrishnan, C.: Semantic (Web) Technology In Action: Ontology Driven Information Systems for Search, Integration and Analysis. In IEEE Data Engineering Bulletin, Special issue on Making the Semantic Web Real, December 2003. http://wwwt.semagix.com/documents/SemanticWebTechinAction.pdf
14. Sheth, A., et al.: Semantic Association Identification and Knowledge Discovery for National Security Applications, Journal of Database Management, 2004 (to appear).
15. SWETO: Semantic Web Technology Evaluation Ontology, http://lsdis.cs.uga.edu/Projects/SemDis/sweto/

16. Townley, J.: The Streaming Search Engine That Reads Your Mind, Streaming Media World, August 10, 2000. http://smw.internet.com/gen/reviews/searchassociation/index.html