SEMPL: A Semantic Portal

Matt Perry
LSDIS Lab, Computer Science,
The University of Georgia
415 Boyd Graduate Studies Research Center
Athens, GA, 30602-7404
(706) 542-2911
mperry@cs.uga.edu

Eric Stiles
LSDIS Lab, Computer Science,
The University of Georgia
415 Boyd Graduate Studies Research Center
Athens, GA, 30602-7404
(706) 542-2911
stiles@cs.uga.edu

ABSTRACT

Semantic Web technology is intended for the retrieval, collection, and analysis of meaningful data with significant automation afforded by machine understandability of data [1]. As one illustration of semantic web technology in action, we present SEMPL, a semantic web portal for the Large Scale Distributed Information Systems lab (LSDIS) at the University of Georgia. SEMPL, which is powered by a state of the art commercial system, Semagix Freedom [7], uses an ontology-driven approach to provide semantic browsing, linking, and contextual querying of content within the portal. By using the ontology based information integration technique, SEMPL can specify the context of a particular piece of research information, annotate web pages, and provide links to semantically related areas enabling a rich contextual retrieval of information.

Categories & Subject Descriptors: H.3.5
[**Information Storage and Retrieval**]: Online Information Services – *Commercial services, Data sharing, Web-based services.*

General Terms: Management, Performance, Design.

Keywords: Semantic Portal, Semantic Web, Ontology.

1. INTRODUCTION

Semantic technologies empower three classes of enterprise and web applications involving heterogeneous content [8]: (a) semantic search and browsing, (b) semantic integration, and (c) knowledge discovery and semantic analytics. Techniques and capabilities that comprise the state of the art in semantic technology include ontology design and population, automatic classification, semantic metadata extraction and annotation, and semantic query processing and/or inferencing.

A Semantic Portal utilizes semantic technology to provide semantic search, browsing, and integration of content. Early portals such as Ontoweb [9] and MindSwap [2] have utilized ontology representation standards devised as part of the Semantic Web activity. Although the vision for Semantic Web involves machine understandability of content and a very high degree of automation [1], functionality of early efforts have been limited, constrained by limited population of ontology and limited use of computation (querying or inferencing) over the more expressive representation theoretically possible with an ontology representation.

In this student-authored poster based on a Semantic Web course project, we discuss a Semantic Portal that better demonstrates the power of semantics in information search, browsing, integration, and presentation. We focus on complex, ontological relationships rather than simple, direct relationships and utilize the commercial system Semagix Freedom to expand the scale and reach of our implementation. Freedom allows for powerful, automated ontology population, and its Semantic Enhancement Engine (SEE) [3] allows us to place any document in the context of our domain.

2. PORTAL FEATURES

Essentially, we leverage semantics in our portal for enhanced browsing, searching, and linking of relevant content. We provide ontology-based browsing and searching of information within the portal and implement novel features such as links to semantically related entities and a semantically enhanced web search

2.1 Ontology-based Browsing and Searching

SEMPL relies on a domain-specific ontology covering academia based computer science research for browsing and searching content within the portal. Since the ontology schema and its inherent taxonomy clearly organize research papers, people, presentations, etc., it can provide a coherent presentation and method of access to the heterogeneous content of heterogeneous format within the portal.

For a visual interface with the ontology schema, SEMPL uses TouchGraph [10]. The upper third of the portal web page is used to display a subset of the ontology schema. Clicking a node in the schema can have two different actions depending on the user mode (browse or search). In *semantic browsing mode*, all instances of the selected class are listed in the bottom third of the page. A click on one of the listed instances will open its *entity-view* page. In *semantic search mode* the same click will open a search template in the bottom third of the page allowing the user to search based on class-specific attributes.

2.2 Links to Semantically Related Entities

To truly have a semantic application, one must utilize the (named) relationships between entities, as these relationships provide the most insight. We wanted to capture the power of these relationships to provide otherwise unknown information about a particular entity, mainly which other entities are related to this entity in the context of our domain. Given the large-scale capabilities of Freedom, especially its ability to populate an ontology with large amounts of instance data or knowledge, it was necessary to keep the implementation simple and efficient. We chose not to add the complexities of rule sets and logics to

discover knowledge, as is discussed in SEAL [4]. Instead, we approached the problem from a graph-traversal perspective. Since an instance of a named relationship is essentially a path connecting two vertices, a complex semantic relationship can be thought of as a path of length more than one which connects two distinct vertices (entities).

The ontology schema and instance data each have a clearly defined role. We use the schema to define a complex semantic relationship. The administrator can configure any number of complex relationships for a given entity class by selecting the paths at the schema level. At the entity instance level, we use instances of these paths to provide links to semantically related entities. Our design provides the ability to take a given entity and efficiently traverse several nodes to find semantically related entity instances. Also, a simple ranking mechanism based on the number of distinct relationship instances connecting two entities is used. This is fundamentally different from what has been done with earlier semantic portal implementations as we go beyond simple, direct relationships and utilize the real power of our ontology in a general and efficient way.

2.3 Semantically Enhanced Web Search

SEMPL's web search component allows searching (through a MetaCrawler-implemented [5], keyword-based web search) and enhancement (through Freedom's SEE) of external content with respect to the given knowledgebase. In the search result pages, recognized entity instances are highlighted and embedded with links to their corresponding *entity-view* pages. Figure 1 demonstrates an enhanced web search result. After clicking the term "RDF," an *entity-view* page presents all relevant portal content for the entity "RDF."
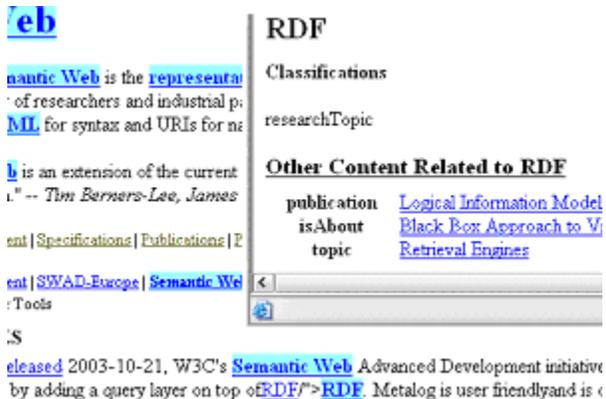
Figure 1. Semantic Enhancement

3. SYSTEM ARCHITECTURE

The modularity of SEMPL's architecture is designed to easily integrate various semantic technologies while allowing for scalable growth. The SEMPL code is developed in three distinct layers: a backend data store, the SEMPL middle layer, and the SEMPL web interface. All communication between the web interface and the data store APIs is through SEMPL's middle layer API.

Even though this project is guided by research, the data store is the commercial semantic software Semagix Freedom. Freedom is designed as a backend tool to be accessed with an application-specific GUI interface. Freedom provides automated ontology population through knowledge extractors and allows for semantic enhancement of documents.

4. CONCLUSIONS

SEMPL addresses three key issues for semantic portals: large scale ontology population, efficient utilization of complex semantic relationships, and enhancement and contextualization of non-annotated web documents. Freedom's automated knowledge extractors do not rely on current knowledge representation languages (i.e., RDF [6]), making large scale ontology population feasible. SEMPL utilizes complex semantic relationships to link relevant entity instances. Finally, SEMPL places any document in the context of its given knowledgebase through its web search component. Our prototype is being hardened for roll out of the next generation of the LSDIS lab web site.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Berners-Lee, T., Hendler, J., and Lassila, O. "The Semantic Web". Scientific American, May 2001

[2] Golbeck, J., Grove, M., Parsia, B., Kalyanpur, A., Hendler, J. New Tools for the Semantic Web. in Proceedings of EKAW '02 (Siguenza Spain, Oct 2002).

[3] Hammond, B., Sheth, A., Kochut, K. Semantic Enhancement Engine: A Modular Document Enhancement Platform for Semantic Applications over Heterogeneous Content. in Real World Semantic Web, Kashyap, V., and Shklar, L., (eds.), IOS Press, 12/2002.

[4] Maedche, A., Staab, S., Stojanovic, N., Studer, R., and Sure, Y. SEAL – A Framework for Developing SEmantic Web PortALs. LNCS-Springer, vol. 2097, 2001.

[5] Metacrawler. http://www.metacrawler.com.

[6] Resource Description Framework (RDF) Model and Syntax Specification. http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/

[7] Semagix, inc. http://www.semagix.com.

[8] Sheth, A., and Ramakrishnan, C. Semantic (Web) Technology in Action: Ontology Driven Information Systems for Search, Integration and Analysis. IEEE Data Engineering Bulletin, Special issue on Making the Semantic Web Real, (12/2003).

[9] Spyns, P., Oberle, D., Volz, R., Zheng, J., Jarrar, M., Sure, Y., Studer, R., Meersman, R. OntoWeb – a Semantic Web Community Portal. in Proceedings of PAKM '02 (Vienna Austria, Dec 2002).

[10] TouchGraph. http://www.touchgraph.com.