

SEMANTIC WEB AND ITS TOOLS

Deepak Ahuja

Ph.D Research Scholar

ABSTRACT

Semantic Web will allow for significantly more machine-readable content to be available on the World Wide Web. Getting this content onto the web, and using it once it is there, requires new “metaphors” for working with Semantic Web data. This paper will describe the “Semantic Web Portal” an approach to using Semantic Web content, and some freely available tools that we are developing to make it a reality.

I INTRODUCTION

The Semantic Web is based on making machine-readable content available on the World Wide Web, and designing the appropriate technologies to harness it. Currently, a number of tools developed for traditional artificial intelligence work are being adopted to the Semantic Web. However, these tools are primarily the products of traditional AI research which have been transitioned to use on the World Wide Web. As such, they are very powerful tools, but only focus on some parts of the “lifecycle” of Semantic Web information. These tools are motivated by the idea of a “Semantic Web Portal” which provides a mechanism for tying together many Semantic Web components. In this paper, we first describe the idea of a Semantic Web Portal, and then describe some of the tools we are developing to make this vision a reality. These tools are available for download at the MIND SWAP web page.

1.1 Site for the semantic Web

A particular focus of our group is the creation of Semantic Web Portal technology that will motivate researchers and students in many areas to add semantic markup to documents, images and data. Authors will be able to link their evolving web resources to terms from multiple ontologies (or to define terms that extend the ontological coverage). As these links are added, queries are made to various web back-ends that contain similar pointers from other documents, databases, image archives, etc. The results are displayed to the user, allowing a constant, dynamical web portal to be created. This portal contains pointers to documents that are on similar topics, databases that can answer queries about conceptually related science, and images and other multimedia resources.

This includes links to related photos she can use in her documents, to database queries that can show recent results, and to other documents she might want to cite or link to. By providing useful information and resources, users will be encouraged to mark up their documents so that they make take advantage of the portal. What allows this system to work more fully is the integration of the markup process with the portal. The portal

provides the most advantage to users while they are creating their own semantic web documents. Thus, not only does the portal provide information, but also it is able to create more links based on the user. If she chooses to link to certain terms provided by the portal, a semantic link is created between the two documents.

Research being done at MIND SWAP to implement such a system includes the development of inference engines that can find relationships between entities that are not explicitly stated, the development of backend “triple stores” that can integrate database and knowledge-base processing, and the development of presentation technologies that can present the information in the portals in a way that is appropriate to the needs of the specific user. In addition, we are developing several tools to make it easier to develop Semantic Web content from existing web sources. In the next section, we discuss some tools being developed in MIND SWAP aimed at the eventual creation of a Semantic Web Portal system.

II DEVELOPED TOOLS OF MIND SWAP

They developed two tools for generating DAML and RDF from formatted documents: Convert To RDF which works with delimited files, and the Web Scraper which looks at formatted HTML pages. For generating content from scratch, there are two more tools. The RDF Editor provides a variety of features to aide users in creating RDF in tandem with HTML documents. The RDF Instance Creator (RIC) provides a simple interface for creating RDF for other media, such as pictures. Finally, the PARKA ontology manager works with triples to provide a fast interface for searching and finding relationships among data.

2.1 RDF Editor

The RDF editor provides users with the ability to create Semantic Web markup, using information from multiple ontologies, while they simultaneously create HTML documents. The aim of this software is as follows:

- To provide the user with a flexible environment in which he can create his web page without mark up hindrances;
- To allow the user to semantically classify his data set for annotation and generate mark-up with minimal knowledge of RDF terms and syntax.
- To provide a reference to existing ontologies on the Internet in order to use more precise references in his own web page/text.
- To ensure accurate and complete RDF mark up with scope to make modifications easily.
- To allow extension to ontological concepts by the user, thus creating new ontological content.

To achieve these ends, the application has three functional parts.

1. HTML Editor with Preview Browser – This is a standard WYSIWYG editor for creating and deploying web pages. Users can write some HTML from scratch, or use the editor to add images and create content in a more natural way.
2. Ontology Browser - A particularly innovative feature of the RDF Editor is that it encourages users to work with multiple ontologies. Many existing tools allow users to create their own ontologies for use in RDF documents. This tool encourages users to work with and extend pre-existing ontologies, exploiting the distributed nature of the Semantic Web.

This interface allows the user to browse through existing ontologies on the Internet with the aim of finding relevant terms and properties. Once the appropriate ontology has been located, the user can add it to the local database, and the properties of the ontology are automatically added to the Local Ontology Information where it can be managed.

3. Semantic Data Trees – This part of the interface is what allows users to classify the data semantically into one of four basic elements: Class, Object, Property and Value.

As an interface to the Semantic Web Portal, the RDF Editor is ideal. As users select classes from ontologies, the portal can return results to them in a separate window. The fetched data is then immediately available for reference or incorporation to the current document. When the user publishes their document, the portal can include all of the new references in its knowledge base.

2.2 RDF Instance Creator

The RDF Instance Creator, is a tool designed to ease the process of creating markup, particularly for non-text sources. RIC allows the user to generate RDF simply by filling a series of forms, thus freeing the user from needing to know RDF while still providing all the benefits that it has to offer.

RIC can use any valid ontology that is currently accessible through the Internet. After importing an ontology, the user is presented with a list of available classes from which they can create objects. When defining an object, its properties appear in the workspace. This provides a simple form interface where the user enters data for each of the object's properties.

Resources that cannot be described, let alone searched, in the current web framework suddenly become available and accessible to users who may be interested in them.

2.3 Scraper

Some web pages have regular structure with labelled fields, lists and tables. Often, an analyst can map these structures to ontology and write a program to translate a portion of the web page into the semantic mark up language. The RDF Web Scraper is a tool that helps users specifies how to extract RDF mark up from these kinds of web pages.

Users analyze the HTML in a page and create a wrapper that describes how the tag structure relates to the contents. The scraper parses the page based on the wrapper, and generates a table of data. The user can then indicate ontological specifications for each column of the table and generate the corresponding RDF.

This application has the ability to take information from between tags as well as from within them. This allows users to scrape the URI's of images or links and mark them up.

2.4 Convert to RDF

Convert to RDF is a tool that takes delimited data, from spreadsheets such as Microsoft Excel or from databases, and generates mark up based on the column headers.

2.5 Parka-DB

The other systems described in this paper have been front-end tools for users generating content. Parka allows the user to define a frame-based knowledge base with class, subclass, and property links used to encode the ontology. Property values can themselves be frames, or alternatively can be string, numeric values, or specialized data structures (used primarily in the implementation). The Parka language allows exceptions, in the form of multiple-inheritance, and provides extremely efficient algorithms for performing inheritance using a true inferential-distance-ordering calculation.

Parka can effectively compute recognition, and handle extremely complex "structure matching" queries -- a class of conjunctive queries relating a set of variable and constraints and unifying these against the larger KB. Parka uses DBMS technologies to support inference and data management.

The structure of the Parka system meshes nicely with the triple structure of DAML and RDF. RDF instances are easily converted into Parka assertions and loaded into the database. At that point, it is possible to extract information about relationships within the data that would not be accessible otherwise. Parka's inferencing mechanisms are then used to take advantage of the ontological information defined in DAML.

III CONCLUSION

The Semantic Web requires new tools that can be used in new ways. One important use will be the semantic web portal, allowing people to dynamically create and use Semantic Web information. Building such an application will need a number of new technologies, and we describe some tools aimed at providing this basis. Thus, the tools described in this paper are examples of some of the basic technologies that are needed to create this new portal technology. These include tools for generating Semantic Web instances from structured sources and from HTML pages, a tool for creating marked up pages easily, a tool for creating instance data easily, especially for non-text sources and a back-end ontology management tool.

REFERENCES

1. Berners-Lee, T. and M. Fischetti, "Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web" by its Inventor, Harper, San Francisco, 1999.
2. Berners-Lee, T., Hendler, J. and Lassila, O. "The Semantic Web," Scientific American, May, 2001
3. Brickley, D and R.V. Guha, "Resource Description Framework (RDF) Model and Syntax Specification", W3C Recommendation submitted 22 February 1999, <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/> (current May 2002).
4. Connolly, D, van Harmelen, F., Horrocks, I, McGuinness, D., Patel-Schneider, P., and Stein, L. DAML+OIL (March 2001) Reference Description, W3C Note 18 December 2001 (<http://www.w3.org/TR/daml+oil-reference>).
5. Hendler, Jim, "Agents and the Semantic Web," IEEE Intelligent Systems. March/April 2001 (Vol. 16, 2).
6. Horrocks, I. Et al – OilED, available on the WWW at <http://img.cs.man.ac.uk/oil/>

7. M. A. Musen, R. W. Ferguson, W. E. Grosso, N. F. Noy, M. Crubezy, & J. H. Gennari. "*Component-Based Support for Building Knowledge-Acquisition Systems*". In Conference on Intelligent Information Processing (IIP 2000) of the International Federation for Information Processing World Computer Congress (WCC 2000), Beijing, 2000.
8. McGuinness, D. "*Conceptual Modeling for Distributed Ontology Environments.*" Proceedings of the Eighth International Conference on Conceptual Structures Logical, Linguistic, and Computational Issues (ICCS 2000). Darmstadt, Germany. August 14-18, 2000.
9. K. Stoffel, M. Taylor, J. Hendler. "*Efficient Management of Very Large Ontologies.*" In Proceedings of American Association for Artificial Intelligence Conference (AAAI-97), AAAI/MIT Press 1997.
10. Y. Sure, M. Erdmann, J. Angele, S. Staab, R. Studer, D. Wenke. OntoEdit: "*Collaborative Ontology Development for the Semantic Web.*" In Proceedings of the 1st International Semantic Web Conference - ISWC2002, Springer, LNCS.

IJARSE