SMORE – Semantic Markup, Ontology, and RDF Editor

Aditya Kalyanpur
University of Maryland, College Park
A. V. Williams Building
College Park, Maryland 20742
301-314-6642

aditkal@yahoo.com

James Hendler
University of Maryland, College Park
MIND Lab, 8400 Baltimore Ave
College Park, Maryland 20740
301-314-6642

hendler@cs.umd.edu

Bijan Parsia
University of Maryland, College Park
MIND Lab, 8400 Baltimore Ave
College Park, Maryland 20740
301-314-6642

bparsia@isr.umd.edu

Jennifer Golbeck
University of Maryland, College Park
A. V. Williams Building
College Park, Maryland 20742
301-314-6642

golbeck@cs.umd.edu

ABSTRACT

The promise of the Semantic Web is founded on the principle that online content will be semantically annotated, creating machine-understandable content using interlinking ontologies. In keeping with this principle, we introduce SMORE, the Semantic Markup, Ontology, and RDF Editor. It provides users with an integrated environment for creating web pages, email, and other online content while facilitating inline, seamless semantic markup.

The rich features of SMORE extend its capabilities beyond that of other annotation tools available. For instance, in addition to combining content creation and annotation, SMORE allows users to mark up parts of images using SVG. Users also have a number of options to collect information from the web, including an advanced ontology search capability, web scraping, and a semantic Virtual portal that provides links to semantically related material. This, combined with the unique ability to defer markup using place holders, use and extend multiple ontologies, infer classification for ad hoc objects, and interlink concepts makes SMORE a unique tool that will benefit both users and the future of the semantic web.

Categories and Subject Descriptors

I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods – *Representation languages*

General Terms

Management, Design, Human Factors

Keywords

Editor, Semantic Web, RDF

1 INTRODUCTION: RATIONALE AND GOAL

One commonly articulated design constraint for the Semantic Web [1,2] is that "Anyone must be able to say anything about anything." Often, this constraint is invoked to support semantics for Semantic Web markup languages having a certain breath of expressivity so that there as few constraints as possible on what can be said (about what). It remains an open question whether given these expressive languages "anyone" can use them effectively. Most people are not ontological engineers, domain experts, or logicians, or even programmers, so its unlikely that they will be able to read, sort through, and grasp how to apply large ontologies, much less construct their own. More to the point, few will bother when they just want to get their web page up, or send that next email, or put a caption on a particularly striking photo. Aside from the difficulty of learning how to model content in a reasonably correct and formal way, current Web focused knowledge engineering tends to involve either an interruption of normal workflow and techniques

1. REPORT DATE **2006**	2. REPORT TYPE	3. DATES COVERED **00-00-2006 to 00-00-2006**

4. TITLE AND SUBTITLE **SMORE -Semantic Markup, Ontology, and RDF Editor**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Maryland,8400 Baltimore Avenue,College Park,MD,20742**	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES **5**	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**			

(e.g., switching to an RDF [3] editor to create RDF content which is then linked to an HTML page) or a wholesale abandonment of prior practice. In other words, the author is forced into a two-step situation where either "the author must *first* create the content and *second* annotate the content"(Authoring and Annotation of Web Pages in CREAM [8]) or they must create all of their content in a knowledge creation context and then render it to HTML in some fashion.

While there are many tools for easing ontology creation and knowledge acquisition, few focus on how normal Web authors work. For instance, Protégé-2000 [6] is a strong Ontology editor with a straightforward, forms based knowledge acquisition system. Using it, however, is much more like entering information in a database, albeit an extraordinarily flexible one than producing a Web page. OILEd [7] and OntoEdit [8] are strongly biased to ontology creation and editing, and there exist numerous RDF Editors such as RDFedt [10] and RIC [11] that allow users to build complex RDF documents. These are fine if your focus is creating an ontology or RDF document, but tend to encourage a two (or more) step processes where content generation and Semantic markup are rigidly distinct.

Figure 1. The SMORE interface

The latest version of Ont-O-Mat [9], on the other hand, addresses at least some of the "integration with normal Web behavior" issues, trying to "hide the border between authoring and annotation as far as possible". While we endorse this move, we suspect that Ont-O-Mat is still too deeply rooted in the traditional knowledge acquisition mindset. For example, their first principal requirement is: "Consistency: Semantic structures should adhere to a given ontology in order to allow for better sharing of knowledge." [*Authoring and Annotation of Web Pages in CREAM*]. In support of this requirement, Ont-O-Mat has many components and modes which focus on ontology driven markup. This tends to reintroduce the impulse to set up the "right" ontologies in advance. This seems contrary to letting "anyone say anything about anything", or, perhaps, it simply raises the burden of generating Semantic Web content to an inhibitory level.

In this paper, we present SMORE (Semantic Markup, Ontology and RDF Editor), a tool whose design is driven by the idea that much Semantic Web based knowledge acquisition will look more like Web page authoring than traditional knowledge engineering. Like Onto-O-Mat, SMORE blurs the line between normal content creation and Semantic annotation, but SMORE also supports *ad hoc* ontology use, modification, combination, and extension.

2 ADDING SEMANTICS: NEEDS AND METHODS

In keeping with the main design principle mentioned earlier--seamless integration of content creation and annotation--SMORE provides built-in support for performing routine web-oriented tasks in the context of semantic markup. For instance, SMORE contains a fully featured *WYSIWYG text/html editor* that allows users to create and deploy web pages. Besides providing standard features for web page design, the editor facilitates the generation of semantic markup by acting as a medium through which the user can compose semantic triples of his data. Users can select portions of text from the web page and insert them into triple placeholders (that follow the standard *subject-predicate-object* model). A crucial point to note here is that the markup obtained from these triples is inaccurate since the triples are composed of plain natural language based textual data without containing any specific ontological references. SMORE leaves open the option of when to perform the task of linking established ontological elements to the user-defined terms. Thus, if users have a pre-determined set of ontologies to work with, they could insert terms from these ontologies directly into a triple using the *Triple Specification Window,* or alternately, they could defer the process of finding the "right" ontology to a later stage. The issue of deferral is discussed later in the paper (Section 3.1).

Another example of how the workflow in SMORE supports the main design principle is the functioning of the *MailSMORE* module that allows users to compose and send e-mails with context-based semantic markup. The importance of semantic mail in today's world is best illustrated by the mass usage of e-mails making tasks such as searching, sorting, filtering, and blocking of SPAM invaluable. *MailSMORE* facilitates semantic mail creation by semi-automating the process of triples creation based on the standardized structure of e-mails. Thus, users compose e-mails normally and *MailSMORE* uses standard e-mail attributes (*subject, to, from, body* etc.) as placeholders to create triples, which are linked to an associated *e-mail ontology*. Moreover, users can specify additional triples pertaining to the body of the message (as is done using the *html editor* described earlier) and link these triples to any external ontology (such as an agenda ontology, for instance, when the

user is sending his agenda via e-mail). The triples set can be converted to RDF and sent as an attachment with the mail, allowing external agents to process this for use in a variety of applications.
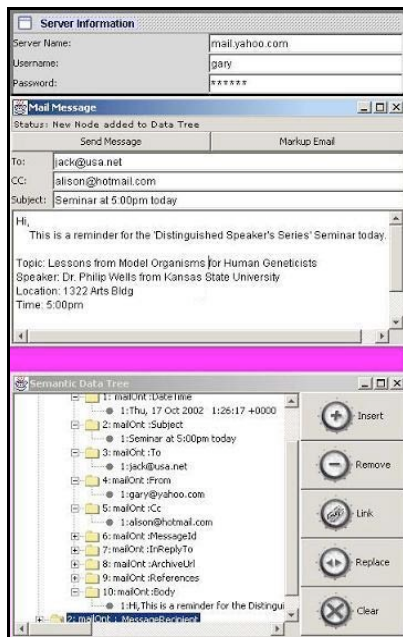

Figure 2. MailSMORE

A final example of the realization of the main design principle of SMORE is *PhotoSMORE*, a module that allows users to view images in a semantic context. *PhotoSMORE* supports the annotation of specific areas of an image with RDF, a concept illustrated in Jim Ley's SVG-based image markup tool [13]. The expressivity of SVG [12] allows the notion of unique *paths* to represent different parts of an image, and *PhotoSMORE* facilitates integration of these *paths* in semantic triples in order to generate a substantial amount of metadata from a single image source. Thus, for instance, users viewing an image of a group of people can separately highlight individual persons (and their respective features) and describe them semantically, using terms from multiple external ontologies (unlike [13]). Annotated images created in this manner can lead to many interesting applications such as the co-depiction experiment at RDF Web [14].
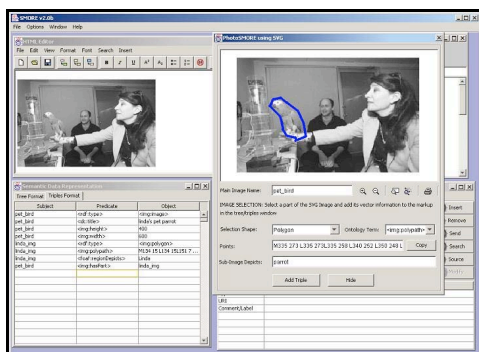
Figure 3. PhotoSMORE

Semantic triples created as a by-product of using any of the built-in toolkits mentioned above can be stored in the *Data Classification Window*, which features an additional visualization perspective - a data tree. The tree format supports operations such as node replacement (with existing ontological elements) and node linkage (specifying equivalence between nodes), enabling conversion of the tree to a complex semantic graph, the notion of which is fundamental in allowing users to see and understand the relationships among their metadata.

The user-defined semantic triples need to be referenced with established ontological elements in order to generate accurate markup. As noted in the introduction, one of the key design initiatives of SMORE is to ensure ad hoc ontology use, modification and extension during this referencing and subsequent markup process. SMORE provides the user with an elaborate *Ontology Management* interface in order to achieve these tasks. It includes a built-in web browser and a *Semantic Virtual Portal* (discussed in section 4.2) to aid the user in searching for existing ontologies, a database for storing established web ontologies locally (in addition to providing support for ontology creation, described in section 3.2) and an associated local-ontology search engine to help users select and link to relevant concepts in various ontologies (also discussed in section 3.2). An important side-note is the need for an ontology information table associated with the database. In order to determine the "right" ontology to use (in terms of relevance and expressivity), users must get as much conceptual information as possible about existing ontological elements. This information available from the original URI of the ontology is displayed in this table for user reference.

3 IMPORTANT ISSUES AND CONSIDERATIONS IN THE SEMANTIC MARKUP PROCESS

3.1 Deferral
With regard to semantic triple composition, it is important to note that while the subject and object of a triple need not have ontological references, the predicates need to be established ontological elements (properties). Hence, SMORE helps the user distinguish between user-defined predicates and properties from an ontology by prefixing user-defined predicates with an asterisk (*). This acts as a reminder for the user to replace them with established properties at some point in the markup process to ensure accurate markup of the document. The deferral of the task of associating elements with ontologies is important since the user can delay the process of finding the "right" ontology until sufficient contextual information is available.

3.2 Multiple Ontology Use and subsequent Implications

As noted earlier, the user may not have a pre-determined set of ontologies that he wishes to use for markup; in this case, the fully interactive nature of the tool allows the user to search for relevant ontologies dynamically, as a side step to the data classification process. In order to aid the user in choosing the "right" ontology, SMORE contains an advanced search engine that can scan the local ontology database for specific classes and properties. In this search, the user provides parameters to extend or restrict the search domain. For instance, the user can search for ontologies having a particular class and/or a particular property; or filter properties having a specific domain and/or range; or extend the result of class search using equivalence relations between classes. The presence of such an advanced ontology search engine is critical. Not only does it help the user save time and effort while marking up the document (by finding an established ontology), but, more importantly, it facilitates linking between documents, a key component in the functioning of the semantic web.

Alternatively, the user can create a DAML [4]-based ontology by defining new terms and/or borrowing elements from various other web ontologies [5]. Terms from this user-defined ontology can be directly inserted into the triples dataset, for conversion to RDF markup. This notion of using and creating ontologies 'inline' is central to the working of SMORE. It also illustrates the conceptual difference between linking to an external ontology (which could signify full trust in the external ontology and its creator) and merging two or more ontologies (which could signify trust in specific ontological terms only).

3.3 Manifested Inferences

An interesting feature of SMORE is the auto-manifestation of inferences based on the semantic markup specified by the user. When the user creates a semantic triple and replaces one of its components with an established ontological term, additional triples inferred from this association are subsequently added to the dataset. For instance, given the triple *Michael Jordan-plays-Basketball*, if we replace the user-defined predicate *plays* by a specific ontological property 'plays' that has a domain of class *Athlete* and a range of class *Sport,* two additional triples can be inferred, namely, *Michael Jordan* is an instance of class *Athlete,* and *Basketball* is an instance of class *Sport*. SMORE directly adds these triples to the dataset thus making it obvious to the user. Furthermore, an interesting point to note is that the user can delete any of these inferred triples, thereby implying that he or she doesn't adhere to these added claims, and the tool then adds the necessary semantic markup to ensure external reasoning agents don't make the same inferences.

4 ADVANCED FEATURES OF SMORE

4.1 Screen Scraper

Often, users browse information in regularly structured web pages that have labeled fields, lists and tables (e.g. Yahoo People). The *Screen Scraper* in SMORE is used to extract semantic markup from these kinds of web pages by allowing users to map the structures to an ontology and translate a portion of the web page into the semantic markup language. The resultant markup can be added to the user's knowledge base. In this manner, the *Screen Scraper* can act as a source of extra metadata generation (from external means). Alternatively, users composing web pages with structured formats can use this tool to facilitate the creation of markup.

An interesting feature of the scraper is its 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. For example, if a faculty list html document contains pictures of each faculty member, the scraper can grab the URI's of those pictures and include markup that indicates who is pictured in the image.

4.2 Semantic Virtual Portal

The Semantic Virtual Portal in SMORE is a dynamic source of rich contextual data. As users edit their pages in SMORE, the portal can be used to return pages with similar markup, related images and data, or references to other material. The underlying concept here is to store links to ontological elements made by a user while marking up his data, and later use these links as pointers to provide other users referencing the same or equivalent ontological elements with that data. The presence of this portal motivates the semantic markup of documents, images and other data with the aim that someone else can access and use this information dynamically. Additionally, it can be used to retrieve related ontologies defined by other users, which can be inserted into the local ontology database and subsequently used in markup.

For example, if a scientist authoring a paper or web page uses a particular term from an online ontology, the semantic web portal will return other sources with similar markup. 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.

5 CONCLUSION

SMORE embodies the underlying design principles stated in the paper by providing a seamless integration

of content creation and annotation. It facilitates the semantic markup of various types of media (photos, html, e-mail) and in doing so provides a high degree of flexibility in the use, modification and extension of ontologies, all of which can be done ad hoc. It integrates a wealth of features into one software package, and introduces many new features that are unavailable anywhere else such as the semantic virtual portal that helps users find related data. Thus, by creating an easy to use and highly useful tool for creating markup, we believe that everyday users will be more likely to use and benefit from the semantic web.

6 ACKNOWLEDGMENTS

7 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] M. A. Musen, R. W. Fergerson, 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.

[7] Horrocks, I. Et al – OilED, available on the WWW at http://img.cs.man.ac.uk/oil/

[8] Sure, Y Et Al, "OntoEdit: Collaborative Ontology Development for the Semantic Web". *Proceedings of the first International Semantic Web Conference 2002 (ISWC 2002), June 9-12 2002, Sardinia, Italia*

[9] Handschuh, S, Staab, S, Mädche, A. "CREAM - Creating relational metadata with a component-based, ontology driven annotation framework" *K-CAP 2001*

[10] Jan Winkler, RDFedt, available at http://www.jan-winkler.de/dev/e_rdfe.htm

[11] Michael Grove, RDF Instance Creator, available at http://www.mindswap.org/~mhgrove/RIC/RIC.sh tml

[12] Scalable Vector Graphics (SVG) 1.0 Specification W3C Recommendation *04 September 2001*

[13] Jim Ley, SVG-based Image Annotation Tool, available at: http://www.jibbering.com/svg/AnnotateImage.ht ml

[14] Co-depiction experiment at RDFWeb: http://rdfweb.org/2002/01/photo/