Visualizing An Historical Semantic Web with Heml

Bruce G. Robertson
Dept. of Classics, Mount Allison University
402 Hart Hall, 63D York St.
Sackville, NB, Canada E4L 1G9
brobertson@mta.ca

ABSTRACT

This poster presents ongoing efforts to enrich the RDF-based semantic Web with the tools of the Historical Event Markup and Linking Project (Heml). An experimental RDF vocabulary for Heml data is illustrated, as well as its use in storing and querying encoded historical events. Finally, the practical use of Heml-RDF is illustrated with a toolkit for the Piggy Bank semantic browser plugin.

Categories and Subject Descriptors: J.5 [Computer Applications]: Arts and Humanities

General Terms: Experimentation

Keywords: ACM proceedings, Heml, RDF, History, Chronology

1. INTRODUCTION TO HEML

Since 2001 the Historical Event Markup and Linking Project (Heml) has explored the use of W3C markup technologies to encode and visualize historical events on the web.[1] It has sought to provide a reasonably lightweight markup language suitable for encoding events and their associated participants and locations. In addition, the project offers a webapp which renders conforming documents into SVG and HTML visualizations[3, 4].

The project provides these tools in the form of Java code and XSLT style sheets integrated as a component for the Apache Group's Cocoon web development framework. All of this is offered under the LGPL; and it may be downloaded from a SVN repository, available at `svn://heml.mta.ca/home/heml/heml_svn`, and from regularly-published releases, now at version 0.7.2.

Heml was originally conceived within the XML paradigm. Though its XML-based tools have proved useful in presenting large-scale research projects and even visualizing emergency messaging[1, 5], the Heml Project's goal is more universal: as we wrote in our original proposal, Heml aims 'to add a historical component to the growing movement for a Semantic Web.' To this end, in the past year, the project has refocused itself around RDF technologies. This poster presents the results of this work and some of the opportunities Heml's visualization tools offer to RDF data.

[1]The project website is at `http://heml.mta.ca`.

WWW 2006, May 23–26, 2006, Edinburgh, Scotland.
ACM 1-59593-332-9/06/0005.

2. THE HEML SCHEMA

The RDF representation used by the Heml webapp is based on the third and most recent schema developed by the project, assigned the namespace `http://www.heml.org/schemas/2003-09-17/heml`. Originally written in W3C XML Schema, it has been translated to Relax-NG and integrated into an XHTML profile language. These resources and their documentation are available at `http://heml.mta.ca/Schemas/2003-09-17/`.

Because this version associates the definitions of locations and persons with the events in which they participate through URI attributes, only a short XSLT stylesheet is required to transform conforming documents into similarly-intentioned RDF. In both cases, the basis of Heml markup is the historical `Event`, which, in its simplest form, associates a text label with an indicator of time and a reference to evidence, and it assigns a URI to the union. In RDF terms, this resource is the root node of the directed graph.

The schema permits a single location to be associated with each event. Locations are provided labels and `geo:lat` and `geo:long` coordinates. Events may be associated with one or more persons, either directly, or in combination with a 'role'. Finally, one or more `KeywordClassification`s, with values such as 'death' or 'battle', are possible properties of events.

2.1 Chronology

Naturally the encoding of time is critical to this project. Heml takes a pragmatic approach to this problem and deploys existing computational tools as they might best serve the purpose of historical indexing. In its schema, temporal primitives include dates recorded in W3C Schema's `DateTime`, `Date` or `gYear` formats. A more complex time primitive is the 'bounded date' which entails the properties `TerminusPostQuem` and `TerminusAnteQuem` in one of the above three formats. Finally all of these four elements can be used as property values of the starting or ending point in a `DateRange`, a construct which expresses a span of time.

3. VISUALIZING RDF WITH THE HEML WEB SERVICE

The Heml webapp transforms conforming data into various representations, both HTML and SVG. However, the most popular representations are the graphical ones: scaled timelines and, when appropriate, historical maps, both animated and rendered as in figure 1. Through our project's REST-based web services, these visualizations are available to conforming data across the web.

Figure 1: Map generated With a SPARQL query

4. ACCESSING HEML DATA IN RDF

To explore the use of RDF for storing and accessing Heml data, we have employed the Redland RDF database and used SPARQL queries to retrieve events in RDF. The results of this process is passed to the visualization web service. Figure 1 shows a map generated through this process. To produce this map, the SPARQL query retrieved events whose labels contained the word 'Henry'. These events included the election and death of U.S. president William Henry Harrison, Spinoza writing Henry Oldenburg, and the escape of the indentured servant, Henry Watkins. The map-drawing routine selects an appropriate SVG base map, plots dots for each location associated with an event and draws the window in which event label text appears. This text is hyperlinked to pop-up HTML which provides links to the web resources encoded as evidence for the event.

4.1 Heml and Piggy Bank

The Heml project aims not merely to produce code that generates historical visualizations, but it also aims through that code to germinate a web of structured data relating to history. In this way it hopes to provide students and scholars with a truly historical means of searching the web, using search the axes of time, place and text. The true extensibility of RDF and the many experiments in data harvesting that go with it make the project's goal closer to reality. Our poster presents a Heml-aware version of Piggy Bank, the semantic browser plugin devised by Simile at MIT[2]. This acts on invisible links and/or xslt stylesheets which export current (XML) Heml data as RDF. Using this, the student calls up a hyperlinked timeline of the events represented in the documents she has read (and whose events she has stored in her personal 'bank'). These might range from tags embedded in an archaeological new bulletin pertaining to early hominids in Europe, through to the Dublin Core publication date tags on an on-line edition of Charles Dickens' *Hard Times*, all the way to current events in Ukraine posted in the notes of a Political Science course. An example is offered in figure 2.

5. FUTURE WORK

The is just the first of many possible applications of the Heml visualization tools. Through an RDF query and transformation system Heml can provide visualizations of other, more complex, RDF vocabularies such as MEMECS (Metadonnes et Mmoire Collective Systmatique)[6]. In fact, *any* labelled RDF resource with associated temporal properties of the appropriate property type could be treated as a Heml event and visualized alongside all others. For instance, just as Piggy Bank now provides 'Calendar' and 'Map' views for its data bank, Heml's tools can offer a more deeply historical 'Timeline' view.

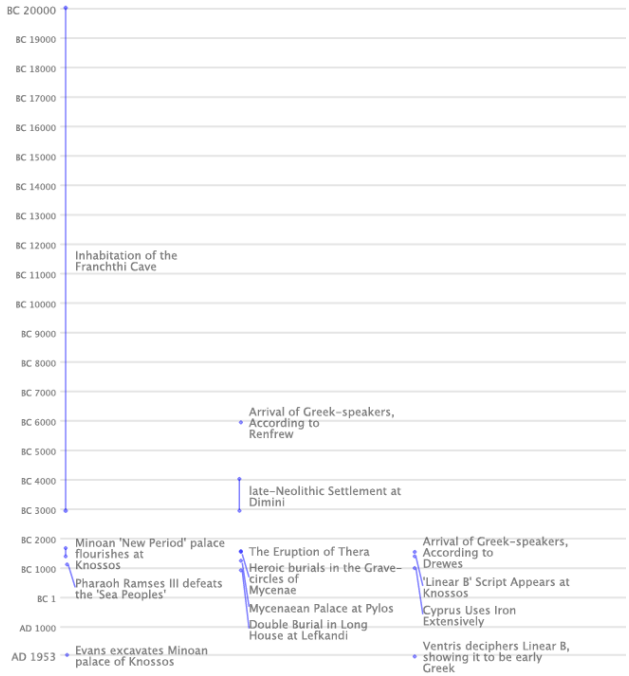

Figure 2: SVG Timeline generated from RDF-encoded data

Similarly, as this project makes its RDF schema more concrete, it is hoped that other RDF vocabularies will find it useful to integrate into their vocabularies. For instance, the FOAF project, `http://www.foaf-project.org`, whose RDF vocabulary describes people and the links between them, might wish to use Heml visualization so that participants can publish Heml events in their life, and thereby provide each other with a brief life histories to compare and explore.

6. REFERENCES

[1] T. Costa, editor. *The Geography of Slavery in Virginia*. University of Virginia, 2005. `http://www.vcdh.virginia.edu/gos/index.html`

[2] D. F. Huynh, S. Mazzocchi, and R. Lee. Piggy Bank. Technical report, MIT and W3C, 2005. `http://simile.mit.edu/piggy-bank/`

[3] B. Robertson. *Dalla Fonte all Rete: Il linguaggio XML e la codifica dei documenti storici, archeologici e archivistici*, An Overview of the Historical Event Markup and Linking Project, pages 37–50. Pisa, 2002.

[4] B. Robertson. Improving Ancient History Online with Heml. *Classics@*, 2, 2004.

[5] B. Robertson and K. Green. Visualizing APRS Messaging with Heml. In *Proceedings of the 24th Digitial Communications Conference*. TAPR and ARRL, 2005.

[6] K. H. Veltman. Cultural and Historical Metadata, MEMECS (Metadonnées et Mémoire Collective Systématique). *Cultivate Interactive*, 1, 2000. `http://www.cultivate-int.org/issue1/memecs/`