EMPLOYING SEMANTIC WEB TECHNOLOGIES TO PROVIDE DYNAMIC CONTENT FOR ELECTRONIC MUSEUM GUIDES

Heinz-Günter Kuper

Informatik in Bildung und Gesellschaft
Institut für Informatik
Humboldt-Universität zu Berlin
10099 Berlin, Germany
kuper@informatik.hu-berlin.de
http://waste.informatik.hu-berlin.de/kuper/

Abstract: *Tim Berners-Lee's vision of the Semantic Web promises to make the targeted retrieval of online information far more accurate. As Semantic Web technologies mature, software systems will be able to deliver detailed information regarding particular artefacts on the fly in a museum environment, thus answering visitor queries that were not anticipated by the content creators. This paper explores some of the issues associated with providing such a solution.*

Keywords: *Semantic Web, RFID, electronic museum guide, faceted browsing.*

1. INTRODUCTION

The research presented in this paper has been conducted under the auspices of the EU/EFRE-funded POSEIDON [1] project, whose main goal is to investigate the potential of radio-frequency identification (RFID) technology in museums. In pursuit of this goal it will be necessary to develop a flexible content management and delivery system in order to accommodate the RFID technology which is at the base of the project.

The Semantic Web and its supporting technologies are particularly attractive in this sense, because by their very nature they permit the management and querying of large amounts of data in a robust manner, leveraging all the benefits associated with linked data, i.e. exposing data that would ordinarily not be revealed using standard search technologies.

Semantic Web technologies can be applied to address the issue of limited knowledge within the context of an exhibition. A museum curator traditionally collates information regarding an exhibition and makes it available to the museum visitor in a static form, either as labels on an exhibit or information presented on an electronic guide or in an accompanying catalogue. It is impossible for the curator to anticipate all the visitors' questions regarding the artefacts in a particular exhibition. Semantic Web technologies will enable a visitor to ask questions not originally anticipated by the

content creators, while affording the content creators a degree of control over the answers returned by the visitor's queries using, for example, the appropriate ontologies to ensure the quality and accuracy of the answers. In other words, the search possibilities are not completely unstructured, but can be tailored to the topic of the exhibition and the returned results focussed on the themes of the exhibition, as envisaged by the curator.

The visitor must be able to search for information without constructing a complicated query, and to this end faceted browsing can be of great use, beginning with a broad query and then drilling down through the masses of information, iteratively refining the search, until the desired information can be presented to the visitor. The dataset used to return results on a query could be maintained by the museum itself, which would limit the scope of the information but improve its reliability, or the Semantic Web at large could be queried.

2. THE DEVELOPMENT OF THE SEMANTIC WEB

The World Wide Web as we know it today, despite its already massive impact on society in general and the scientific community in particular, is little more than an incunabulum, a technology still very much wrapped in swaddling clothes and which only now is beginning to realise its revolutionary potential. Tim Berners-Lee's original proposal [2]

described a system that would not only use hyperlinks, but would also use metadata to enrich the information stored within the system and assist researchers in their queries. The Web has had a long path to travel before Berners-Lee was able re-introduce his concept of the Semantic Web [3].

Essentially, the idea behind the Semantic Web is to enrich data with metadata, thus allowing it to be processed algorithmically by machines, taking some of the eyeball work out of sorting through the large number of mostly irrelevant hits that are returned by the standard keyword search option common to modern search engines. Through the application of Semantic Web technologies, the various data on a Web page can be encapsulated in semantic nuggets which can then be used to find specific information in a broader context.
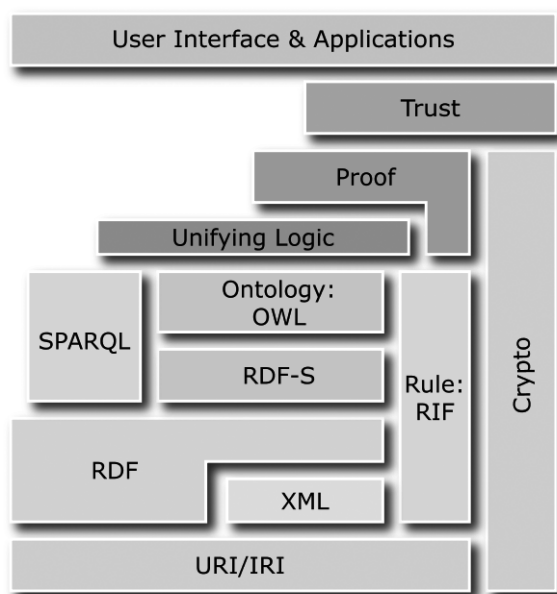

Fig. 1 – The Semantic Web Stack [4]

The Semantic Web is realised by a number of technologies. It would be beyond the scope of this paper to attempt even to outline the field, which currently is populated with a number of competing approaches. Nevertheless a brief excursion into the history of its development might give an indication of the path this technology will take in the future.

The fundamental technology underlying the Semantic Web is the Resource Description Framework (RDF) [5]. Statements in RDF take the form of subject-predicate-object triples, e.g. in Turtle (Terse RDF Triple Language) [6] notation:

```
painting:Mona_Lisa
dc:creator
person:Leonardo_da_Vinci
```

This statement records the fact that Leonardo da Vinci painted the Mona Lisa. Precursors of this format can be found in LISP association lists and in the entity-attribute-value (EAV) data model. The EAV model was developed to cope with the fact that

clinical databases essentially represent sparse matrices, since there are many null values in a conventional relational database designed to record patient data. No patient suffers from every imaginable disease nor has the patient been given every possible test, and new treatments are constantly being developed and new diseases discovered, which would require continual updates to the schema of a relational database. The EAV model allows data to be recorded in simple statements that can then be aggregated to retrieve all the clinical information regarding a particular patient. The EAV model is thus very robust when it comes to continually expanding datasets that are also under constant revision.

In 1995 Ramanathan V. Guha developed the Meta-Content Framework (MCF) while working at Apple [7]. MCF was intended to record metadata associated with Web sites and to be used in conjunction with the Hot Sauce Web navigator, which was to allow the user to navigate through 3D representations of a Web site. Guha moved to Netscape in 1997, where he was inspired to represent MCF in XML, which was then rechristened as the Resource Description Framework.

RDF records the data, while the Web Ontology Language OWL [8] describes classes of resources and the relationships between them and defines various constraints. For example the following OWL snippet states that a museum is a cultural institution that must possess at least one collection:

```
<owl:Class rdf:ID="Museum">
<rdfs:subClassOf
rdf:resource="&institution;Cultural"/>
<rdfs:subClassOf>
<owl:Restriction>
<owl:onProperty
rdf:resource="#Collection"/>
<owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">
1
</owl:minCardinality>
</owl:Restriction>
</rdfs:subClassOf>
...
</owl:Class>
```

OWL is not the only ontology language, but it is a W3C standard and the second version was released in October 2009.

The SPARQL Protocol and RDF Query Language [9] (a recursive acronym) defines a syntax for querying RDF datasets under consideration of the ontologies that apply to that dataset. SPARQL bears similarities to the Structured Query Language SQL. For example, the following SPARQL query over the DBpedia [10] dataset returns the works for which da Vinci is famous:

```
SELECT ?works WHERE {
:Leonardo_da_Vinci
```

```
dbpedia2:works
?works . }
```

By leveraging these technologies a visitor's museum experience will be enriched by the ability to make queries regarding the artefacts and thus to discover information beyond that which was provided statically by the exhibition's curator.

3. THE SEMANTIC WEB IN THE MUSEUM

The overhead involved in implementing Semantic Web technologies in a museum environment is justified by the improved delivery of quality information to the museum visitor, since these technologies support the development of extensive knowledge bases which would permit content to be presented to the visitor in ways not originally anticipated by the content providers.

For example, information regarding the original owner of certain artefacts could be stored as metadata in the knowledge base, i.e. the RDF dataset. It would then be possible for the visitor (or indeed the researcher) to search for all artefacts that originally belonged to a particular individual by querying the knowledge base, a process which in standard hyperlinked pages would involve a keyword search on the individual's name, without being able to specify that the query is to focus on that individual's ownership of items and not merely on any mention of the individual's name on a page.

The ability to query structured information in this way is familiar to anyone who has ever used a standard database. The major advantage of the Semantic Web is that this information can be presented online without tying the user to any particular database management system.

One of the fundamental principles of the Semantic Web is that "Anyone can say Anything about Any topic", referred to as the "AAA Slogan" [11]. In other words, a particular knowledge base could be expanded by merging it with data from another knowledge base. Reference to the appropriate ontologies, written for example in OWL, would ensure the integrity of the merged data.

4. MANAGING AND DELIVERING CONTENT

The are two potential sources of information that could provide content for an electronic museum guide: an internally managed dataset and the Semantic Web itself. Since the fundamental technologies underlying these two sources are the same, one interface can query both sources.

The Semantic MediaWiki [12], which is an extension to the popular and mature MediaWiki [13] software, fulfils the fundamental requirements for a

system to gather the content for the electronic guide. Wiki technology is ubiquitous and it is not unreasonable to assume that most people understand the principles behind the software, if indeed they have not already collaborated using a wiki. The barrier to entry is thus minimised and the content creators are not required to learn a new system, although the Semantic MediaWiki does extend the syntax employed by the standard MediaWiki. Forms may simplify the data entry and relieve some of the cognitive load associated with learning a new syntax.
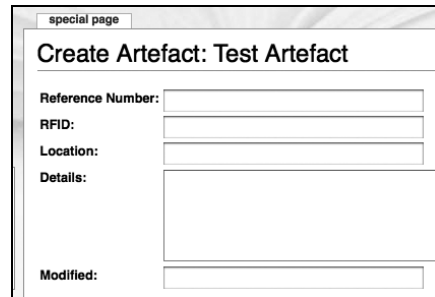

Fig. 2 – Forms can be used to simplify the input of metadata

All data input into a Semantic MediaWiki can be exported into an RDF file, allowing it to be processed by other Semantic Web tools as required. Content gathered in the Semantic MediaWiki will be made available to the visitor on the electronic guide by means of a software that implements a faceted browsing interface.
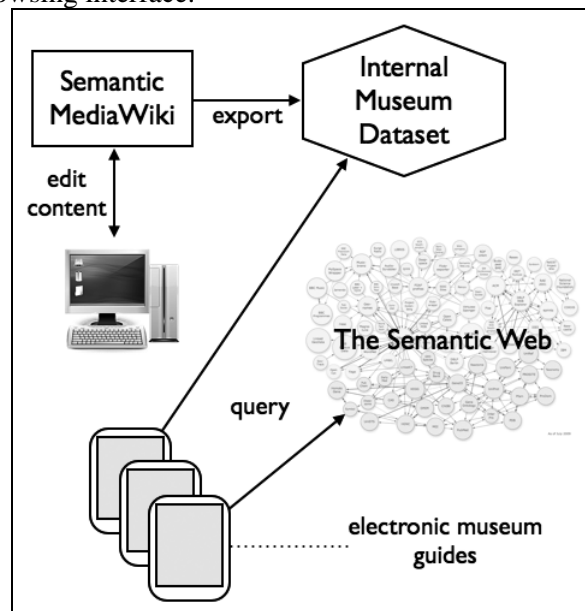

Fig. 3 – Queries can be made over data managed in-house as well as over the Semantic Web in general

Faceted browsing relies on faceted classification, whereby a particular subject can be described using various facets, depending on the classifier's perspective. For example, a collection of artefacts in an exhibition might be classified according to diverse facets such as epoch, artist, manufacturer,

religious or cultural significance, language, country of origin, etc. Faceted browsing thus involves traversing a path through the facets by applying the categories sequentially and simultaneously, allowing the user to iteratively filter the data until the desired information is revealed.

The Semantic Drilldown [14] is another extension to the MediaWiki specifically tailored to browse content gathered with the Semantic MediaWiki. It is an interface based on the concept of faceted browsing, allowing the user to 'drill down' through the layers of information to discover specific data by applying successive filters and categories to a knowledge base. The initial filtering could be based on the RFID returned by the sensors to the electronic guide, providing the visitor with an immediate overview of the information available on the artefacts arrayed before her in a particular vitrine.

5. RELATED WORK

The Semantic Computing Research Group [15] of Aalto University is researching the application of Semantic Web technology to cultural heritage and the creation of "semantic portals" (see [16]).

6. CONCLUSION

This paper represents the preliminary approach to providing a system that will allow multiple content providers to input content into a knowledge base in a flexible and collaborative manner. Flexibility is a major requirement considering the research nature of the POSEIDON project, as well as the constantly changing environment of the exhibiting museum. By relying on Semantic Web technologies, it is hoped the museum visitor will be able to explore the knowledge base provided on the electronic museum guide in an unrestrained and personalised manner.

A major avenue of future research will be to enable the system to browse the wider Semantic Web and not just to restrict it to the content developed in-house. This necessitates the development of ontologies geared towards a particular exhibition, as well as the development of higher level ontologies that capture the fundamental nature of the museum.

The technologies surrounding the Semantic Web are designed to make the vast amount of information available online more conducive to machine processing by developing frameworks for recording metadata and the relationships between metadata (ontologies). The museum, whose purpose is not only to preserve artefacts but also to educate the populace, can profit from the Semantic Web by harnessing its technologies to allow visitors to actively explore the issues associated with any particular exhibition.

8. REFERENCES

[1] http://www.poseidon-projekt.de/
[2] T. Berners-Lee. Information Management: A Proposal. Retrieved online at http://www.w3.org/History/1989/proposal.html on 2010-06-28.
[3] T. Berners-Lee. J. Hendler. O. Lassila. The Semantic Web, *Scientific American*, May 2001.
[4] Graphic courtesy of Steve Bratt, http://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/#(24)
[5] http://www.w3.org/RDF/
[6] http://www.w3.org/TeamSubmission/turtle/
[7] M. Andreessen. Innovators of the Net: Ramanathan V. Guha and RDF. Retrieved online at http://web.archive.org/web/20080205163659/http://wp.netscape.com/columns/techvision/innovators_rg.html on 2010-06-28.
[8] http://www.w3.org/TR/owl2-overview/
[9] http://www.w3.org/TR/rdf-sparql-query/
[10] http://dbpedia.org/
[11] D. Allemang. J. Hendler. *Semantic Web for the Working Ontologist: Modeling in RDF, RDFS and OWL.* Elsevier, 2008, p. 7.
[12] http://semantic-mediawiki.org/
[13] http://www.mediawiki.org/
[14] http://www.mediawiki.org/wiki/Extension:Semantic_Drilldown
[15] http://www.seco.tkk.fi/
[16] E. Hyvönen. Semantic Portals for Cultural Heritage. *Handbook on Ontologies*. S. Staab & R. Studer (Eds.). Springer-Verlag, 2009, pp. 757–778.

Heinz-Günter Kuper *completed his B.A. (1ˢᵗ Class Honours) at the University of Western Australia in 1995 before moving to Germany, where he received his Informatik Diplom in 2005 from the Humboldt University of Berlin. He is currently working as a researcher on the POSEIDON project developing an electronic museum guide based on RFID technology. His research interests include the Semantic Web and the application of RFID in the museum environment.*