Publishing Museum Collections on the Semantic Web – The MuseumFinland Portal

Eero Hyvönen, Miikka Junnila, Suvi Kettula, Eetu Mäkelä,
Samppa Saarela, Mirva Salminen, Ahti Syreeni, Arttu Valo, and Kim Viljanen
Helsinki Institute for Information Technology (HIIT), University of Helsinki
P.O. Box 26, 00014 UNIVERSITY OF HELSINKI, FINLAND
http://www.cs.helsinki.fi/group/seco/
firstname.lastname@cs.helsinki.fi

ABSTRACT

Museum collections contain large amounts of data and semantically rich, mutually interrelated metadata in heterogeneous databases. The publication of museum collections on the web is therefore a very promising application domain for semantic web techniques. We present a semantic web portal called "MuseumFinland — Finnish Museums on the Semantic Web"[1] [3] that contains some 4,000 cultural artifacts from the collections of three museums using three different database schemas and database systems. The system is based on seven RDF(S) ontologies consisting of some 10,000 classes and individuals.

Categories and Subject Descriptors: H.3.6 Information Systems: Online Information Services

General Terms: Design

Keywords: Semantic Web, ontology, content publishing

1. THE GOALS

The goals of developing the system were: 1) Provide the public a global view to the heterogeneous collections in Finland. 2) Provide the end-user with a content-based search-engine for finding objects of interest, and a semantic recommendation system for browsing the collections. 3) Create for the museums a national publication channel for publishing contents on the Semantic Web. In this paper, these goals and solutions developed in our work are summarized.

2. GLOBAL VIEW OF COLLECTIONS

The heterogeneity of museum collection databases creates a severe obstacle to information retrieval. The traditional solution to this problem is to use multi-search[2]. Here the query is send to the local museum database systems, answered there *locally*, and the results are then appended into the global answer (hit list) by the portal. In contrast, MuseumFinland solves the problem by creating a dynamic seamless space of web pages depicting collection objects. For example, figure 1 shows the web page of a distaff, a part used in a spinning wheel. The metadata is seen in the middle and links to other object pages on the right. The web space can 1) be searched by a semantic search-engine and 2) be browsed using the links on the right. The multi-search paradigm cannot facilitate such services based on global inter-collection associations between objects.

3. VIEW-BASED SEARCH ENGINE

The semantic search engine is a new server-based version of Ontogator[3] [4]. It is based on the *multi-facet* search paradigm [6, 2]. The idea is to organize the concepts and individuals of the underlying knowledge base into orthogonal category taxonomies called *views*. Views are used extensively in the user interface in helping the user to formulate the queries. The user can express the query easily in the right terminology by selecting (sub)categories from the views. For example, by selecting "carpet" from an artifact type taxonomy and "silk" from a material taxonomy, silk carpets are found.

There are nine view hierarchies in use grouped under the four headings of "Artifact Characteristics" (object type and object material views), "Artifact Creation" (manufacturer, manufacturing time, and manufacturing location views), "Usage" (user, place of usage, and usage event views), and "Collections" (museum collection view). The view hierarchies are projected from the underlying seven ontologies used in annotating the collection data. The projection is based on logical predicates that define how the hierarchy is formed. For example, in the object type view the relations rdfs:subClassOf and rdf:type are used, while in the manufacturing location and place of usage views the part-of relation is used. More complex projection rules are needed, e.g., for associating objects with the usage event view. SWI-Prolog[4] with its RDF parser is used for creating the projections.

Finding relevant categories becomes a search problem of its own when dealing with thousands of categories. To help the user, Ontogator also contains a search engine for finding view categories.

4. SEMANTIC RECOMMENDATIONS

A major goal of the MuseumFinland is to reveal the rich semantic linkage connecting the collection objects with each

[1] http://museosuomi.cs.helsinki.fi

[2] Artefacts Canada, http://www.chin.gc.ca/, Australian Museums and Galleries Online, http://www.amol.org.au/, etc.

[3] Ontogator is written in Java and Jena 2.0, http://www.hpl.hp.com/semweb/jena.htm.

[4] http://www.swi-prolog.org

WWW2004, May 17–22, 2004, New York, New York, USA.
ACM 1-58113-912-8/04/0005.

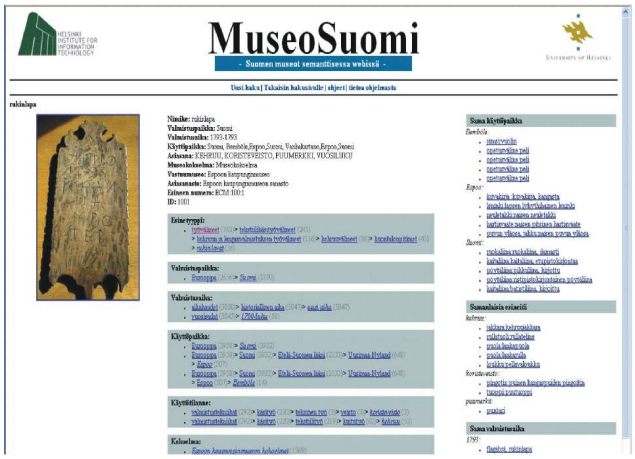

Figure 1: A collection object web page.

other. The linkage is shown to the end-user as hypertext links. The links are defined declaratively in terms of Prolog predicates. Each predicate defines a semantic association and gives it a label as an explanatory name for the link.

For example, in figure 1 the semantic links on the right point to to objects used at the "same location" (categorized according to the name of the common location), to objects related to "similar events" (e.g., objects used in spinning, and decorative objects, because distaffs are usually beautifully decorated), to objects manufactured at the "same time", and so on. Since a decoratively carved distaff used to be a typical wedding gift in Finland, it is also possible to recommend links to other objects used as wedding gifts, such as wedding rings.

The semantic recommendation system is implemented as a logic server called Ontodella based on the SWI-Prolog HTTP server version. The MuseumFinland system itself is a Cocoon-based server[5] that communicates with Ontogator and Ontodella servers with XML/RDF messages.

5. CONTENT PUBLICATION PROCESS

A practical goal of our work is to design a *process* for Finnish museums to publish their collections on the Semantic Web. In our process scheme [5], the museum first transforms its collection data into XML (cf. figure 2). Each collection object is represented as an *XML card* that describes the object in terms of 22 properties whose values are strings and numbers read from the underlying database. The XML Schema used is agreed upon the participating museums and guarantees syntactic interoperability of the collections.

Next, each XML Card is transformed into an *RDF card* with similar RDF properties, but where up to 16 string values are transformed into the URIs of the corresponding classes and individuals in the ontologies. This transformation is based on a set of *term cards* that map terms with ontology resources. MuseumFinland provides the museums with 1) the RDF(S) ontologies, 2) a set of term cards. The museums can adapt their terminological conventions to the portal by creating new term cards of their own. Two special tools has been developed for creating terminologies

[5]http://cocoon.apache.org

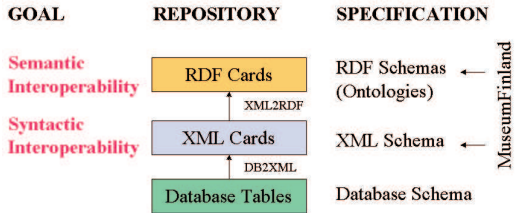

Figure 2: Data transformations in MuseumFinland.

and RDF annotations, and Protégé-2000[6] is used for manual editing.

6. DISCUSSION

MuseumFinland is an application of the idea of semantic portals to solving interoperability problems of museum collection databases when publishing their content on the Semantic Web. The novelty of the view-based search engine of MuseumFinland with respect to other view-based systems [6, 2] lies in its capability of using RDF(S) ontologies as the basis of search [4]. The idea of semantic recommendations is related to Topic Maps, Open Hypermedia [1], and the HyperMuseum [7].

Acknowledgments

Our work is funded mainly by the National Technology Agency Tekes, Nokia, TietoEnator, the Espoo City Museum, the Foundation of the Helsinki University Museum, the National Board of Antiquities, and the Antikvaria Group consisting of some 20 Finnish museums.

7. REFERENCES

[1] C. Goble, S. Bechhofer, L. Carr, D. De Roure, and W. Hall. Conceptual open hypermedia = the semantic web? In *Proceedings of the WWW2001, Semantic Web Workshop, Hongkong*, 2001.

[2] M. Hearst, A. Elliott, J. English, R. Sinha, K. Swearingen, and K.-P. Lee. Finding the flow in web site search. *CACM*, 45(9):42–49, 2002.

[3] E. Hyvönen, M. Junnila, S. Kettula, , E. Mäkelä, S. Saarela, M. Salminen, A. Syreeni, A. Valo, and K. Viljanen. MuseumFinland—Finnish Museums on the Semantic Web. User's perspective. In *Proceedings of Museums and the Web 2004 (MW2004), Arlington, Virginia, USA*, 2004.

[4] E. Hyvönen, S. Saarela, and K. Viljanen. Ontogator: combining view- and ontology-based search with semantic browsing. In *Proceedings of the XML Finland 2003 conference. Kuopio, Finland*, 2003.

[5] E. Hyvönen, M. Salminen, S. Kettula, and M. Junnila. A content creation process for the semantic web. In *Proceeding of OntoLex 2004: Ontologies and Lexical Resources in Distributed Environments*, Lisbon, Portugal, 2004 (forthcoming).

[6] A. S. Pollitt. The key role of classification and indexing in view-based searching. Technical report, University of Huddersfield, UK, 1998. http://www.ifla.org/IV/ifla63/63polst.pdf.

[7] P. Stuer, R. Meersman, and S. De Bruyne. The HyperMuseum theme generator system: Ontology-based internet support for active use of digital museum data for teaching and presentations. In *Museums and the Web 2001: Selected Papers*. Archieves & Museum Informatics, 2001.

[6]http://protege.stanford.edu/