Cultural Semantic Interoperability on the Web: Case Finnish Museums Online

Eero Hyvönen, Suvi Kettula, Vilho Raatikka, and Kim Viljanen

University of Helsinki and Helsinki Institute for Information Technology (HIIT)

P.O. Box 26, 00014 UNIVERSITY OF HELSINKI

email: firstName.lastName@cs.helsinki.fi

The Problem

More and more museum collections are published on the Web. A problem in this work is the heterogeneous nature of the (meta)data catalogued in the databases. From the *data perspective*, heterogeneity is due to the different non-standard cataloguing formats and practices used, the incomplete or partly erroneus data entered, the different and ad hoc semantic taxonomies in use, and different (natural) languages employed in descriptions. From the *system perspective*, the collections are incompatible: the databases have different schemas, they use different DBMSs provided by different vendors, and the data are geographically distributed.

A goal of our Semantic Computing Research Group [1] is to develop technology enabling the creation of the "Finnish Museums Online" (FMO)[2] semantic portal. FMO makes heterogeneous museum collections (databases) semantically interoperable on the Web. For the end-users, this would facilitate easy-to-use content-based information retrieval from consolidated cultural repositories. A result of our 2-year project will be a demonstrational ontology-based information retrieval system combining two major collection cataloguing database systems used in Finland: Escoll/Antikvaria[3] and Musketti[4].

Contributions

We argue that Semantic Web technologies provide a useful basis for realizing the FMO vision: (1) XML Schemas can be used to define mediating languages for syntactic interoperability between collection data in different museums. (2) RDF and RDF Schemas with ontology languages, such as DAML+OIL, provide means for extending syntactic interoperability into semantic interoperability. (3) Ontologies can be employed in helping the catalogists to enter the data in a semantically valid form. (4) Ontologies can be used a basis for user-friendly content-based information retrieval [2].

Overview of FMO

In our system, each museum provides the collection data to be published on the Web as an XML repository conforming to an XML Schema. XML data is transformed into RDF (instances) defining the metadata of the collection records. Semantic validity is checked against a set of RDF Schemas defining the collection ontologies, such as ObjectType, Material, Place, Time etc. By associating the ontology concepts with cataloging terms of linguistic thesauri in use, such as the Finnish General Thesaurus (YSA), the Outline of Cultural Materials (OCM) standard, and the Finnish Museum Thesaurus, human classification work can be eased. Data values used in the RDF instance descriptions belong to the different classes of the ontologies. The ontologies are hence the "glue" that makes the different terminologies used by the museum catalogists semantically compatible with each other.

Ontologies also make semantic navigation possible for the end-user. We are implementing a semantic information retrieval system Ontogator based on the view-based HiBrowse model [1]. The user can select classes from ontologies, and the system finds the instances that match the selected class restrictions. By constraining classes (views) further, the desired collection instance data are eventually found. In addition, Ontogator is a kind of intelligent dynamic browser for the ontologies and related instances. At each web page, a set of "interesting links" is generated based on the ontology context and the related collection records. These links recommend next steps for browsing in the same spirit as Topic Map associations are used for navigation by binding related topics with each other.

Status and Partners

The project started in the spring 2002 and is finished in the spring 2004. The first experimental demonstration of the Ontogator is scheduled for May, 2002.

We thank Espoo City Museum, National Board of Antiquities, Helsinki University Museum, Nokia Research Centre, TietoEnator, and National Technology Agency (TEKES) for co-operation and funding the research.

[1] http://www.cs.helsinki.fi/group/seco/

[2] As visioned in http://www.museoliitto.fi/kuldi.htm.

[3] http://www.ouka.fi/taidemuseo/antikvaria.htm

[4] http://www.nba.fi/DEVELOP/jarjest.htm

References

[1] A Steven 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.

[2] A. Schreiber, B. Dubbeldam, and B. Wielinga. Ontology-based photo-annotation. *IEEE Intelligent Systems*, May/June 2001.