A Portal for Publishing Museum Collections on the Semantic Web

Eero Hyvönen and Samppa Saarela and Kim Viljanen and Eetu Mäkelä
and Arttu Valo and Mirva Salminen and Suvi Kettula and Miikka Junnila [1]

Abstract.

This paper presents the semantic portal MUSEUMFINLAND[2] for publishing museum collections on the Semantic Web. It is shown how museums with their semantically rich and interrelated collection content can create a large, consolidated semantic collection portal together on the web. By semantic web techniques, it is possible to make collections semantically interoperable and provide the museum visitors with intelligent content-based search and browsing services to the global collection base.

1 THE GOALS

"MUSEUMFINLAND — Finnish Museums on the Semantic Web" is a semantic portal that contains metadata from the collection databases of the National Museum[3], Espoo City Museum[4], and Lahti City Museum[5], and more content is being ported into the system. The application is intended for the public in the large to use.

The goals for developing the system were the following:

Global view to distributed collections It is possible to use the heterogeneous distributed collections of the museums participating in the system as if the collections were in a single uniform repository.

Content-based information retrieval The system supports intelligent information retrieval based on ontological concepts, not on simple keyword string matching as is customary with current search engines.

Semantically linked contents A most interesting aspect of the collection items to the end-user are the implicit semantic relations that relate collection data with their context and to each other. In MUSEUMFINLAND, such associations are exposed to the end-user by defining them in terms of logical predicate rules that make use of the underlying ontologies and collection metadata.

Easy local content publication The portal should provide the museums with a cost-effective publication channel.

In the following, these goals and solutions developed in our work are described. After this, main results of the work are summarized, lessons learned discussed, and directions for further research outlined.

2 GLOBAL VIEW TO COLLECTIONS

Museum databases are usually situated at different locations and use different database systems and schemas. This creates a severe obstacle to information retrieval. To address the problem, the web can be used for creating a single interface and access point through which a search query can be sent to distributed local databases and the results combined into a global hit list. This "multi-search" approach is widely applied and there are many cultural collection systems on the web based on it, such as the portals Australian Museums & Galleries Online[6] and Artefacts Canada[7].
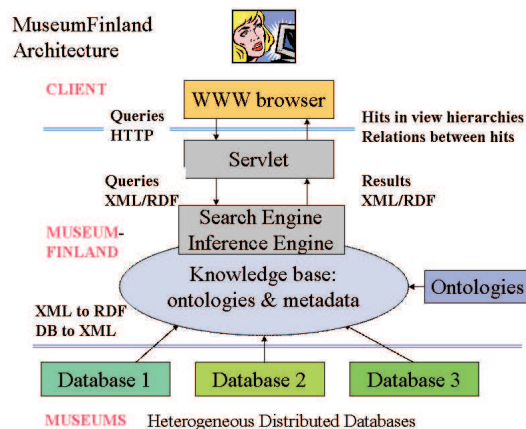

Figure 1. Information retrieval in MUSEUMFINLAND. Local database contents are first merged and the query is evaluated with respect to the global interrelated data.

A problem of multi-search is that by processing the query independently at each *local database*, the *global* dependencies, associations between objects in different collections are difficult to found. Since exposing semantic associations between

collections items is one of our main goals, MUSEUMFINLAND cannot be based on the multi-search paradigm. Instead, the local collections are first consolidated into a global repository, and the queries are answered based on it (cf. figure 1). Mutually shared conceptual models, ontologies, are used for enriching the content and for making the collections interoperable. To show the associations to the end-user, the collection items are represented as web pages interlinked with each other through the semantic associations. The MUSEUM-FINLAND home page is the single entry point through which the end-user enters the global semantic WWW space.
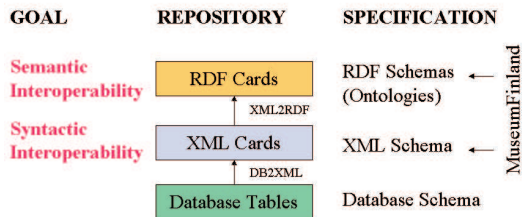

Figure 2. Data transformations in MUSEUMFINLAND.

The challenge in consolidating the collections is how to make them interoperable in syntax and, especially, in semantics. In our solution [9], 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 by 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 a set of underlying RDF(S) ontologies. This transformation is based on a set of *term cards* that map terms with ontology resources. MUSEUMFINLAND provides the museums with the ontologies and 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 have been developed for creating terminologies (Terminator) and RDF annotations (Annomobile) semi-automatically. Protégé-2000[8] is used for the manual editing part.

3 MULTI-FACET SEARCH BASED ON ONTOLOGIES

The content-based search engine of MUSEUMFINLAND is a server called Ontogator. Ontogator is based on the multi-facet view-based search paradigm developed within the information retrieval research community [13, 4, 8]. Multi-facet search is based on a set of *categories* that are organized into a set hierarchical, orthogonal taxonomies called subject *facets* or *views*.

A search query in multi-facet search is formulated by selecting categories of interest from the different facets. For example, by selecting the category "Furniture" from the Artifact s

View category	View	Ontology
Object	Artifact	Artifacts
	Material	Materials
Creation	Creator	Actors
	Location of creation	Locations
	Time of creation	Times
Usage	User	Actors
	Location of usage	Locations
	Situation of usage	Situations
Museum	Collection	Collections

Table 1. View facets in the MUSEUMFINLAND portal.

facet, and "Eero Saarinen" from the Creator facet, the user can express the query for retrieving all kinds of furniture, such as chairs, tables, etc., created by Eero Saarinen. Intuitively, the query is a conjunctive constraint over the facets with disjunctive constraints over the sub-categories in each facet.

More formally, if the categories selected are $C_1, ..., C_n$ and the subcategories of $C_i, i = 1...n$, including C_i itself are $S_{i,1}, S_{i,2}..., S_{i,k}$, respectively, then this selection corresponds to the following boolean AND-OR-constraint:

$$(S_{1,1} \vee ... \vee S_{1,k}) \wedge (S_{2,1} \vee ... \vee S_{2,l}) \wedge ... \wedge (S_{n,1} \vee ... \vee S_{n,m}) \quad (1)$$

Facets can be used for helping the user in information retrieval in many ways. Firstly, the facet hierarchies give the user an overview of what kind of information there is in the repository. Secondly, the hierarchies can guide the user in formulating the query in terms of appropriate keywords. Thirdly, the hierarchies can be used to disambiguate homonymous query terms. Fourthly, the facets can be used as a navigational aid when browsing the database content [4]. Fourthly, the number of hits in every category that can be selected next can be computed *beforehand* and be shown to the user [13]. In this way, the user can be hindered from making a selection leading to an empty result set—a recurring problem in IR systems—and is guided toward selections that are likely to constrain (or relax) the search appropriately.

Table 1 depicts the 9 views used in MUSEUMFINLAND and their underlying 7 ontologies. The Artifacts ontology is a taxonomy of the tangible collection objects such as pottery, cloths, weapons, etc. All exhibits in the system belong to some class in this ontology. The Materials ontology is a taxonomy of the artifact materials, such as steel, silk, tree, etc. The Actors ontology defines classes of agents, such as persons, companies etc., and individuals as instances of these classes. The Events ontology include intangible happenings, situations, events, and processes that take place in the society, such as farming, feasts, sports, war, etc. Locations is an ontology representing areas and places on the Earth and in Finland in particular. The Times ontology is a taxonomy of various predefined historical periods, and the Collections ontology classifies the museums and collections in the portal. The Artifacts, Materials, and Events ontologies are subsets of a larger cultural ontology called MAO (6768 classes) that we created based on the Finnish cultural thesaurus MASA [11].

Figure 3 shows the search interface of MUSEUMFINLAND. The nine facet hierarchies of table 1 are shown (in Finnish) on the left. For each facet hierarchy, the next level of sub-
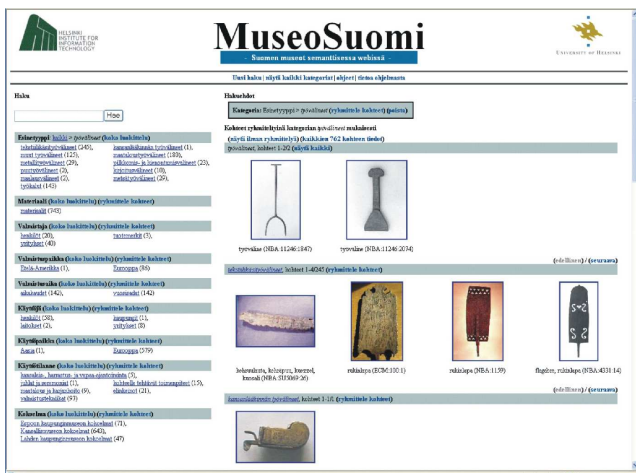
Figure 3. The search interface of MuseumFinland.

categories is shown as links. A query is formulated by selecting a sub-category by clicking on its name. When the user selects a category c in a facet f, the system constrains the search by leaving in the result set only such objects that are annotated in facet f with some sub-category of c. The figure depicts the situation after selecting the sub-category Tools ("työvälineet") from the Artifact facet ("Esinetyyppi"). The result set is shown on the right grouped by the sub-categories of Tools, such as Textile making tools ("tekstiilityövälineet") and Tools of folk medicine ("kansanlääkinnän työvälineet"). Hits in different categories are separated by horizontal bars and can be scrolled independently in each category. In this case, all categories do not fit in the screen shot.

When answering the query, the result set for each direct sub-category in the facets seen on the screen is recomputed, and a number (n) is shown to the user after the category name. It tells that if the sub-category is selected next, then there will be n hits in the result set. For example, the number 643 in the Collection facet on the bottom ("Kokoelma") tells that there are 643 tools in the collections of the National Museum ("Kansallismuseon kokoelmat"). A selection leading to an empty result set ($n=0$) is removed from its facet (or alternatively disabled and shown in gray color, depending on the user's preference). In this way, the user can be hindered from making a selection leading to an empty result set, and is guided toward selections that are likely to constrain the search appropriately. The query can be relaxed by making a new selection on a higher level of the facets or by dismissing the facet totally from the query.

In above, the category selection was made among the direct sub-categories listed in the facets. An alternative way is to click on the link Whole facet ("koko luokittelu") on a facet. The system then shows all possible selections in the facet with hit counts. In this way, the user can easily formulate the query using the right categories exposed to her as links, and can get easily overviews of the database contents along different classifications in different situations.

User studies [10, 2] indicate that if the user does not precisely know what (s)he is looking for, then the multi-facet search method with its "browsing the shelves" sensation is clearly preferred over keyword search (or using a single facet search). Otherwise, a direct Google-like keyword search interface is preferred. To support word-based search, too, an additional search engine was implemented in MuseumFinland (upper left corner in figure 3). This engine is used for two purposes at the same time: for searching categories to be used in multi-facet search and for searching collection objects with matching metadata values in the conventional way. [7]

4   SEMANTIC LINKAGE

One of the main goals of the MuseumFinland portal is to reveal the rich semantic linkage connecting the collection objects with each other. The links can be *explicit* or *implicit*. Explicit links correspond to the RDF statements (triples) in the underlying knowledge base and are based on the collection domain ontologies (classes and their properties) and the actual collection data (instance data). For example, an instance of a painting may have the RDF property dc:creator linking the art work to an individual artist. Implicit links can be defined in terms of explicit ones but are not present in the RDF graph. For example, if there are explicit links linking children with their mothers and fathers, then implicit links such as "grandfather" or "cousin" can be defined.

In MuseumFinland, implicit links are defined declaratively in terms of logic by using Prolog predicates. Each predicate defines a semantic association and gives it an explanatory label, such as "cousin of". By applying such a predicate to a collection item resource, implicitly related other resources with respect to the semantic association can be found. On the HTML level in the user interface, the label of the association is used as the name for the link and the found resource as the target. For example, if the family relations of artists are known in the ontology, then such a predicate could infer links to other pages depicting paintings whose creator is of the same family.

Figure 4. Collection item metadata with semantic recommendations.

For example, figure 4 depicts an collection object page found by multi-facet search. The object is a distaff ("rukinlapa" in Finnish) used in a spinning wheel. On the left, a photo of the object is shown. The metadata of the object is

shown in the middle on top. All facet categories of the object are listed in the middle bottom as hierarchical link paths. A new search can be started by selecting any link from there. On the right, the system displays links to other recommended collection items. i.e., *semantic recommendations.*

The recommendation links provide a semantic browsing facility to the end-user. For example, in figure 4 there are links 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. In MUSEUM-FINLAND, such associations can be exposed to the end-user as link groups whose titles and link names explain to the user the reason for the recommendation. The possibilities for creating such associations are intriguing. Of course, only links that can be inferred based on the metadata and ontologies available can be created.

Recommendations are defined in terms of flexible logical predicate rules using the methods described in [6]. The semantic recommendation system of MUSEUMFINLAND is implemented as a logic server called Ontodella. This system is based on the HTTP server version of SWI-Prolog[9] [15].

There is also a prototype implementation of MUSEUM-FINLAND that can be used with WAP 2.0 compatible mobile telephones. The current prototype recreates all functionality of the web interface in a layout more suitable to the limited screen space of mobile devices. When the user makes a selection for the multi-facet search, impossible category choices leading to empty results can be pruned out. This is a very useful feature for devices that have a small screen to display choices.

5 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 power of MUSEUMFINLAND comes from the use of ontologies:

Exact definitions By using ontologies, the museums can define the concepts used in cataloging in a precise, machine understandable way.

Terminological interoperability The terms used in different institutions can be made mutually interoperable by mapping them onto common shared ontologies. The ontologies are not used as a norm for telling the museums what terms to use, but rather to make it possible to tolerate terminological variance as far as the terminology mapping from the local term conventions to the global ontology is provided.

Ontology sharing Ontologies provide means for making exact references to the external world. For example, in MUSEUMFINLAND, the location ontology (villages, cities, countries, etc.) and the actor ontology (persons, companies, etc.) is shared by the museums in order to make the right

and interoperable references. For example, two persons who happen to have the same name should be disambiguated by different URIs, and a person whose name can be written in many ways, should be identified by a single URI to which the alternative terms refer.

Automatic content enrichment Ontological class definitions, rules, and consolidated metadata enrich collection data semantically.

Intelligent services Ontologies can be used as a basis for intelligent services to the end-user. In MUSEUMFINLAND, the view-based search engine is based on the underlying ontological structures and the semantic link recommendation systems reveals to the end-user the underlying semantical context of the collection items and their mutual relations.

The novelty of the content-based search engine with respect to other view-based systems [13, 4] is based on its capability of using RDF(S) ontologies as the basis of search. The main benefits obtained are: 1) Ontological logical inference can be employed in projecting the views from the ontologies (e.g., the location meronymy and various concept hyponymies). 2) The implicit complicated relations between view categories and the underlying data resources to be searched for can be specified flexibly in terms of logical predicates. Ontogator combines virtues of the view- and ontology-based search paradigms [8].

The idea of linking collection items with semantic associations was inspired by Topic Maps [12]. However, in our case the links are not given by a map but are determined by logical inference using the underlying RDFS ontology and RDF metadata. Another application of this idea to generating semantically linked static HTML sites from RDF(S) repositories is presented in [6]. Logic and dynamic link creation on the semantic web has been discussed, e.g., in the work on Open Hypermedia [3, 1]. In the HyperMuseum [14], collection items are also semantically linked with each other. Here linking is based on shared words in the metadata and their linguistic relations, such as synonymy and antonymy. In contrast, our system is not based on words but on ontological references in the underlying RDF(S) knowledge base and the links can be defined freely in terms of logical rules. The idea of annotating cultural artifacts in terms of multiple ontologies has been explored, e.g., in [5].

5.1 Lessons Learned

The main problem encountered in the content work was that the original museum collection data in the databases was not systematically annotated. Various conventions are in use in different museum systems and museums. Much of the metadata is not based on a keywords but is free text. The Terminator and Annomobile tools developed for the XML to RDF transformation [9] are only semi-automatic, and a human editor is often needed to make the right annotations. Due to homonymy, not even thesaurus keywords can always be to mapped unambiguously to RDF concepts by the machine. However, the homonymy problem turned out to be less severe than expected, because disambiguation could be based on the facet/ontology to which the database field was related.

The view-based search method can be implemented quite efficiently. The current system scales up to the order of 10,000 RDF cards and 10,000 ontological concepts on an ordinary PC

server. From the user's perspective, the idea of multi-facet search seems useful and a natural next step a head from the single facet systems on the web today, such as Yahoo[10] and Open Directrory Project[11]. Using Prolog and RDF together for projecting the facets and for creating the semantic recommendation links was powerful and flexible. It is possible compute and store the results of some inferences before running system in order to speed up reasoning. In our case, the mappings between facet categories and RDF resources are determined in Prolog beforehand and are compiled into an RDF tree that can be used more efficiently by the view-based search engine. The semantic recommendations are currently determined dynamically.

MUSEUMFINLAND user interface was first implemented as a Java servlet using XSLT transformations. The system was then re-designed and re-implemented as a Cocoon-based server[12] that queries the Ontogator search engine server and Ontodella logic server with XML/RDF messages. It is possible to do this over HTTP. With Cocoon, the implementation could be made in a couple of months and can be modified easily. For example, the mobile telephone interface was created by modifying the PC version.

5.2 Further Work

We are investigating how new kind of RDF material, conforming to different ontologies, such as art collections using the Iconclass[13] system and educational videos based on the IEEE Learning Objects Metadata standard, can be merged in the portal. More work is needed in developing a set of recommendation predicates that would be of most interest to the users.

Ways of collaboration between museum content providers and portal maintenance people need to be developed in order to develop MUSEUMFINLAND from an application into a continuous publication *process*. For example, protocols for adding, modifying, and retracting RDF cards and ontology resources according to the wishes of the museums need to be developed.

The pilot version of MUSEUMFINLAND was opened on the public web in March 2004 at:

`http://museosuomi.cs.helsinki.fi`

Acknowledgments

REFERENCES

[1] P. Dolong, N. Henze, and W. Neijdl, 'Logic-based open hypermedia for the semantic web', in *Proceedings of the Int. Workshop on Hypermedia and the Semantic Web, Hypertext 2003 Conference, Nottinghan, UK*, (2003).

[2] J. English, M. Hearst, R. Sinha, K. Swearingen, and K.-P. Lee, 'Flexible search and navigation using faceted metadata', Technical report, University of Berkeley, School of Information Management and Systems, (2003). Submitted for publication.

[3] 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).

[4] 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).

[5] L. Hollink, A. Th. Schereiber, J. Wielemaker, and B.J. Wielinga, 'Semantic annotations of image collections', in *Proceedings KCAP'03, Florida*, (2003).

[6] E. Hyvönen, M. Holi, and K. Viljanen, 'Designing and creating a web site based on RDF content', in *Proceedings of the WWW2004 workshop, Application design, development and implementation isssues in the semantic web*, (2004). http://www.cs.helsinki.fi/u/eahyvone/publications/swehg_www2004.pdf.

[7] E. Hyvönen, M. Junnila, S. Kettula, E. Mäkelä, S. Saarela, M. Salminen, A. Syreeni, A. Valo, and K. Viljanen, 'Finnish Museums on the Semantic Web. User's perspective on MuseumFinland', in *Proceedings of Museums and the Web 2004 (MW2004), Arlington, Virginia, USA*, (2004). http://www.archimuse.com/mw2004/papers/hyvonen/hyvonen.html.

[8] E. Hyvönen, S. Saarela, and K. Viljanen, 'Application of ontology techniques to view-based semantic search and browsing', in *The Semantic Web: Research and Applications*, eds., C. Bussler, J. Davies, D. Fensel, and R. Studer, LNCS 3053. Springer–Verlag, Berlin, (2004).

[9] E. Hyvönen, M. Salminen, S. Kettula, and M. Junnila. A content creation process for the Semantic Web, 2004. Proceeding of OntoLex 2004: Ontologies and Lexical Resources in Distributed Environments, May 29, Lisbon, Portugal.

[10] K.-P. Lee, K. Swearingen, K. Li, and M. Hearst, 'Faceted metadata for image search and browsing', in *Proceedings of CHI 2003, April 5-10, Fort Lauderdale, USA*. Association for Computing Machinery (ACM), USA, (2003).

[11] *Museoalan asiasanasto*, ed., R. L. Leskinen, Museovirasto, Helsinki, Finland, 1997.

[12] Steve Pepper, 'The TAO of Topic Maps', in *Proceedings of XML Europe 2000, Paris, France*, (2000). http://www.ontopia.net/topicmaps/materials/rdf.html.

[13] 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.

[14] Peter Stuer, Robert Meersman, and Steven 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*, eds., D. Bearman and J. Trant, Archieves & Museum Informatics, (2001). http://www.archimuse.com/mw2001/papers/stuer/stuer.html.

[15] J. Wielemaker, A. Th. Schereiber, and B. J. Wielinga, 'Prolog-based infrastructure for RDF: performance and scalability', in *Proceedings ISWC'03, Florida*. Springer–Verlag, Berlin, (2003).

---

[10] www.yahoo.com
[11] www.dmox.org
[12] http://cocoon.apache.org
[13] http://www.iconclass.nl