CultureSampo—Finnish Culture on the Semantic Web: The Vision and First Results

Eero Hyvönen, Tuukka Ruotsalo, Thomas Häggström,
Mirva Salminen, Miikka Junnila, Mikko Virkkilä, Mikko Haaramo,
Eetu Mäkelä, Tomi Kauppinen, and Kim Viljanen

*Semantic Computing Research Group
Helsinki University of Technology (TKK), Laboratory of Media Technology
University of Helsinki, Department of Computer Science
http://www.seco.tkk.fi/
first.last@tkk.fi

Abstract

This paper concerns the idea of publishing heterogenous cultural content on the Semantic Web. By heterogenous content we mean metadata describing potentially any kind of cultural objects, including artifacts, photos, paintings, videos, folklore, cultural sites, cultural process descriptions, biographies, history etc. The metadata schemas used are different and the metadata may be represented at different levels of semantic granularity. This work is an extension to previous research on semantic cultural portals, such as MuseumFinland, that are usually based on a shared homogeneous schema, such as Dublin Core, and focus on content of similar kinds, such as artifacts. Our experiences suggest that a semantically richer event-based knowledge representation scheme than traditional metadata schemas is needed in order to support reasoning when performing semantic search and browsing. The new key idea is to transform different forms of metadata into event-based knowledge about the entities and events that take place in the world or in fiction. This approach facilitates semantic interoperability and reasoning about the world and stories at the same time, which enables implementation of intelligent services for the end-user. These ideas are addressed by presenting the vision and solution approaches taken in two prototype implementations of a new kind of cross-domain semantic cultural portal "CULTURESAMPO—Finnish Culture on the Semantic Web".

1 Towards Semantic Cross-domain Interoperability

A widely shared goal of cultural institutions is to provide the general public and the researchers with aggregated views to cultural heritage, where the users are able to access the contents of several heterogenous distributed collections of institutions *simultaneously*. In this way, the organizational and technical obstacles for information retrieval between collections and organizations, even between countries and languages could be crossed.

Content aggregation may occur at the syntactic or semantic level. The basis for *syntactic interoperability* is sharing syntactic forms between different data sources, i.e., the metadata schemas such as the Dublin Core Metadata Element Set[1] or the Visual Resource Association's (VRA) Core Categories[2]. Such schemas make it possible to identify different aspects of the search objects, such as the "author", "title", and "subject" of a document, and focus search according to these. Syntactic interoperabilty facilitates, for example, multi- or metasearch[3]. Here the user types in a query in a metaportal. The query is then distributed to a set of underlying systems and the results are aggregated for the end-user. For example, the Australian Museums and Galleries Online[4] and Artefacts Canada[5] are multi-search engines over nation-wide distributed cultural collections. Here the content includes metadata about museum artifacts, publications etc. represented using shared metadata schemas.

Content aggregation at the *semantic level* means that not only the form of the data is shared and in-

[1] http://dublincore.org/documents/1998/09/dces/

[2] http://www.vraweb.org/vracore3.htm

[3] http://en.wikipedia.org/wiki/Metasearch_engine

[4] http://www.amonline.net.au/

[5] http://www.chin.gc.ca/

teroperable, but also the values used in the metadata schema, and that the meanings of the values are semantically defined in terms of ontological structures. The values of metadata fields, such as authors, material types, and geographical locations are taken from a set of shared vocabularies, i.e., ontologies, or if different vocabularies are used, then the mappings between them are available. At this level of content aggregation, reasoning about the ontological relations between content items becomes possible enabling semantic search, semantic browsing, recommendations, explanations, and other "intelligent" services. A prototypical example of this approach is the portal "MUSEUMFINLAND—Finnish Museums on the Semantic Web"[6] (Hyvönen et al., 2005a), where distributed, syntactically heterogeneous museum collection databases are integrated by a set of seven shared ontologies, and semantic search and browsing services are provided to end-users based on the aggregated knowledge base.

Another distinctive feature between cultural content aggregation systems is whether they deal with metadata that conforms to a *single metadata schema* or *multiple schemas*. For example, the Helmet library system[7] aggregates public library collections of neighboring cities for the public by using a single metadata format. In the same vein, an online record shop may deal with CD/DVDs whose metadata is represented in a homogeneous way. On the other hand, in a system such as Artefacts Canada, the underlying databases contain items of different kinds, such as art, furniture, photos, magazines etc. whose metadata conform to different schemas. For example, a metadata field representing physical the material of an object is essential for a piece of furniture or artifact but not for a publication.

Semantic web portals have tackled the problem of semantic interoperability usually by sharing metadata schemas. For example, in MUSEUMFINLAND heterogeneous artifact collection databases were made semantically interoperable, but the content was of a single domain (artifacts), and the metadata was based on a single, Dublin core like schema of artifacts. There are paintings and some other forms of art in MuseumFinland collections, but they have been cataloged as pieces of artifacts in the cultural museums participating in the portal, and not as pieces of art. The reasoning routines were based on the annotation schema and the ontologies.

In this paper we investigate the problem of *semantic cross-domain interoperability*, i.e. how content of different kinds conforming to multiple metadata schemas could be made semantically interoperable. The focus is the cultural domain and content types studied in our work include artifacts, paintings, photographs, videos, audios, narratives (stories, biographies, epics), cultural processes (e.g., farming, booth making), cultural sites, historical events, and learning objects. In this case, the content is cross-domain in nature and, as a result, comes in forms that may be quite different from each other. Mapping them into a Dublin Core like generic metadata framework is problematic. Instead, we propose content aggregation at a semantically more foundational and rich level based on events and thematic roles (Sowa, 2000). The research is being carried out not only in theory, but by implementing real portal prototypes. More specifically, we show how the idea of MUSEUMFINLAND can be extended into a cross-domain semantic cultural portal called "CULTURESAMPO—Finnish Culture on the Semantic Web". Figure 1 illustrates the positioning of CULTURESAMPO along the distinctions discussed above and its relation to some other portal systems mentioned.

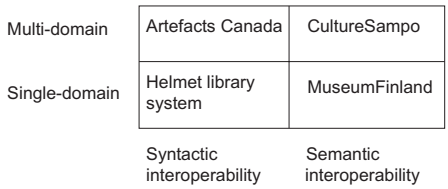

Figure 1: Portals can be classified in terms of the number of metadata schemas used (vertical axis) and the level of interoperability (horizontal axis).

In the following we first state the vision and goals of creating CULTURESAMPO. After this problems of obtaining semantic cross-domain interoperability are discussed and the idea of using event-based descriptions is proposed as a solution. The discussion is based on experiences gathered in creating two experimental prototypes of CULTURESAMPO. In conclusion, contributions of the work are summarized and directions for further research are proposed.

2 The Vision and Goals of CultureSampo

CULTURESAMPO shares the general goals of MUSEUMFINLAND:

[6] This application is operational at http://www.museosuomi.fi with a tutorial in English.

[7] http://www.helmet.fi

Global view to heterogeneous, distributed contents The portal supports the usage of heterogenous and distributed collections and contents of the participating organizations as if there were a single uniform repository.

Intelligent end-user services The system supports *semantic search* based on ontological concepts and *semantic browsing*, where semantic associations between search objects are exposed dynamically to the end-user as recommendation links with explicit explanations. These links are defined in terms of logical rules that make use of the underlying ontologies and collection metadata.

Shared content publication channel The portal should provide the participating organizations with a shared, cost-effective publication channel.

CULTURESAMPO focuses, from the content perspective, especially on material related to the "Golden Era" of the Finnish culture in the 19th century. During this period the notion of Finland as a nation with an original cultural background and history was formed, and the development resulted in the independence of the country in 1917.[8] A central component of the Finnish cultural heritage has been the national epic Kalevala[9]. It was published originally in 1835 and has been translated into some 60 languages. This epic, based on large collections of folklore[10] collected especially in the eastern parts of Finland, Karelia, has been a continuous source of inspiration in Finnish fine arts, music, sculpture, literature, and other branches of culture. The world of Kalevala also nicely relates to the original agrarian Finnish life and artifacts that are available in museums.

In CULTURESAMPO the Karelian culture is central also because one goal of the work is to reunite Karelian collections using semantic web techniques. These collections have now been distributed in several museums due to the result of the World War II where eastern parts of Finland were annexed to the Soviet Union. The semantic web provides means for re-uniting cultural entities virtually on the semantic web. The problem of distributed cultural heritage due to wars and other reasons is very common in Europe. We envision, that the ideas and techniques developed in CULTURESAMPO could later contribute to creation of cross-national and multi-lingual cultural portals, a kind "CultureEurope".

The system will also contribute, in a sense, to the tradition of Kaleva translations. It provides first excerpts of Kalevala translated, not into natural languages for the humans to use but for the machine to "understand" in the formal languages of the semantic web, RDF and OWL.[11]

The latter part of the portal name "Sampo" is the name of the mythical machine-like entity of the Kalevala epic. Sampo gives its owner power, prosperity, everything, but its actual construction and nature is semantically ambiguous and remains a mystery — tens of academic theories about its meaning have been presented. CULTURESAMPO adds still another modern interpretation of what a "Sampo" could be based on the semantic web.

3 Making a Cultural Portal More Intelligent

A major focus of our work in CULTURESAMPO is to study how to provide the end-user with intelligent search and browsing services based on semantically rich cross-domain content originating from different kind of cultural institutions. For example, consider the painting "Kullervo departs for the war" in figure 2 depicting an event in Kalevala. From the end-users' viewpoint, it could probably be interesting, if this piece of art could be linked with other paintings and photos, with the war theme in art museums, with weapons and accessories in cultural museums, with academic studies about Kalevala and Kullervo, with information about dogs and horses in the museum of natural history, with other (external) information on the web about Kullervo, with the actual poem in Kalevala and related pieces of folk poetry, with movies and videos on a media server, with biographies of the artist, and so on. An interesting line of associations could be created by considering the events, processes, and the Kalevala story that takes place in the picture. In this way, for example, the painting could be linked with content concerning the next or previous events in the Kalevala story. Such associations and viewpoints could be insightful, useful, and even entertaining both when searching for content and when browsing it.

To investigate and test the feasibility of this idea in practise, we are extending the portal MUSEUMFINLAND into CULTURESAMPO by a sequence of new prototypes. In 2005, the first prototype to be called "CULTURESAMPO I" was designed and

[8] Before that Finland had been a part of Sweden (until 1809) and Russia (1809-1917).

[9] http://www.finlit.fi/kalevala/index.php?m=163&l=2

[10] http://www.finlit.fi/english/kra/collections.htm

[11] http://www.w3.org/2001/sw/

Figure 2: Kullervo departs for the war. A painting at the Finnish National Gallery.

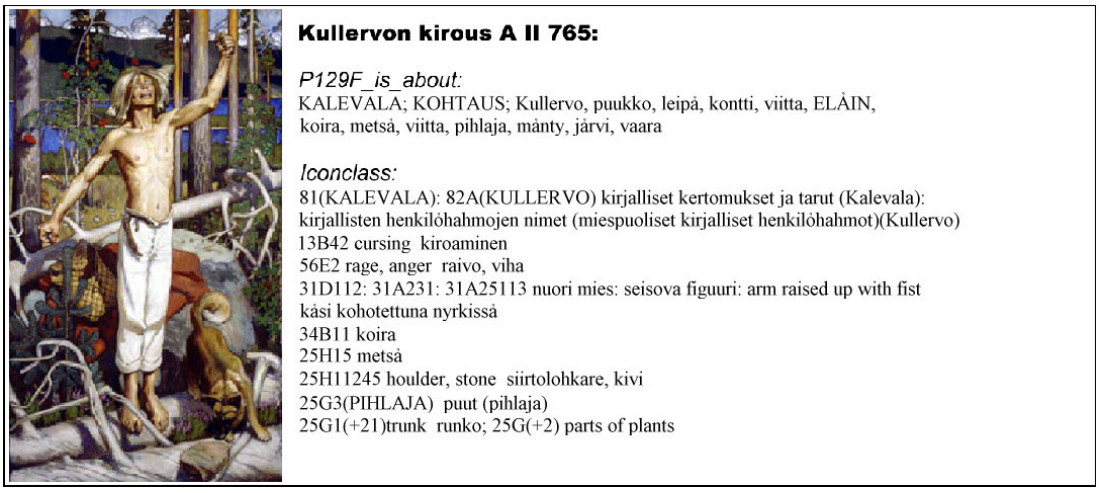

Figure 3: The painting “Kullervo cursing” and its metadata from the Finnish National Gallery.

implemented (Junnila et al., 2006; Junnila, 2006; Salminen, 2006). Figure 3 depicts a painting and its metadata in CULTURESAMPO I. The metadata shown originates from the Finnish National Gallery[12] and describes the subject of the painting in the following way: First, the CIDOC CRM[13] (Doerr, 2003) property `P129F_is_about` lists the following set of keywords (in Finnish): "Kalevala", "event", "Kullervo", "knife", "bread", "knapsack", "robe", "animal", "dog", "forest", "rowan", "pine", "lake", and "mountain". Second, the property "Iconclass" lists a set of ICONCLASS[14] (van den Berg, 1995) notations (categories) describing the subject. This description is partly redundant with the Finnish keywords.

In figure 4 this painting is viewed in CULTURESAMPO I. On the right column, a set of semantic links to other search objects are recommended with explanations created by the logical linking server Ontodella (Viljanen et al., 2006). The figure illustrates a link to a knapsack in the collections of the National Museum of Finland[15], a link to a biography of the artist, and a link to the point in the Kalevala epic where the event of the painting actually takes place.

CULTURESAMPO I was implemented using the same framework as MUSEUMFINLAND, i.e., the OntoViews framework (Mäkelä et al., 2004) including the view-based semantic search engine Ontogator (Mäkelä et al., 2006) and Ontodella (Viljanen et al., 2006). However, in this case much richer cross-domain metadata was used. The test material was limited in size but included examples of artifacts, paintings, photos, videos, biographical information, and narratives such as poems of Kalevala, and descriptions of traditional agrarian processes, such as farming by the slash and burn method.

During this experiment we identified two major obstacles for creating cross-domain semantic cultural portals:

Semantic Interoperability of metadata schemas. The problem of integrating metadata schemas occurs 1) *horizontally* when integrating schemas of different form semantically and 2) *vertically* when integrating content annotated at different levels of granularity.

Expressive power of metadata schemas. A central research hypotheses underlying CULTURESAMPO is that, from the end-user's viewpoint, different processes and events that take place in the society and history should be used as a kind semantic glue by which "insightful semantic links" could be created for the user to browse. This idea was already tested to some extent in MUSEUMFINLAND by creating an artificial event view for the end-user, and by mapping contents of it using logical rules. However, it seemed that a richer and a more accurate knowledge representation method was needed in annotating the contents than traditional metadata schemas.

In the following, our approaches to addressing these problems are outlined.

4 Semantic Interoperability of Metadata Schemas

Re-consider the figure 2. Its metadata may tell e.g. that this painting was created by A. Gallen-Kallela in 1901 in Helsinki. This metadata can be represented, by using RDF triples in Turtle notation[16], in the following way (this example is not based on the actual metadata but is for illustration only):

```
:Kullervo_departs_war
   dc:creator persons:A.Gallen-Kallela ;
   dc:date "1901" ;
   dc:spatial places:Helsinki .
```

The metadata record in a biographical repository, such as the ULAN[17] of the Getty Foundation, may tell us more about the artist in a very different metadata format, e.g.:

```
persons:A.Gallen-Kallela
   :placeOfBirth places:Pori ;
   :timeOfBirth "1865" ;
   :placeOfDeath places:Stockholm ;
   :timeOfDeath "1931" .
```

A problem here is that the number of different properties in metadata schemas easily gets large in cross-domain applications. Furthermore, the meaning of many properties, such as `dc:date` and `dc:spatial` in the metadata schema of paintings and `timeOfBirth/Death` and `placeOfBirth/Death` in the biographical metadata schema of persons may share some meaning, but are still different. We soon realized that when using the schemas for reasoning tasks, the logical rules accounting properly all kinds of combinations

[12] http://www.fng.fi
[13] http://cidoc.ics.forth.gr/
[14] http://www.iconclass.nl
[15] http://www.nba.fi/en/nmf
[16] http://www.dajobe.org/2004/01/turtle/
[17] http://www.getty.edu/vow/ULANSearchPage.jsp

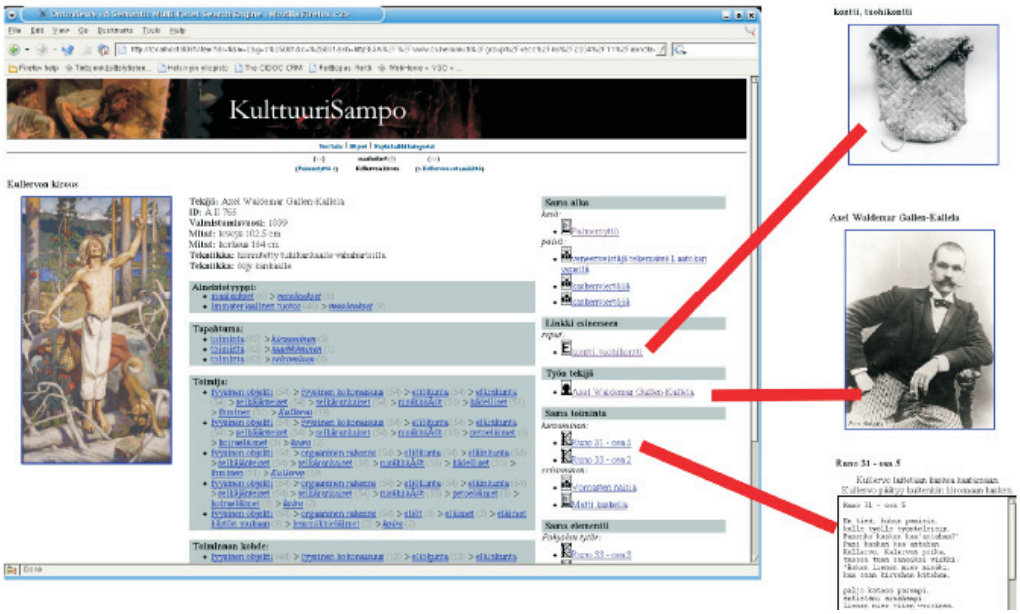

Figure 4: The painting of figure 3 viewed in the semantic portal CULTURESAMPO I. Three semantic recommendation links created by the system are visualized on top of the screenshot.

of properties become complicated, and the number of rules becomes large due to combinatorial explosion. It seems that a more primitive representation of knowledge than traditional metadata schemas is needed for reasoning.

A potential solution approach to solve the problem is to use the CIDOC CRM ontology. The system "provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation"[18]. The framework includes some 80 classes, such as "E22 Man-Made Object", "E53 Place", and "E52 Time-Span", and a large set of some 130 properties relating the entities with each other, such as "P4 Has Time-Span" and "P87 Is Identified By". Interoperability of cultural content can be obtained by mapping metadata standards to CIDOC CRM.

The focus in CIDOC CRM is in modeling concepts necessary for representing the documentation semantics of different metadata schemas used in the cultural domain, such as Dublin Core. In contrast, in CULTURESAMPO our main focus is to represent *real world knowledge* related to cultural heritage, e.g., the subjects that the paintings in figures 2 and 3 depict. For this purpose, a different kind of knowledge representation scheme and large domain ontologies containing tens of thousands of domain concepts and events are needed.

Our solution to the problem of semantic interoperability is to transform different metadata schemas into a shared, more primitive knowledge representation of the real world. In this way, the meaning of `dc:date`, `:timeOfBirth` and `:timeOfDeath` can be made interoperable. By basing reasoning on the more primitive representation, more generic and fewer rules operating a smaller set of properties can be devised. As for the knowledge representation scheme, the idea of representing knowledge in terms of actions and thematic relations between actions and entities was adopted. This general idea has been applied widely in computational linguistics and natural language processing (cf. e.g. (Zarri, 2003)), in knowledge representation research (Sowa, 2000), and also in CIDOC CRM, where events are of central importance, too.

For example, in CULTURESAMPO the three time-relations of the above examples are reduced into only one time-relation relating an instance of an event type, such as "painting_event", "birth_event", or "death_event" to a time entity. The meaning of semantically complex properties in metadata schemas

[18] http://cidoc.ics.forth.gr/

is essentially represented in terms of different events and related entities. For example, the metadata about the painting "Kullervo departs for the war" means that there was a painting event related with A. Gallen-Kallela, the year 1901, and Helsinki by the thematic roles "agent", "time", and "place":

```
:painting_event_45
   rdf:type cs:painting_event ;
   cs:agent persons:A.Gallen-Kallela ;
   cs:time "1901" ;
   cs:place places:Helsinki .
```

Information about the artist's birth and death dates can be transformed in a similar manner into birth and death events, respectively. In this way, we can not only eliminate various time-related properties from the knowledge base but also aggregate knowledge form different sources on the more primite knowledge representation level. In this case, for example, event-based biographical knowledge about the life events of A. Gallen-Kallela can be enriched with the knowledge about the paintings he painted.

Solving the semantic interoperability problem of metadata schemas by using a primitive event-based knowledge representation scheme was one of the major challenges in creating the CULTURESAMPO II prototype in 2006. This idea will be described, especially from the semantic browsing viewpoint, in more detail in (Ruotsalo and Hyvönen, 2006).

5 Extending Semantic Representational Power of Metadata Schemas

The idea of using event-based knowledge representations in annotation provides also a solution for creating semantically richer annotations. Event-based annotations have been studied before, e.g., in the context of annotating the subject of photographs (Schreiber et al., 2001) and in representing narratives (Zarri, 2003).

To illustrate this idea, re-consider the painting "Kullervo departs for the war" of figure 2. The subject of content is here annotated by a set of keywords (in Finnish) including "Kullervo", "horse" and "dog". A problem from the knowledge representation viewpoint is that the mutual relations of the subject annotations are not known. For example, it is not known whether Kullervo rides a horse, a dog, both of them, or none of them. It is also possible that the dog rides Kullervo, and so on. Events can be used for elaborating the description, if needed, by specifying values for their thematic roles. In this case, for example, Kullervo would be in the agent role and the horse in the patient role in a riding event. This kind of information can be essential when searching the contents (e.g. to distinguish between riders and riding entities) or when providing the end-user with semantic links and explanations (e.g. to distinguish links to other riding paintings in contrast to other horse paintings).

In CULTURESAMPO content comes not only in different forms but is also annotated at different levels of detail "vertically". For example, the metadata from a museum database is given as it is and may contain only minimal metadata while some other content may be described in a very detailed manner by using lots of Iconclass notations or manually annotated events. In our case, detailed semantic descriptions are being created, for instance, when translating the Kalevala story into RDF and OWL. Here each Kalevala part of potential interest to the end-user is annotated in terms of events, thematic roles and other metadata. Furthermore, the events may constitute larger entities and have some additional semantic relations with each other. In CULTURESAMPO I this idea was experimented by representing processes and events of two Kalevala poems, in paintings, photos, and cultural processes (Junnila et al., 2006).

In CULTURESAMPO II this work continues with a new modified event-based model. Furthermore, in the new scheme, annotations can be given at three levels of granularity in order to enable vertical interoperability:

Keywords In some cases only keywords are available as subject metadata. At this level the annotation is a set of literal values. Even if ontological annotations have been used (cf. below), literal keywords may be needed for free indexing words.

Keyconcepts Here the annotation is a set of URIs or other unique references to an ontology or a classification system, such as Iconclass. The additional knowledge introduced by keyconcepts w.r.t. using literal keywords is their ontological connections. This enables semantic interoperability, as discussed earlier.

Thematic roles At this level thematic roles between activities and other entities can be specified. The additional knowledge w.r.t. using only keyconcepts is the distinction of the roles in which the keyconcepts are at different metadata descriptions.

Each new level of annotation granularity only adds

new information with respect to the previous level. This means that semantically richer representations can be easily interpreted at the lower level. Event-based descriptions mean at the keyconcept level that only the entity resources that are used in the events are considered, not the properties. At the keyword level, only the literal labels of the annotations at the keyconcept level are considered. This strategy enables, e.g., application and integration of traditional text-based search methods with ontological annotations—a useful feature since much of the content in semantic portals is in textual form in any case (e.g., free text descriptions of collection items, biographical articles, poems etc.).

The main ontology underlying CULTURESAMPO II is the General Finnish Upper Ontology YSO (Hyvönen et al., 2005b) of about 20,000 concepts. This lightweight ontology has been created based on the widely used General Finnish Thesaurus YSA[19]. CULTURESAMPO also makes use of extended versions of the ontologies used in MUSEUMFINLAND.

6 The Portal

CULTURESAMPO II provides the end-user with semantic search and browsing facilities in a way similar to MUSEUMFINLAND. Semantic multi-facet search can be used. Since the ontological model is event-centric, the user is provided with a view classifying verb-like event concepts in addition to more traditional views (persons, collections, etc.). Figure 5 illustrates the search interface.

When a search object is selected to viewing, recommended semantic links with explanations are provided for browsing. Also here the event-centric model is evident: most recommendations are based on sharing events and roles. Figure 6 shows a search object page of a photograph for illustration.

In addition, CULTURESAMPO II includes many new forms of semantic visualization, especially w.r.t. geographical information and time lines (Kauppinen et al., 2006). For visualizing search results on the map, Google Maps[20] service is used (cf. figure 7). It will be used as a search interface, too, later on. In the same vein, the Simile Time Line[21] has been incorporated in the user interface using Ajax-technology (cf. figure 8.

CultureSampo I was implemented on our old OntoViews architecture, based on Apache Cocoon[22]. However, when adding many more cross-linked components to the system in CULTURESAMPO II, such as the timeline, map views, and the new recommendation system, severe limits in the old architecture became apparent.

A major guideline in our work has been to create applications that can be configured to work with a wide variety of RDF data. To accomplish this, we have endeavored to build our applications out of modular components that combine to provide advanced functionality. As CULTURESAMPO II became more complex and started to incorporate components from other projects, there appeared a need for making the individual components smaller and supporting a more complex multidirectional control and configuration flow between them. Apache Cocoon, however, is based on a generally sequential pipeline architecture, which is very limited in its ability to support any multidirectional communication. And while it was possible to make use of modular component libraries on the Java level, there was no architectural support for keeping these components either universal or configurable, which in general resulted in them not being such.

To solve these problems, a new architecture was developed for CultureSampo II based on the well-known Service Oriented Architecture, Inversion of Control and Dependency Injection principles. Specifically, the new platform was built on top of the Apache HiveMind[23] services and configuration microkernel.

7 Discussion

Our work on CULTURESAMPO suggests that using event-based annotations is a promising approach to creating cross-domain semantic portals for several reasons:

1. By using more accurate semantic descriptions semantic search and browsing (recommendation) can be made more accurate and explained in more detail. The semantic accuracy of annotations can be extended in a natural way by the new layer of relational event annotations that explicate the thematic roles between activities and other entities in the description. First tests on CULTURESAMPO I and II seem to indicate that this kind semantic knowledge is vital for semantic information retrieval tasks (search) and for creating insightful semantic linking of contents

[19] http://www.vesa.lib.helsinki.fi
[20] http://maps.google.com/
[21] http://simile.mit.edu/timeline/
[22] http://cocoon.apache.org/
[23] http://jakarta.apache.org/hivemind/

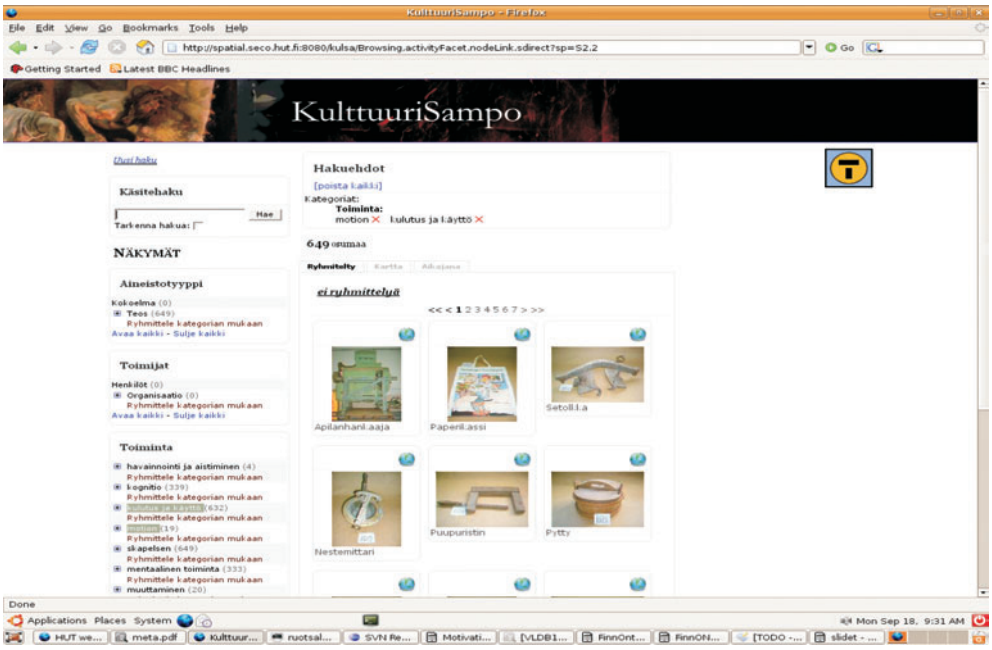

Figure 5: CULTURESAMPO II search page. Views are on left and hits on the right.

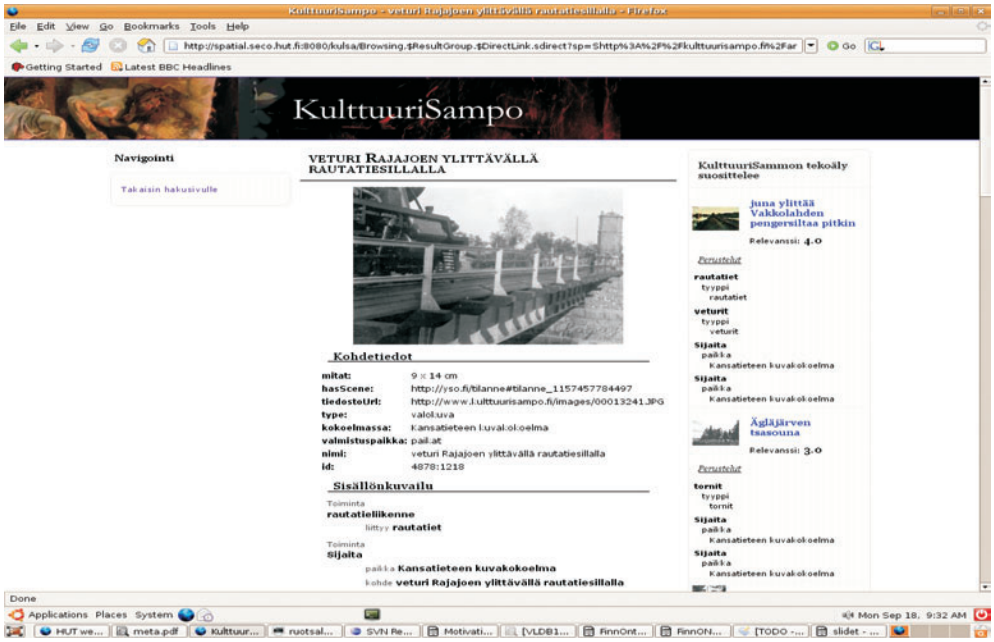

Figure 6: CULTURESAMPO II item page. Metadata is on the left and recommendation links on the right.

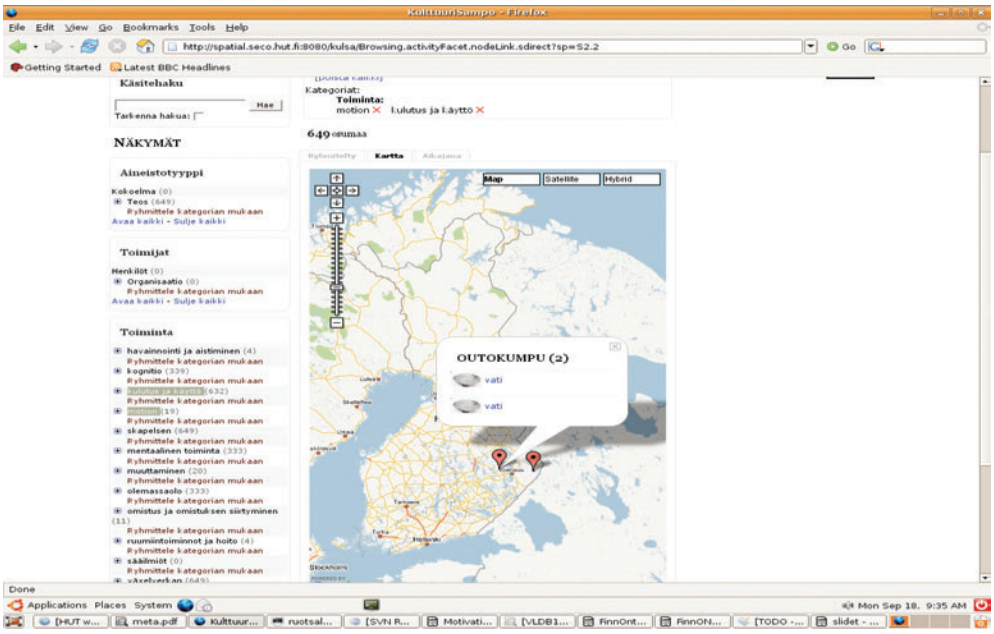

Figure 7: Using Google Maps in CULTURESAMPO II for visualizing search items on the map. The items are positioned based on a place ontology, and interactive to obtain additional information.

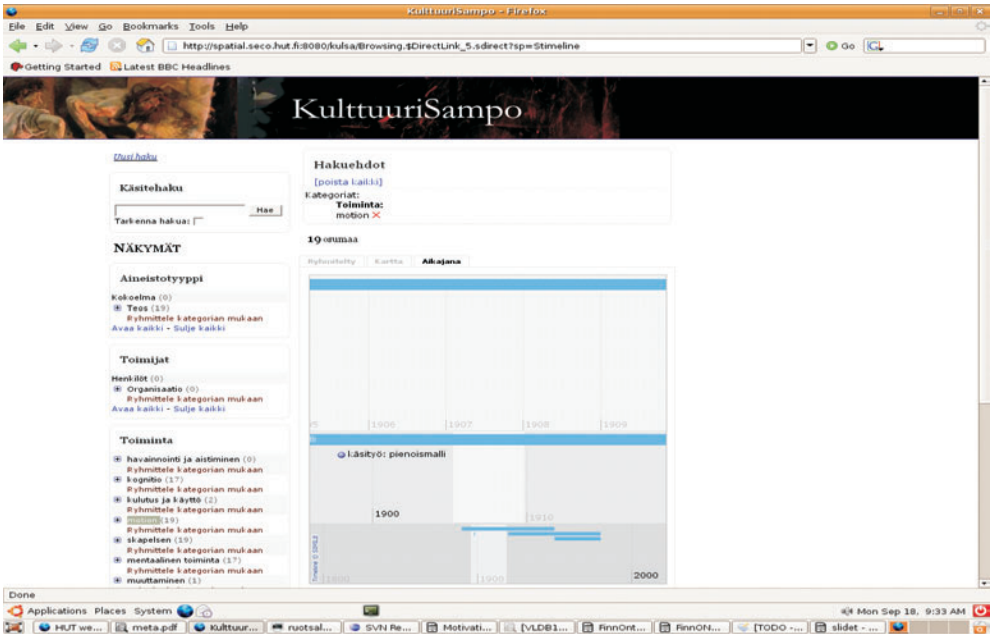

Figure 8: Using Simile Time Line in CULTURESAMPO II for visualizing search items on the time line, and for selecting them for a closer look.

automatically (Junnila et al., 2006; Ruotsalo and Hyvönen, 2006).

2. Event-based descriptions can be used for representing the meanings in terms of happenings and entities of the real world based on different metadata schemas. This enables semantic interoperability.

3. The resulting knowledge representation scheme is simpler in terms of the number of properties than the original set of metadata schemas. This makes it simpler to implement reasoning rules needed for the intelligent services for the end-user.

The price for more accurate semantics is the extra cost of creating the annotations. In CULTURESAMPO I all content was manually crafted. In CULTURESAMPO II a semi-automatic process has been used. At the schema level, the content has been enriched automatically by a separate, rule-based knowledge transformation module. This system transforms, e.g., the metadata of paintings into painting events. At the level of enriching the subject descriptions of the content, enriching has been mostly manual by adding thematic role relations between the entities used in the original databases. For example, to tell that Kullervo rides a horse and not vice versa in figure 2, a riding event with Kullervo and an instance of horse in the proper thematic roles has to be created. In principle, the machine and ontologies could help the annotator in her work, if it is known that usually humans ride horses.

The work of annotating narratives, such as the Kullervo poem in Kalevala and the process of farming by the slash and burn method in CULTURESAMPO I (Junnila et al., 2006) has been done completely manually. However, we are also investigating how language technology can be applied to creating semi-automatically annotations for textual contents (Vehviläinen et al., 2006). It is clear, that development of tools that could help in creating annotations will be of outmost importance in the future.

In some cases, like when annotating unique important materials such as Kalevala, the price for detailed annotations can be paid, while in many other cases it is unrealistic to assume that such representations will be available. In CULTURESAMPO this problem of dealing with materials annotated at different levels of semantic accuracy is addressed by using three layers of annotations: keywords, keyconcepts and thematic roles.

The success of the CULTURESAMPO will finally be judged by the end-users. Empirical usability tests are needed in order to evaluate the added value of the semantic approach. The first test, based on the current CULTURESAMPO II, has been scheduled for the autumn 2006. The goal of this experiment is to test whether the end-users really find the semantic recommendations generated by the event-based model feasible and helpful.

CULTURESAMPO II is still a research prototype and its current version contains only a few content types and less that 10,000 search objects. For example, in contrast to CULTURESAMPO I, there are no narratives in the system yet, only events. However, new types of content are being included in the scheme and in the system. Another line of development in the system is designing additional conceptual visualization tools. On the reasoning side, spatiotemporal reasoning under uncertainty is being studied (Kauppinen and Hyvönen, 2006) and is being implemented in the system.

We plan to publish CULTURESAMPO on the public web in 2007.

Acknowledgments

Our research is a part of the National Semantic Web Ontology Project in Finland[24] (FinnONTO) 2003-2007 funded mainly by the Finnish Funding Agency for Technology and Innovation (Tekes).

References

M. Doerr. The CIDOC CRM - an ontological approach to semantic interoperability of metadata. *AI Magazine*, 24(3):75–92, 2003.

E. Hyvönen, E. Mäkela, M. Salminen, A. Valo, K. Viljanen, S. Saarela, M. Junnila, and S. Kettula. MuseumFinland – Finnish museums on the semantic web. *Journal of Web Semantics*, 3(2):224–241, 2005a.

E. Hyvönen, A. Valo, V. Komulainen, K. Seppälä, T. Kauppinen, T. Ruotsalo, M. Salminen, and A. Ylisalmi. Finnish national ontologies for the semantic web - towards a content and service infrastructure. In *Proceedings of International Conference on Dublin Core and Metadata Applications (DC 2005)*, Nov 2005b.

M. Junnila. Tietosisältöjen semanttinen yhdistäminen toimintakuvausten avulla (Event-based approach to semantic linking of data content). Master's thesis, University of Helsinki, March 6 2006.

[24] http://www.seco.tkk.fi/projects/finnonto/

M. Junnila, E. Hyvönen, and M. Salminen. Describing and linking cultural semantic content by using situations and actions. In *Semantic Web at Work - Proceedings of the 12th Finnish Artificial Intelligence Conference STeP 2006, Volume 1.*, Nov 2006.

T. Kauppinen, R. Henriksson, J. Väätäinen, C. Deichstetter, and E. Hyvönen. Ontology-based modeling and visualization of cultural spatio-temporal knowledge. In *Semantic Web at Work - Proceedings of the 12th Finnish Artificial Intelligence Conference STeP 2006, Volume 1.*, Nov 2006.

T. Kauppinen and E. Hyvönen. Modeling and reasoning about changes in ontology time series. In R. Kishore, R. Ramesh, and R. Sharman, editors, *Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems*. Springer-Verlag, Dec 2006. In press.

E. Mäkelä, E. Hyvönen, and S. Saarela. Ontogator — a semantic view-based search engine service for web applications. In *Proceedings of the 5th International Semantic Web Conference (ISWC 2006)*, Nov 2006.

E. Mäkelä, E. Hyvönen, S. Saarela, and K. Viljanen. OntoViews – a tool for creating semantic web portals. In *Proceedings of 3rd International Semantic Web Conference (ISWC 2004), Hiroshima, Japan*, November 2004.

T. Ruotsalo and E. Hyvönen. Knowledge-based recommendation based on heterogenous metadata schemas, 2006. Paper under contruction.

M. Salminen. Kuvien ja videoiden semanttinen sisällönkuvailu (Semantic content description of images and videos. Master's thesis, University of Helsinki, May 2006.

A. T. Schreiber, B. Dubbeldam, J. Wielemaker, and B. J. Wielinga. Ontology-based photo annotation. *IEEE Intelligent Systems*, 16:66–74, May/June 2001.

J. Sowa. *Knowledge Representation. Logical, Philosophical, and Computational Foundations*. Brooks/Cole, 2000.

J. van den Berg. Subject retrieval in pictorial information systems. In *Proceedings of the 18th international congress of historical sciences, Montreal, Canada*, pages 21–29, 1995. http://www.iconclass.nl/texts/history05.html.

A. Vehviläinen, E. Hyvönen, and O. Alm. A semi-automatic semantic annotation and authoring tool for a library help desk service. In *Proceedings of the first Semantic Authoring and Annotation Workshop, ISWC-2006, Athens, GA, USA*, November 2006. To be published.

K. Viljanen, T. Känsälä, E. Hyvönen, and E. Mäkelä. Ontodella - a projection and linking service for semantic web applications. In *Proceedings of the 17th International Conference on Database and Expert Systems Applications (DEXA 2006), Krakow, Poland*. IEEE, September 4-8 2006.

G. P. Zarri. Semantic annotations and semantic web using nkrl (narrative knowledge representation language). In *Proceedings of the 5th International Conference on Enterprise Information Systems, Angers, France (ICEIS 2003)*, pages 387–394, 2003.