

CultureSampo: A National Publication System of Cultural Heritage on the Semantic Web 2.0

Eero Hyvönen, Eetu Mäkelä, Tomi Kauppinen, Olli Alm, Jussi Kurki,
Tuukka Ruotsalo, Katri Seppälä, Joeli Takala, Kimmo Puputti, Heini Kuitinen,
Kim Viljanen, Jouni Tuominen, Tuomas Palonen, Matias Frosterus,
Reetta Sinkkilä, Panu Paakkanen, Joonas Laitio, and Katariina Nyberg

Semantic Computing Research Group (SeCo)
Helsinki University of Technology (TKK) and University of Helsinki
first.last@tkk.fi
<http://www.seco.tkk.fi/>

Abstract. CULTURESAMPO is an application demonstration of a national level publication system of cultural heritage contents on the Web, based on ideas and technologies of the Semantic (Web and) Web 2.0. On the semantic side, the system presents new solutions to interoperability problems of dealing with multiple ontologies of different domains, and to problems of integrating multiple metadata schemas and cross-domain content into a homogeneous semantic portal. A novelty of the system is to use semantic models based on events and narrative process descriptions for modeling and visualizing cultural phenomena, and for semantic recommendations. On the Web 2.0 side, CULTURESAMPO proposes and demonstrates a content creation process for collaborative, distributed ontology and content development including different memory organizations and citizens. The system provides the cultural heritage contents to end-users in a new way through multiple (nine) thematic perspectives, based on semantic visualizations. Furthermore, CULTURESAMPO services are available for external web-applications to use through semantic AJAX widgets.

1 A Basis for Cultural Heritage on the Semantic Web

In our view, a cross-domain semantic cultural heritage portal [1] should be built on three pillars: First we need a cross-domain content infrastructure of ontologies, metadata standards, and related services, that is developed and maintained on a global level through collaborative local efforts. Second, the process of producing ontologically harmonized metadata should be organized in a collaborative fashion, where distributed content producers create semantically correct annotations cost-efficiently through centralized services. Third, the contents should be made available to human end-users and machines through intelligent search, browsing, and visualization techniques. For machines, easy to use mash-up APIs and web services should be available. In this way, the collaboratively aggregated, semantically enriched knowledge base can be exposed and reused easily as services in other portals and applications in the same vein as Google Ads or Maps¹.

¹ <http://code.google.com/apis/maps/>

CULTURESAMPO² is an operational online demonstration on a national Finnish level of implementing of such a semantic portal. The system consists of a national ontology infrastructure [2], content creation process, and an online portal [3]. In this paper, the main parts of CULTURESAMPO are first shortly overviewed. After this contributions of the work are summarized w.r.t. related work.

2 Collaborative Ontology Infrastructure

An integral part of CULTURESAMPO are the ontologies and services of the FinnONTO infrastructure [2]. The general idea here is to extend the generic, logic based W3C recommendations with domain specific ontologies in different domains. Most of the FinnONTO ontologies were developed by transforming nationally used thesauri into light weight ontologies in a semi-automatic process. In the FinnONTO model, the ontologies are developed in a distributed fashion by collaborating expert groups of different fields, and are mapped together to form a large national ontology called KOKO encompassing all domains. At the moment, KOKO includes an upper ontology YSO (20 600 concepts), a museum ontology MAO (6800 concepts), an agriforestry ontology AFO (5500 concepts), an applied art ontology TAO (2600 concepts) and a photography ontology VALO (1900 concepts). In addition, CULTURESAMPO also utilizes a geographical registry of 800 000 places in Finland, a spatiotemporal ontology of Finnish counties 1865–2007 [4], an ontology of persons and organizations, and ontologized international systems, such as the Iconclass³, and the Union List of Artists Names (ULAN)⁴.

3 Distributed, Cross-Domain Content Creation

The contents of CULTURESAMPO come from 22 museums, archives, and libraries, most of which produce their contents independently from each other using heterogeneous cataloging systems and practices. The system also contains external web contents: all Wikipedia articles (in English and Finnish) that have coordinate information, as well as related photographs from the Panoramio service can be found on CULTURESAMPO's map views.

CULTURESAMPO also has an internal Web 2.0 commenting facility by which citizens can contribute new knowledge to individual content items, e.g. identify persons in an old photograph of a museum collection. In these ways, citizens are able to contribute to the “national semantic memory”. Furthermore, interactive content production based on the SAHA editor [5] has been used in the system by participating organizations.

The system contains cultural objects of 67 different content types: artifacts, paintings, drawings, sculpture, abstract art, novels, comics, web pages, three types of folklore, five types of folk music, photos, aerial photos, persons, organizations, biographies, historical events, skills, videos, buildings, archeological sites, and others. The metadata

² The portal prototype is publicly in use at <http://www.kulttuurisampo.fi/>

³ <http://www.iconclass.nl/>

⁴ http://www.getty.edu/research/conducting_research/vocabularies/ulan/

is represented using 204 different properties and 253 additional properties used in ontologies. The content types harvested from the collections of the participating memory organizations are represented using 18 different metadata schemas. The aggregated knowledge base contains 134,000 cultural objects and 285,000 additional cultural resources (persons, places etc.), described by 2.7 million triples. The content is enriched using reasoning, that generates 8,7 million additional RDF triples in the knowledge base.

From a semantic modeling viewpoint, a research focus of our work has been event- and process-based annotations used in artificial intelligence and knowledge representation [6]. Events have been used for modeling cultural processes and narrative stories; a simplified version of the model [7] is in use.

An example of a cultural process is “Production of Ceramics” created by experts at the University of Applied Arts in Helsinki. It illustrates and explains the composition of different work phases when manufacturing ceramics. Our process model has also been applied to videos. A case example available on the portal describes how the shoe maker Onni Wirlander manufactured a pair of traditional leather boots. The video can be viewed directly starting from different points of interest. When watching the video, the recommendation system creates dynamically, for each subsequence separately, recommendation links to materials of interest in the portal, such as tools related to the sequence.

Content creation in CULTURESAMPO is supported by a number of generic tools, such as the National Ontology Library Service ONKI⁵, the metadata editor SAHA, information extraction tool POKA⁶, and the semantic content validator VERA⁷.

4 The Semantic Web 2.0 Portal CULTURESAMPO

The portal is an end-user application for both 1) human users and 2) machines, and is multilingual (Finnish, Swedish, and English). A major novelty of the portal is to provide its contents through nine “thematic perspectives”:

Map views. There are 4 map views [4] available using Google Maps: One for viewing all objects and filtering them in terms of their relation to the places; one for finding old Finnish counties with digitized limits; one for viewing historical maps layered semi-transparently over modern Google Maps ; one for finding near-by objects of interest. It is also possible to find Wikipedia articles and Panoramio photos by using the map interfaces.

Relational search. This view is a demonstration of relational search [8], where the idea is not to search for objects but associative relation chains between objects. We used the ULAN registry of 120,000 artists and organizations with 390,000 names. The user types in two names, using semantic autocompletion, and CULTURESAMPO tells how the persons or organizations are related with each other by the social network based on

⁵ <http://www.yso.fi/>

⁶ <http://www.seco.tkk.fi/tools/poka/>

⁷ <http://www.seco.tkk.fi/services/vera/>

some 50 different social roles (e.g., parent-of, teacher-of, patron-of etc.). The underlying social RDF/OWL network can also be browsed by a graphical network browser.

Search and organize. In this view, a new kind of generalized version of the faceted-search paradigm has been developed, called domain centric search [9]. This advanced search and analysis facility is used for filtering contents and for visualising the results on facets, maps or a timeline. For example, it is possible to find out how the beard fashion, as depicted in art and photographs, has changed in the 19th century in Finland.

Collections view. Here the contents can be accessed based on an organizational view. Each participating organization has an automatically generated home page in the system with links to subcollections and the actual collection items.

Finnish history view. This view is based on an ontology representing events in the Finnish history. These events are of interest of their own, but are also used to create semantic recommendations to other CULTURESAMPO contents, e.g. to biographies of persons participating in the events.

Skills and processes. This view is used for finding cultural procedural descriptions in the system, i.e. semantic models of skills, processes, and documentations on videos.

Biographies. In this view, biographies of the National Biography⁸ are used to access CULTURESAMPO contents. When reading biographies, related contents are shown based on the concepts extracted from the text using the information extraction tool POKA.

Semantic Kalevala. This view [10] contains a semantically annotated version of the national epic of Finland, Kalevala, that is related in many ways to Finnish art and culture. The epic is based on old Finnish folklore; thousands of runes and pieces of folk music are available in the portal, too. When reading Kalevala, annotations related to its subsequences can be viewed to help reading, and semantic recommendations to related materials, such as art, artifacts etc. in CULTURESAMPO are automatically produced.

Karelia view. This view contains Wikipedia articles about the Karelia area in Finland that has been influential to Finnish culture. Like in the biographies view, the POKA system is used for extracting concepts from the texts (here web pages) and for generating semantic recommendations.

In addition to humans, the system can be used by machines via an AJAX interface. In this way, the collaboratively aggregated and semantically enriched national CULTURESAMPO knowledge base can be used by other portals and systems on the web, in the same spirit as Google Ads or Maps.

5 Discussion and Conclusions

The vision and implementation of CULTURESAMPO goes beyond current semantic web portals for cultural heritage [1], such as MuseumFinland [11], MultimediaN 2006 [12],

⁸ <http://www.kansallisbiografia.fi/>

and CHIP Demonstrator 2007 [13]: The CULTURESAMPO system 1) is highly cross-domain with lots of content types and metadata schemas (usually only one schema such as Dublin Core or VRA is used), 2) it makes use of sophisticated semantic annotation models including events and processes, 3) it uses new kind of semantic search and recommendation techniques, 4) it has exceptionally versatile selection of semantic visualizations available (different map views, timelines, graphs, process visualization, semantic video viewing), 5) it is based on a large nation wide collaboratively maintained infrastructure of ontologies and ontology services, 6) it includes a model of and tools for collaborative semantic content creation, and 7) the services are available for machines, too.

Acknowledgements. This work is a part of the national FinnONTO⁹ research project 2003–2007, 2008–2010, funded mainly by Tekes and a consortium of 38 companies and public organizations, partly by the EU FP7 SmartMuseum project 2008–2010, and by the Finnish Cultural Foundation (2008–2010).

References

1. Hyvönen, E.: Semantic portals for cultural heritage. In: Handbook on Ontologies, 2nd edn. Springer, Heidelberg (forthcoming, 2009)
2. Hyvönen, E., Viljanen, K., Tuominen, J., Seppälä, K.: Building a national semantic web ontology and ontology service infrastructure—the FinnONTO approach. In: Bechhofer, S., Hauswirth, M., Hoffmann, J., Koubarakis, M. (eds.) ESWC 2008. LNCS, vol. 5021, pp. 95–109. Springer, Heidelberg (2008)
3. Hyvönen, E., Mäkelä, E., Kauppinen, T., Alm, O., Kurki, J., Ruotsalo, T., Seppälä, K., Takala, J., Puputti, K., Kuittinen, H., Viljanen, K., Tuominen, J., Palonen, T., Frosterus, M., Sinkkilä, R., Paakkari, P., Laitio, J., Nyberg, K.: CultureSampo—Finnish culture on the Semantic Web 2.0. Thematic perspectives for the end-user. In: Museums and the Web 2009 Proceedings, Archives & Museum Informatics, Toronto (2009), <http://www.archimuse.com/mw2009/papers/hyvonen/hyvonen.html>
4. Kauppinen, T., Väättä, J., Hyvönen, E.: Creating and using geospatial ontology time series in a semantic cultural heritage portal. In: Bechhofer, S., Hauswirth, M., Hoffmann, J., Koubarakis, M. (eds.) ESWC 2008. LNCS, vol. 5021, pp. 110–123. Springer, Heidelberg (2008)
5. Valkeapää, O., Alm, O., Hyvönen, E.: Efficient content creation on the semantic web using metadata schemas with domain ontology services. In: Franconi, E., Kifer, M., May, W. (eds.) ESWC 2007. LNCS, vol. 4519, pp. 819–828. Springer, Heidelberg (2007)
6. Sowa, J.: Knowledge Representation. Logical, Philosophical, and Computational Foundations. Brooks/Cole (2000)
7. Junnila, M., Hyvönen, E., Salminen, M.: Describing and linking cultural semantic content by using situations and actions. In: Robering, K. (ed.) Information Technology for the Virtual Museum. LIT Verlag (2008)
8. Kurki, J., Hyvönen, E.: Relational semantic search: Searching social paths on the semantic web. In: Poster Proc. of the ISWC + ASWC 2007, Busan, Korea (2007)
9. Mäkelä, E., Suominen, O., Hyvönen, E.: Automatic exhibition generation based on semantic cultural content. In: Proc. of the Cultural Heritage on the Semantic Web Workshop at ISWC + ASWC 2007 (2007)

⁹ <http://www.seco.tkk.fi/projects/finnonto/>

10. Hyvönen, E., Takala, J., Alm, O., Ruotsalo, T., Mäkelä, E.: Semantic Kalevala—accessing cultural contents through semantically annotated stories. In: Proc. of the Cultural Heritage on the Semantic Web Workshop at ISWC + ASWC 2007, Busan, Korea (2007)
11. Hyvönen, E., Mäkelä, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., Junnila, M., Kettula, S.: MuseumFinland—Finnish museums on the semantic web. *Journal of Web Semantics* 3(2), 224–241 (2005)
12. Schreiber, G., Amin, A., van Assem, M., de Boer, V., Hardman, L., Hildebrand, M., Hollink, L., Huang, Z., van Kersen, J., de Niet, M., Omelayenko, B., van Ossenbruggen, J., Siebes, R., Taekema, J., Wielemaker, J., Wielinga, B.J.: MultimediaN e-culture demonstrator. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 951–958. Springer, Heidelberg (2006)
13. Wang, Y., Stash, N., Aroyo, L., Gorgels, P., Rutledge, L., Schreiber, G.: Recommendations based on semantically-enriched museum collection. *Journal of Web Semantics* 6(4), 43–50 (2008)