

## Publication VIII

Eero Hyvönen, Eetu Mäkelä, Tomi Kauppinen, Olli Alm, Jussi Kurki, Tuukka Ruotsalo, Katri Seppälä, Joeli Takala, Kimmo Puputti, Heini Kuittinen, Kim Viljanen, Jouni Tuominen, Tuomas Palonen, Matias Frosterus, Reetta Sinkkilä, Panu Paakkari, Joonas Laitio, and Katariina Nyberg. 2009. 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 by authors

# CultureSampo—Finnish Culture on the Semantic Web 2.0: Thematic Perspectives for the End-user

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

We present an overview of CultureSampo, an ambitious system for creating a collective semantic memory of the cultural heritage of a nation on the Semantic Web 2.0, combining ideas underlying the Semantic Web and the Web 2.0. The system addresses the semantic web challenge of aggregating highly heterogeneous, cross-domain cultural heritage collections and other contents into a semantically rich intelligent system for human and machine users. At the same time, CultureSampo is an approach to solve the social and practical Web 2.0 challenge of organizing the underlying collaborative ontology development and content creation work of memory organizations and citizens. This paper focuses on CultureSampo's search, recommendation, and visualization services for the end-users. The key idea here is to access cultural heritage on the Semantic Web through nine "thematic perspectives", such as places on the maps, the social network of cultural persons, timelines, and narrative texts, e.g. biographies and literary works.

## 1. Publishing Collections Collaboratively on the Semantic Web 2.0

CultureSampo (<http://www.seco.tkk.fi/applications/kulttuurisampo/>) (Hyvönen et al., 2008) is a publication system and a portal by which memory organizations and citizens can publish their collections on the Semantic Web in a collaborative way. CultureSampo extends our earlier application "Museum-Finland—Finnish Museums on the Semantic Web" (Hyvönen et al., 2004, Hyvönen et al., 2005) (<http://www.museosuomi.fi/>), a system for publishing artifact collections on the Semantic Web, by sup-

porting publication of different kind of cross-domain contents, both material and immaterial. The system also introduces many novelties to both end-users and publishers.

Creating a system like CultureSampo is challenging due to two major reasons:

**Semantic challenges.** Cultural heritage content is semantically heterogeneous and available in various forms (documents, images, audio tracks, videos, collection items, learning objects etc.), concern various topics (art, history, handicraft etc.), is written in different languages, and is targeted to both nonprofessionals and experts. Furthermore, the content is semantically interlinked, as depicted in figure 1.1. For example, the content may contain a person’s narrative biography, works of art she created, places of interest where she lived in, Wikipedia articles or novels about or by the person, social connections to other persons, and events in the history that the person was related to. In our work, semantic web technologies (<http://www.w3.org/2001/sw/>) are used to address these challenges.

**Organizational challenges.** Museums, archives, libraries, media organizations, associations, and individual citizens create cultural heritage content independently from each other. This has led to a situation, where metadata produced by different organizations is usually incompatible with each other in terms of metadata schemas, vocabularies, and cataloging conventions. In our work, ideas underlying the Web 2.0 are used to support collaboration and promote interoperability in distributed content creation.

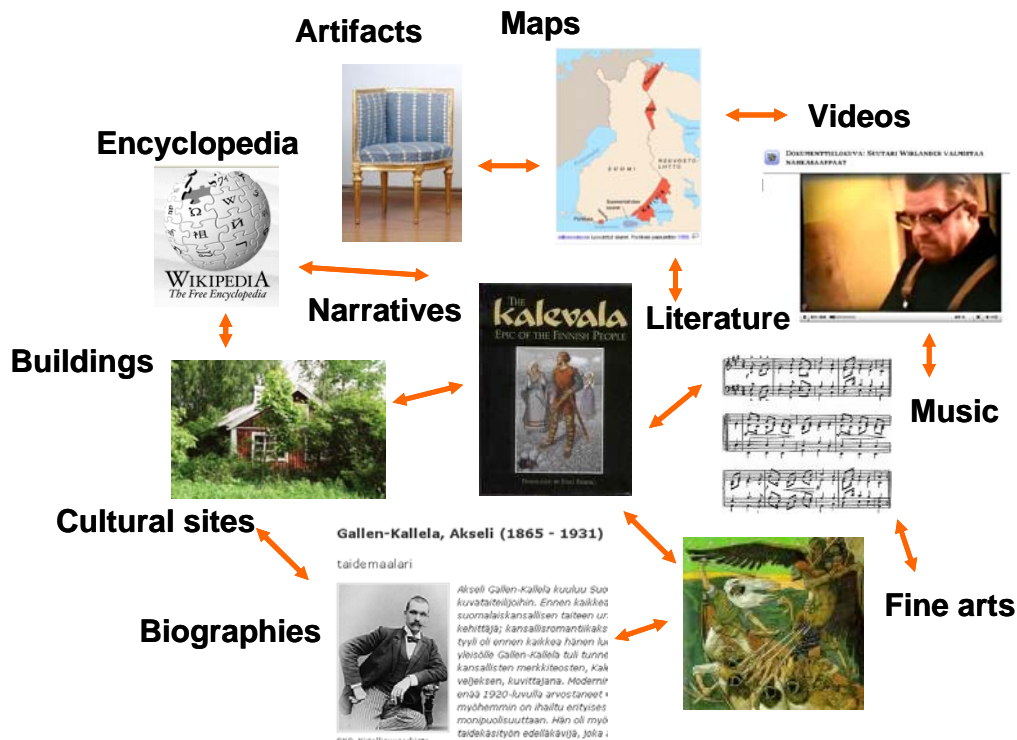


Figure 1.1 Cultural heritage is semantically heterogeneous and mutually linked.

CultureSampo has been developed since 2004 as a part the FinnONTO project (2003–2007, 2008–2010) (<http://www.seco.tkk.fi/projects/finnonto/>). The goal of this work is build a national level semantic web content infrastructure and demonstrate its usefulness in practical applications. The first public prototype of CultureSampo was published in September 2008 and can be used in three languages on the web at <http://www.kulttuurisampo.fi/>.

CultureSampo consists of three components (cf. figure 1.2):

1. **Collaborative ontology infrastructure.** The basis of CultureSampo is the national FinnONTO infrastructure (Hyvönen et al., 2008b). It includes a collaboratively created system of cross-domain ontologies and related ontology services for utilizing them cost-efficiently as services. The ontologies and

the services were published as the National Ontology Service ONKI (<http://www.yso.fi/>) on September 12, 2008.

2. **Content production system.** Our content creation model consists of a set of metadata models and a content creation process for producing and harvesting content from museums, libraries, archives and other organizations, as well as from individual citizens and (inter)national Web 2.0 sources.
3. **Semantic Web 2.0 portal.** The portal itself is unique in its use of versatile cross-domain semantic models, new semantic searching and browsing methods, and semantic visualizations for the end-users.

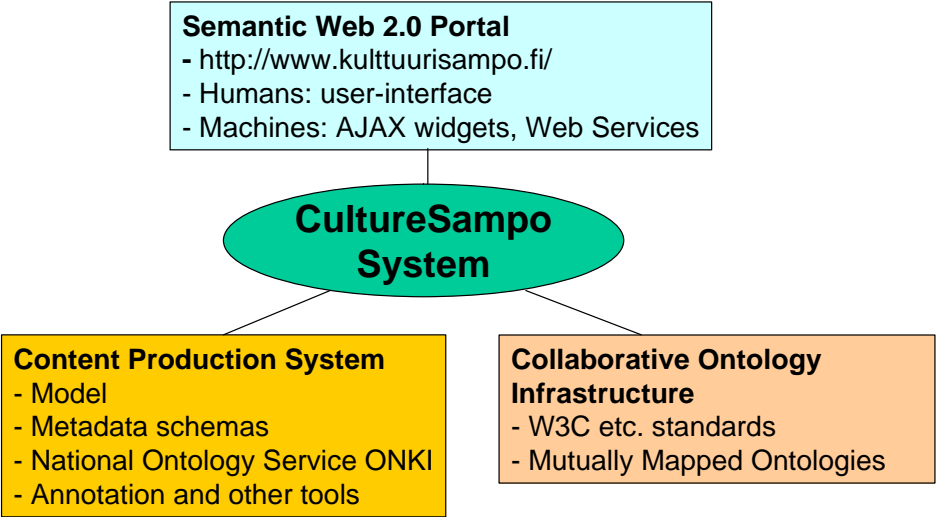


Figure 1.2 Three components of CultureSampo.

In the following, these three components are overviewed with an emphasis on the end-user interface of the portal. After this, contributions of the work to related research are summarized.

## 2. Collaborative Ontology Infrastructure

The ontologies of the FinnONTO infrastructure constitute an integral part of CultureSampo. They complement the generic, logic based W3C semantic web recommendations, such as RDF, OWL, and SPARQL, with domain specific concept descriptions in different domains. Most of the FinnONTO ontologies were developed by transforming nationally used thesauri into lightweight ontologies. The process was not purely mechanical, like e.g. in (van Assem, 2006), but also manual processing was required in order to refine the semantic thesaurus relations into full-blown subsumption hierarchies. In the FinnONTO model, the ontologies in different domains 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. KOKO includes e.g. an upper ontology YSO (20 600 concepts), a cultural heritage ontology MAO (6800 concepts), an agriforestry ontology AFO (5500 concepts), an applied art ontology TAO (2600 concepts), and a photography ontology VALO (1900 concepts). Figure 2.1 illustrates the KOKO system of mutually mapped ontologies where YSO constitutes the upper ontology.

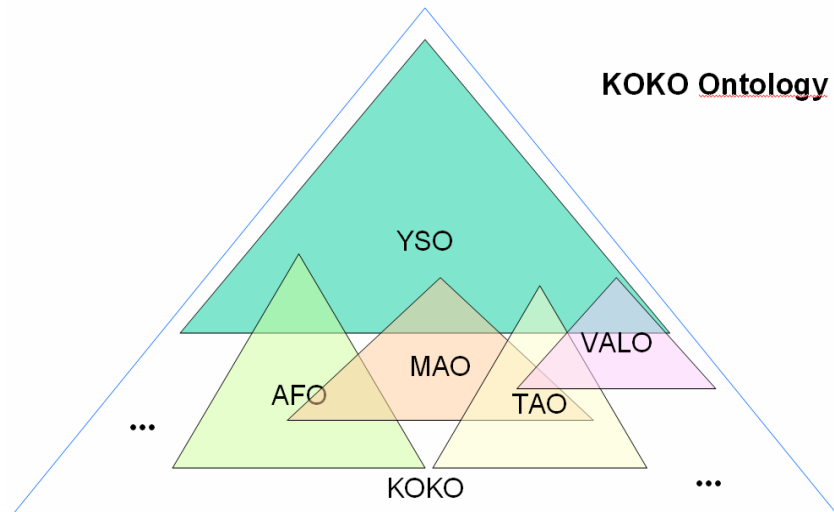


Figure 2.1. KOKO system of mutually mapped cross-domain ontologies.

The ontologies are provided to end-users not only in RDF/OWL form, as usual, but as ready to use semantic web widgets (Mäkelä et al., 2007) using Web 2.0 AJAX APIs, as well as through conventional Web Services. In addition to KOKO, CultureSampo also utilizes a geographical registry of 800,000 places in Finland, a spatiotemporal ontology of historical Finnish municipalities 1865–2007 (Kauppinen et al., 2008), and an ontology of persons and organizations. Furthermore, international vocabularies, such as the Iconclass (<http://www.iconclass.nl/>), the Art and Architecture Thesaurus (AAT) ([http://www.getty.edu/research/conducting\\_research/vocabularies/aat/](http://www.getty.edu/research/conducting_research/vocabularies/aat/)), and the Union List of Artists Names (ULAN) ([http://www.getty.edu/research/conducting\\_research/vocabularies/ulan/](http://www.getty.edu/research/conducting_research/vocabularies/ulan/)) are used.

### 3. Content Production System

CultureSampo contains cultural objects of nearly 30 different content types including artifacts, paintings, drawings, sculptures, pieces of 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, and archeological sites. These content types are represented using 18 different metadata schemas. The aggregated knowledge base contains 52,000 cultural objects and 235,000 other resources, such as ontological class concepts and place instances. The cultural objects are described by 784,000 RDF property triples.

The content is enriched using reasoning, resulting in some 10 million property triples. The enriched knowledge base is used for intelligent information retrieval and for creating semantic recommendation links between objects. The content is represented using RDF and OWL, and SPARQL is used for querying recommendations. The system also utilizes external web resources through web services: all Wikipedia articles in English and Finnish that have coordinate information, as well as photographs of the Panoramio service Panoramio (<http://www.panoramio.com/>) can be found on CultureSampo's map views.

These information sources have diverse ownerships. The contents come from 22 Finnish museums, archives, and libraries, most of which produce their contents independently from each other using heterogeneous cataloging systems and practices. Wikipedia and Panoramio content is created (inter)nationally by the public. CultureSampo also has a commenting facility by which individuals 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, distributed content production based on the SAHA editor (Valkeapää et al., 2007) has been used internally in the system by the participating organizations.



Figure 3.1. CultureSampo system in a nutshell.

Figure 3.1 depicts the CultureSampo system as a whole. In the center is the KOKO ontology and other infrastructure ontologies. The collection items (cf. figure 1.1) around the ontologies are attached to the ontologies by metadata. The content providers depicted around the circle, i.e. CultureSampo, publish metadata locally and independently using shared metadata schemas and ontologies. The result is a large global semantic RDF network linking different contents together in ontologically meaningful ways.

From a semantic modeling viewpoint, one research focus of our work has been event- and process-based annotations used in artificial intelligence and knowledge representation (Sowa, 2000). In our case, events have been used for modeling cultural processes and narrative stories (Junnila et al., 2008) and for metadata schema integration (Ruotsalo, Hyvönen, 2007). The KOKO ontology was designed to support this by clearly separating events and processes from other concepts as in Dolce (Gangemi et al., 2002)

In some metadata schemas of CultureSampo it is possible to annotate content using processes in terms of events, subevents and their sequences. The model in use in the prototype is a simplified version of our earlier schema (Junnila et al., 2008). The portal then automatically generates an interactive representation of the process as a kind of a temporal table of contents. This system is used in the prototype for creating skill models, cultural process models, and documentation of processes in videos:

**Semantic skill models.** An example of a skill model is the model “Production of Ceramics” produced by experts at the University of Applied Arts in Helsinki. It illustrates and explains the composition of different work phases when manufacturing ceramics. At each phase, semantic recommendations to other relevant CultureSampo contents can be created automatically. For example, links to products in collections that were manufactured using the same techniques, are automatically obtained.

**Semantic models of cultural processes.** There is a similar kind of chronological model “A Year on a Farm” of the seasonal farming events and processes taking place at a typical farm in Finland. Again, tools and other materials from CultureSampo are linked automatically with related events. This model was created by a farmer employed at the Finnish Museum of Agriculture. The permanent exhibition of this museum is actually organized using the same idea of presenting farming events taking place during different yearly seasons.

**Semantic documentation on videos.** The annotation model can be applied also to documenting instances of actual skill events or processes documented on a video. The case example available on the

portal describes how the shoemaker Onni Wirlander manufactured a pair of traditional leather boots. The video was produced and the annotation created by the Espoo City Museum using the SAHA editor connected to the ONKI ontology services. The semantic model describes what happens at different (sub)sequences on the video. Semantic search can find not only the video as a black box, as in systems such as YouTube, but also points of interest inside the video. The video can be viewed directly starting from different points of interest. This is important with longer videos.

The examples above demonstrate our new ideas of representing and storing immaterial, procedural cultural heritage in the memory system, here descriptions of handicraft skills and cultural processes. Typical cultural heritage portals contain metadata only about static objects such as artifacts.

Content creation in CultureSampo, both ontologies and the metadata, is based on distributing the work to organizations and citizens in a Web 2.0 fashion. In this strategy, extra costs can be minimized by reorganizing the work done already in the organizations and in public. The work is supported by a number of generic FinnONTO tools, such as the metadata editor SAHA, information extraction tool POKA (Valkeapää et al., 2007), and the semantic content validator VERA (<http://www.seco.tkk.fi/services/vera/>).

**Nine perspectives into cultural heritage**

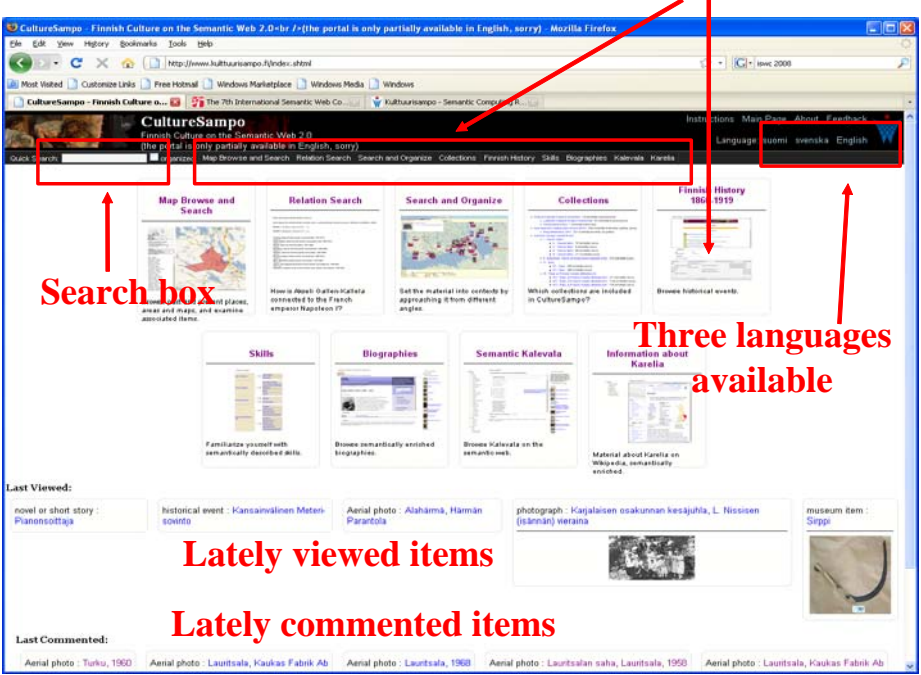


Figure 4.1 Front page of CultureSampo.

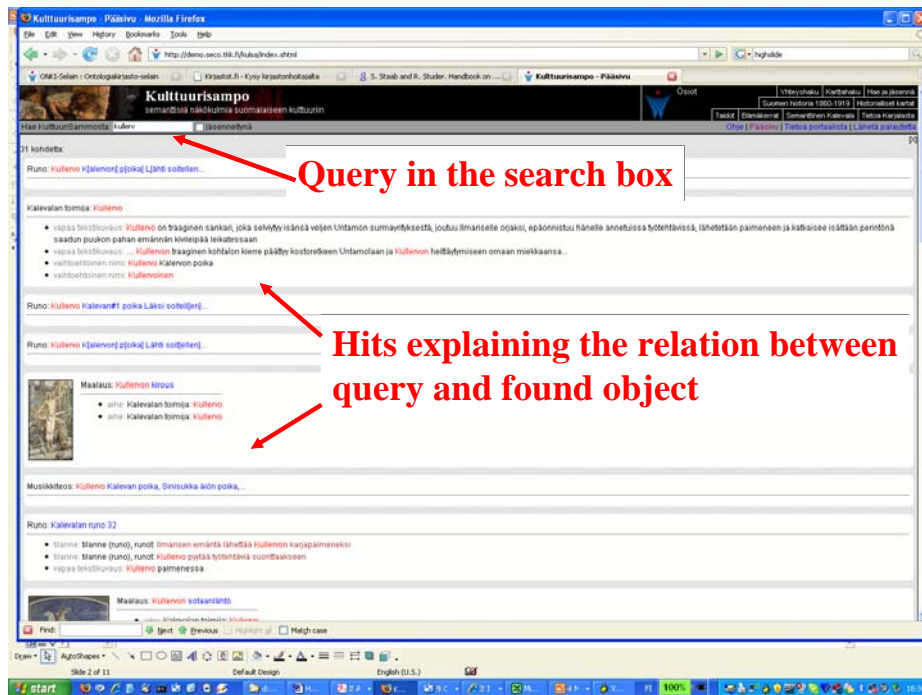


Figure 4.2 Using the semantic search box.

#### 4. Semantic Web 2.0 Portal: Thematic Perspectives

Figure 4.1 depicts the front page of the portal CultureSampo. The system is multilingual: Finnish, Swedish, or English can be selected by the links in the upper right corner. However, because nearly all contents are in Finnish and translations of some parts of the Finnish KOKO ontologies are not available, the system is not equally powerful in other languages. On the bottom of the page, lately viewed objects from the knowledge base are shown, as well as objects that have been commented lately by end-users.

In the upper left corner, there is a Google-like search box for typing in a search query. CultureSampo utilizes semantic autocompletion (Hyvönen, Mäkelä, 2006) in order to guess the possible query words that the user is aiming at. Semantic query expansion based on the ontologies is used in order to enhance recall. Furthermore, the underlying ontologies are used to organize the hit results into meaningful categories. For example, in figure 4.2 the user is typing into the search box “Kullervo...” aiming probably at the hero “Kullervo” of the Kalevala epic. The hit results are categorized by the roles connecting the hero and the matched objects, e.g., paintings depicting him, runes telling about him etc.



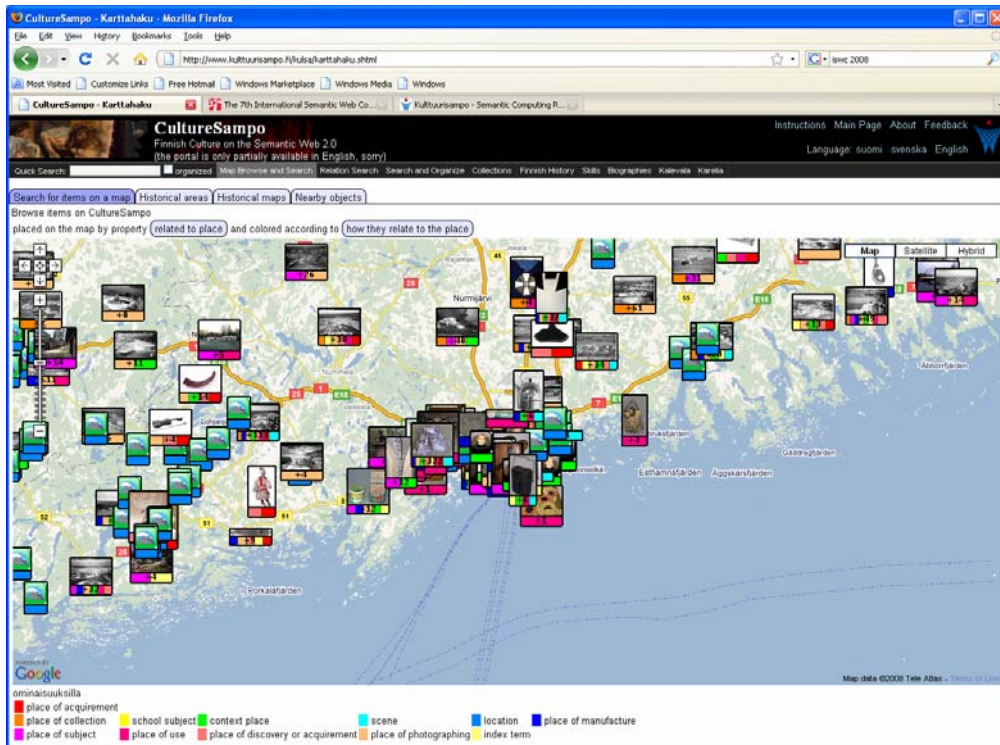


Figure 4.3 Collection objects on the map based on 12 different spatial relations.

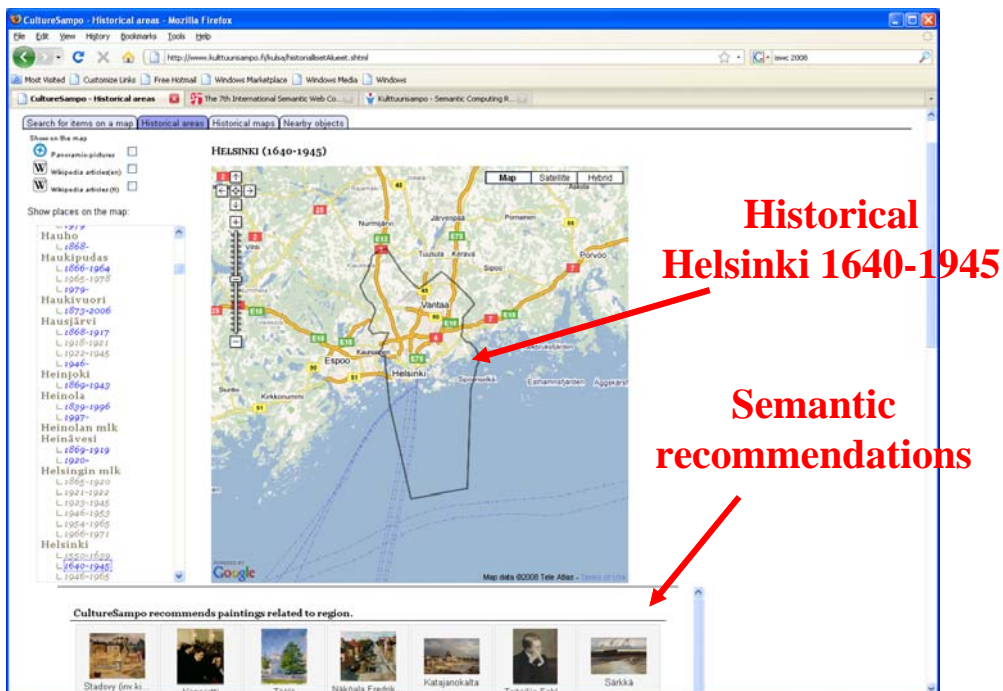


Figure 4.4 Historical areas tab.

A major novelty of the portal is to provide the end-user with an access to the cultural RDF knowledge base through nine “thematic perspectives”. They are available at the main entry page and as choices in the menu bar (cf. figure 4.1 in the middle). These perspectives are overviewed below in order to illustrate the underlying ideas and functionalities available. New perspectives can be added into the system fairly easily, because the underlying architecture is based on a set of general service components, such as the semantic recommending system.

## Perspective 1: Maps Search and Browse Views

There are four map views [14] available using Google Maps. Each view has its own tab:

1. The tab “Search for Items on the Map” shown in figure 4.3 displays any collection object with coordinate information on Google Maps, and tells the semantic relations of objects to the places. There are 12 different spatial relations in use, such as “place of acquirement”, “place of subject”, and “place of manufacture”.
2. The tab “Historical Areas” is used for finding old Finnish counties with their digitized limits on the maps, based on the spatiotemporal Finnish place ontology SAPO [23]. In figure 4.4, a link directory to old places is shown on the left, and the user has selected the historical Helsinki. Semantic recommendation links to related collection items are automatically shown on the bottom of the page with short explanations about their relation to the place.

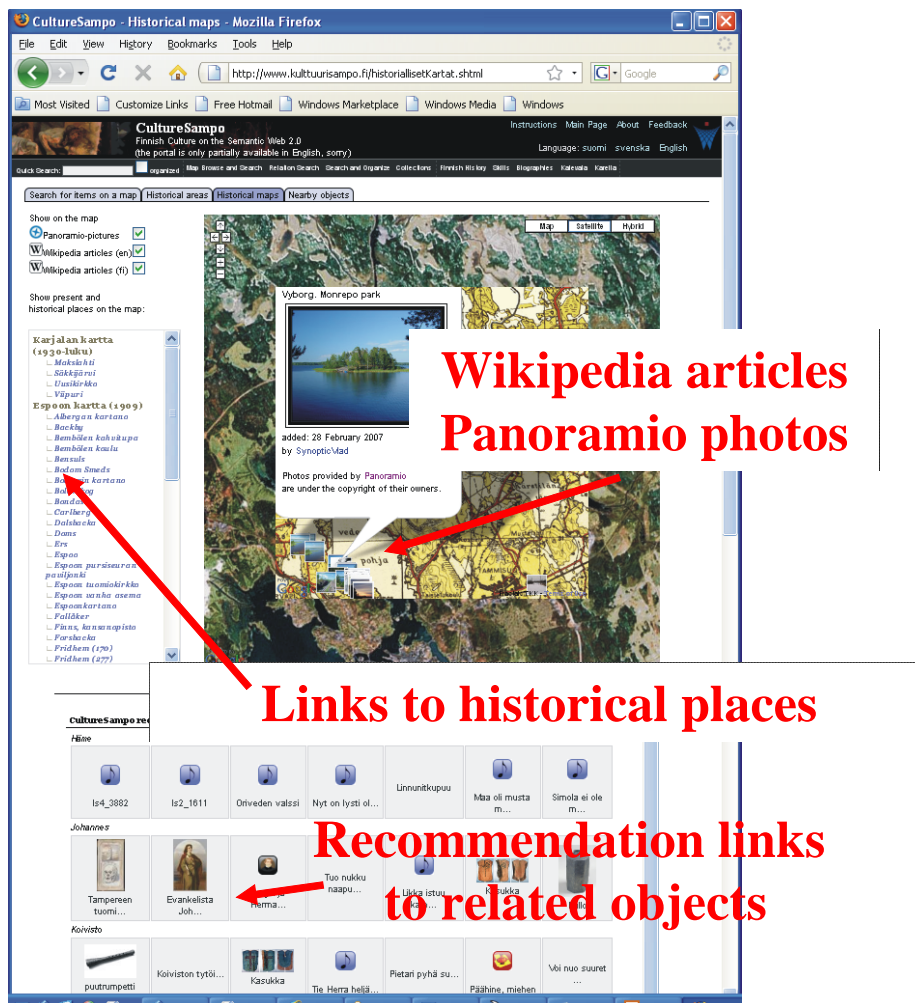


Figure 4.5 Historical maps on Google Maps.

3. The tab “Historical Maps” [14] is used for viewing historical maps layered semi-transparently over modern Google Maps. Wikipedia articles in Finnish or English with coordinate information, and photos from the Panoramio service can be seen on the maps and can be opened by clicking on them. For example, in figure 4.5 old Finnish Karelian maps are viewed semi-transparently on top of modern Russian Google Maps [23] (this area is part of Russia today). The user has located the historical Finnish park “Monrepos”, and found a modern Panoramio photograph there, taken by a contemporary Russian citizen. On the left, there is also the index of old Finnish places on the maps as direct selection links. Semantic recommendations to collection items related to historical places on the maps are displayed

below. In this case, for example, links to pieces of folk music and poems collected from the region can be seen, as well as related paintings, artifacts, and old aerial photographs.

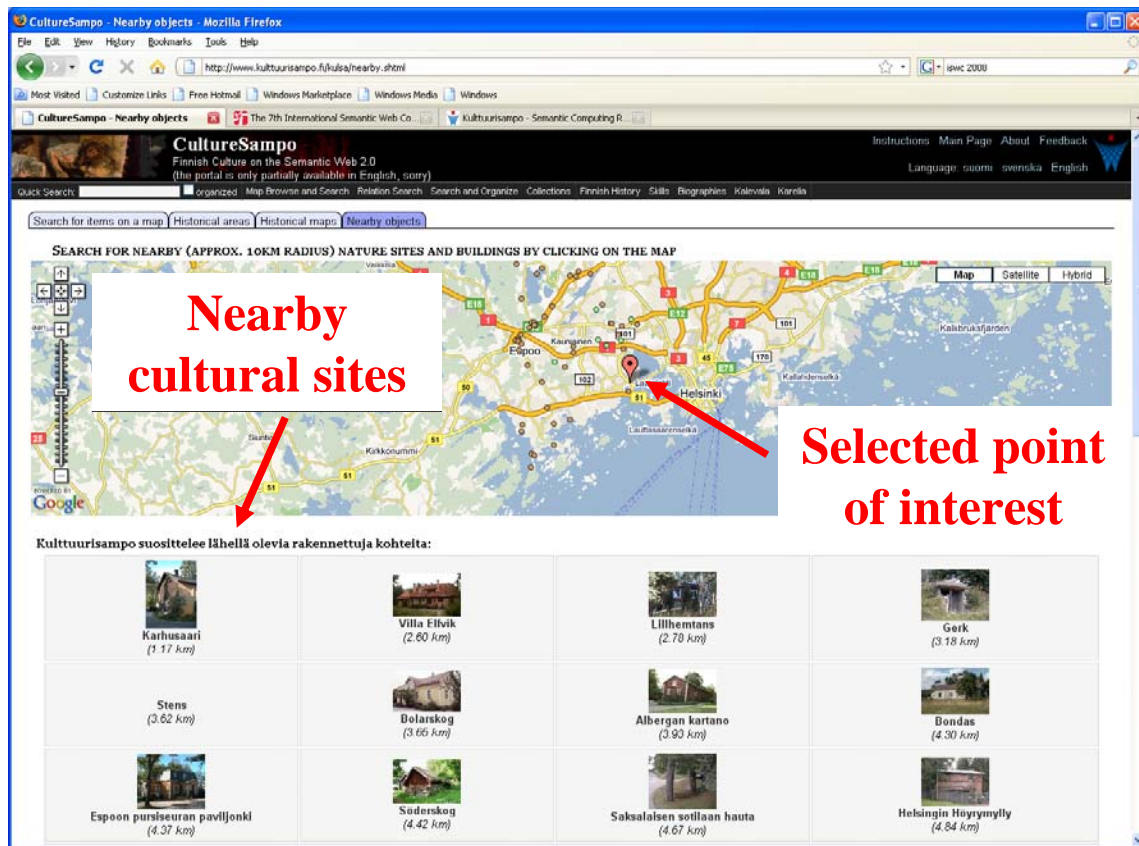


Figure 4.6 Finding nearby cultural sites.

- The fourth tab "Nearby objects" is used for finding nearby objects of interest: the user clicks a point and the system finds nearby cultural sites of interest (cf. figure 4.6).

## Perspective 2: Relational Search

The relational search perspective is a demonstration of relational search [31] also called association identification or knowledge discovery [32]. Here 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. Here the user can type in two names, using semantic autocompletion, and CultureSampo tells how the persons or organizations are related to each other by the social network based on 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. For example, in figure 4.7 the user typed in Napoleon I (the French emperor) and Akseli Gallen-Kallela (a Finnish artist), and CultureSampo found a social path of 7 steps between the persons. The browsable social network of Napoleon I is depicted on the right hand side window.

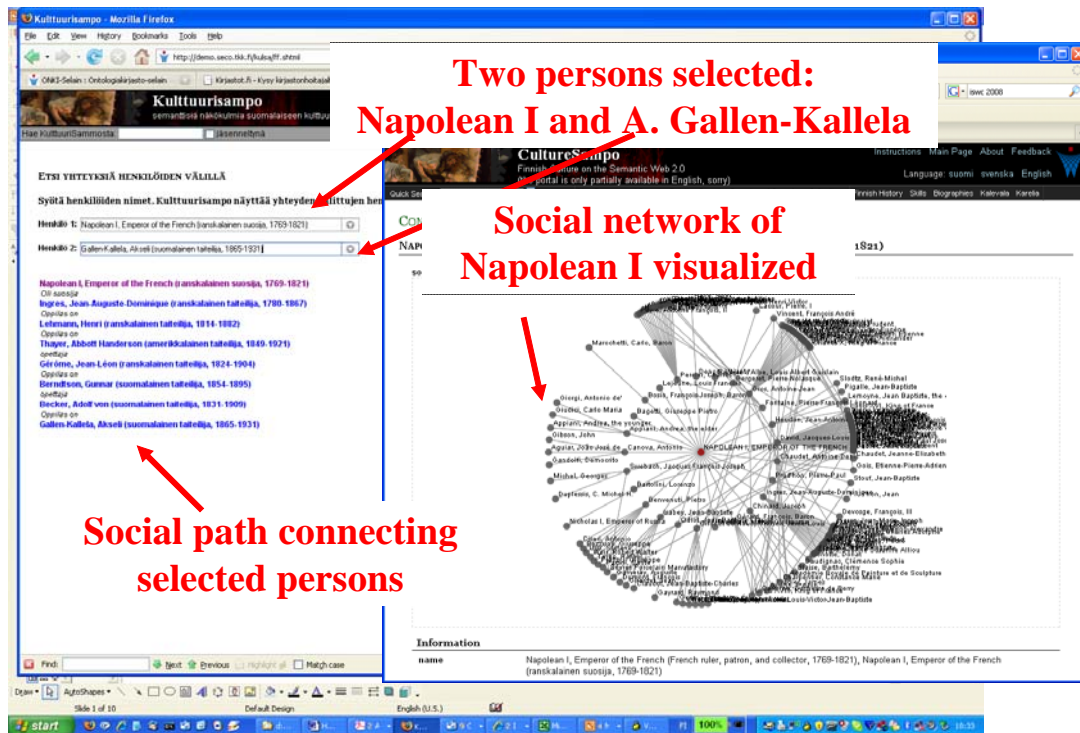


Figure 4.7 Answering the question "How is Napoleon I, the French emperor, related to Akseli Gallen.-Kallela, the Finnish artist?" by relational search.

### Perspective 3: Search and Organize

In the search and organize perspective, the idea is to move beyond locating interesting singular items. Using the tools of this perspective, it is possible to analyze the contents and create organized presentations that are able to bring out interesting trends and information contained in the data as a whole. For example, it is possible to find out what were the most popular themes in Finnish fiction in the year 2007. In figure 4.8, the end-user is analyzing how beard fashions have changed during the ages by first searching for collection items depicting beards in different ways, and then projecting the results on a timeline. In figure 4.9, images of churches are visualized on the map in order to investigate the geographical distribution of different kind of churches in Finland.

The user interface of the perspective is divided into two functional parts. On the top, constraints for the result set are specified and changed. For example, in figure 4.0, the user has formulated the query "Tell me about photographs related to the keyword church".. On the bottom, the constrained result set can be organized, grouped and visualized according to different ontological facets as lists, on a map, or on a timeline. For example, it is possible to view the results "as a list according to the most popular theme", "according to time of manufacture on a timeline" or "colored according to style and placed on a map according to place of manufacture". For both query constraining and organizing, the perspective makes use of domain-centric faceted search (Mäkelä et al., 2007), a generalization of the faceted search paradigm to heterogeneous data.

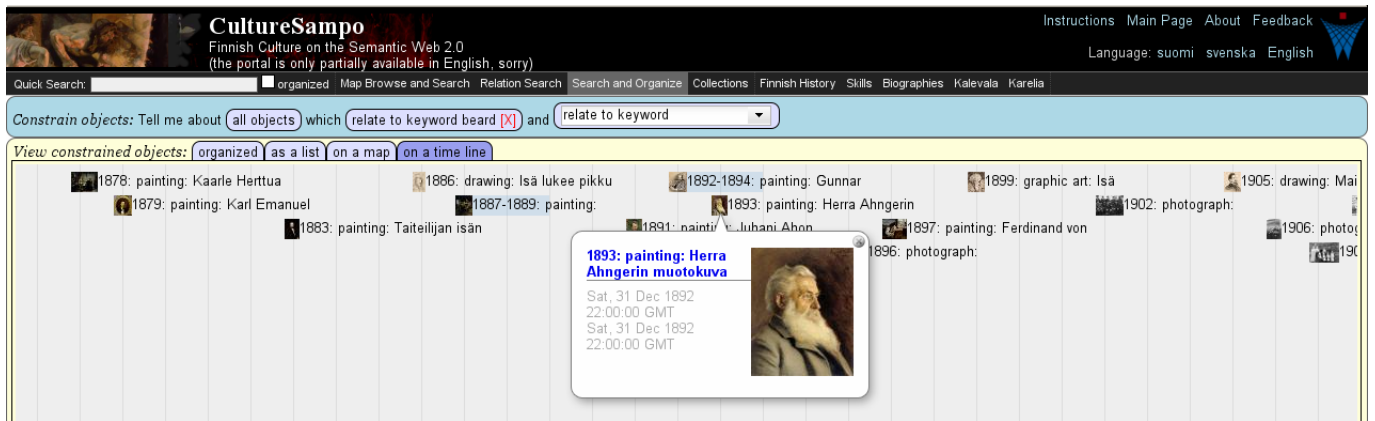


Figure 4.8 How have beard fashions changed during the ages? Paintings, photographs and other visual objects relating to the keyword 'beard' rendered on a timeline.

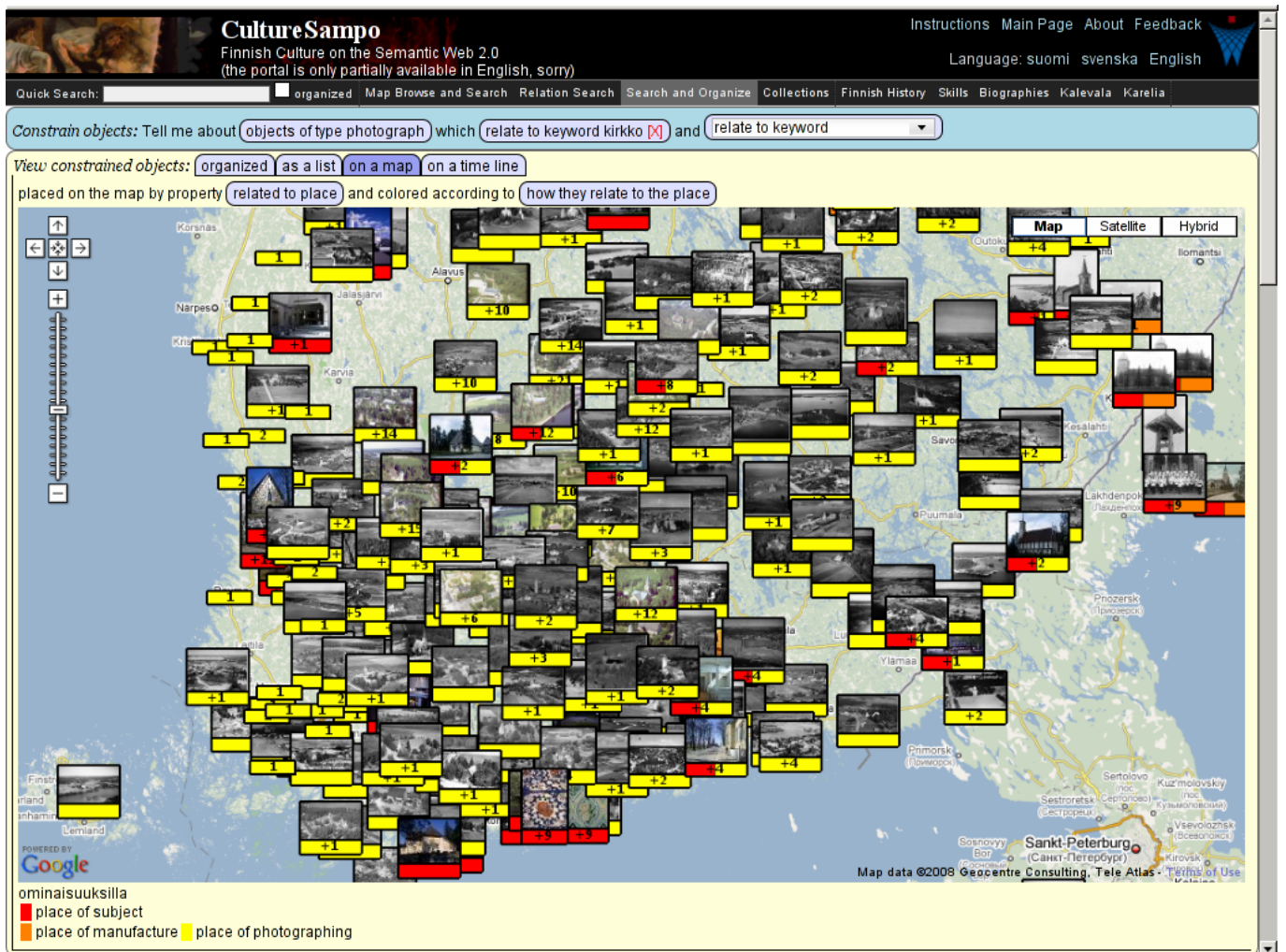


Figure 4.9 Is Finland a land of churches? Photographs relating to the keyword 'kirkko' (church in Finnish) rendered on a map.

#### Perspective 4: Collections

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 its collections and the actual collection items.

## Perspective 5: Finnish History

This view is based on an ontology representing events in the Finnish history [33]. These events are of interest on their own, but are also used to create semantic recommendations to other contents, e.g. to biographies of persons participating in the events. The history ontology in the prototype system contains 220 events on a timeline in 1860–1919. The content originates from the Agricola portal (<http://agricola.utu.fi/>) maintained by history researchers.

The screenshot shows a web browser window displaying the CultureSampo application. The main content is a video player titled "DOCUMENTARY FILM: SUUTARI WIRLANDER VALMISTAA NAHKASAAPPAAT". The video player has a progress bar and a volume control. To the left of the video player is a "Table of Contents" for the video, listing various stages of the shoemaking process with corresponding timestamps. To the right of the video player is a "Semantic process description" box, which is a white box with red text. Below the video player is a "Dynamic information about the video scene" box, which is a white box with red text. To the right of the video player is a "Semantic recommendations" box, which is a white box with red text. The video player shows a person working on a shoe. The "Table of Contents" lists the following items:

- 0:20:40 Pikalangan valmistus
- 0:45 Langan kerääminen
- 1:30 Langan pikaaminen pikilapulla hankaamalla
- 2:40-3:26 Puunsuojan valmistus
- 2:40 Naulalastujen vuoleminen
- 7:14 Nautilan vaihtaminen laatuista

The "Semantic process description" box contains the text: "Semantic process description". The "Dynamic information about the video scene" box contains the text: "Dynamic information about the video scene". The "Semantic recommendations" box contains the text: "Semantic recommendations".

Figure 4.9 Semantic video viewing with dynamic recommendations.

## Perspective 6: Skills and Cultural Narratives

Figure 4.9 depicts the idea of documenting traditional skills semantically on videos, and providing the documentation through a semantic video viewer. The use case here is the shoemaker Wirlander producing a pair of leather boots. This process has been described semantically as a sequence of hierarchical events that take place in the video in different parts. The corresponding “table of contents” is automatically generated for the end-user on the left. The user can then view any part of the video by selecting items from the list.

In the screenshot, the user is viewing the part “sharpening of the knife”. At the same time, the system recommends items of interest in the collections, in this case different knives belonging to the collections of the museums. Each part of the video that forms a meaningful entity is a search object of its own, and can be found through the general search engine of the system.

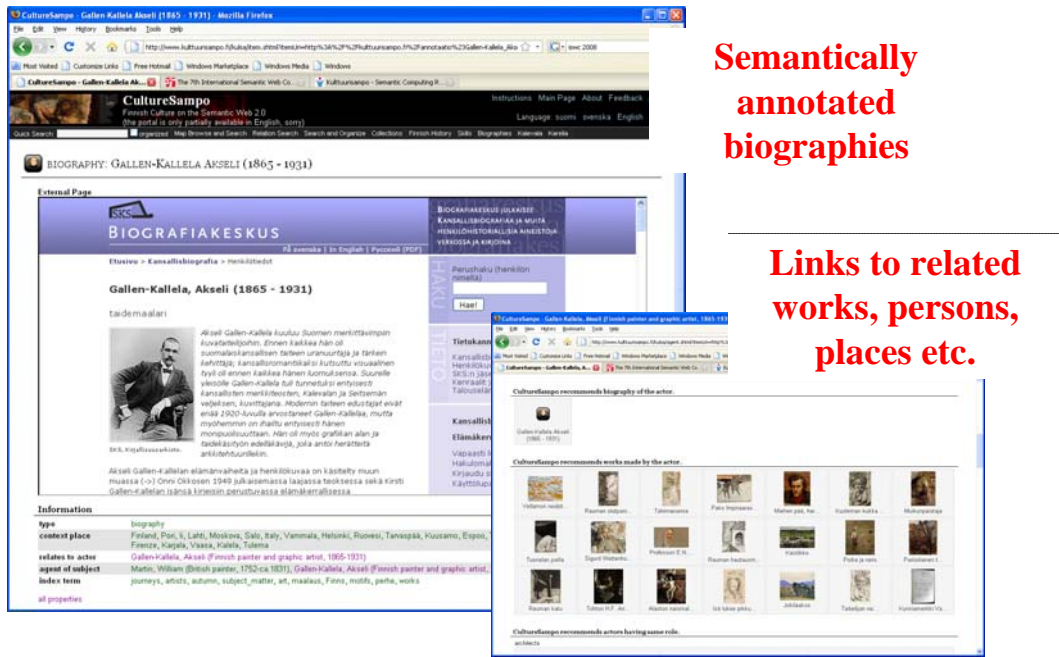


Figure 4.10 Biographical perspectives to cultural heritage.

### Perspective 7: Biographies

In this view, biographies of the National Biography are used to access CultureSampo contents. When reading the biographies, related contents are shown as recommendations based on the concepts extracted from the text using the information extraction tool POKA (<http://www.seco.tkk.fi/tools/poka/>). Figure 4.10 depicts the situation where the user is reading the biography of Akseli Gallen-Kallela, retrieved from the server of the Finnish Literature Society (SKS), and provides her with links to e.g. the art works of the artist as well as biographies of other persons related to his life.

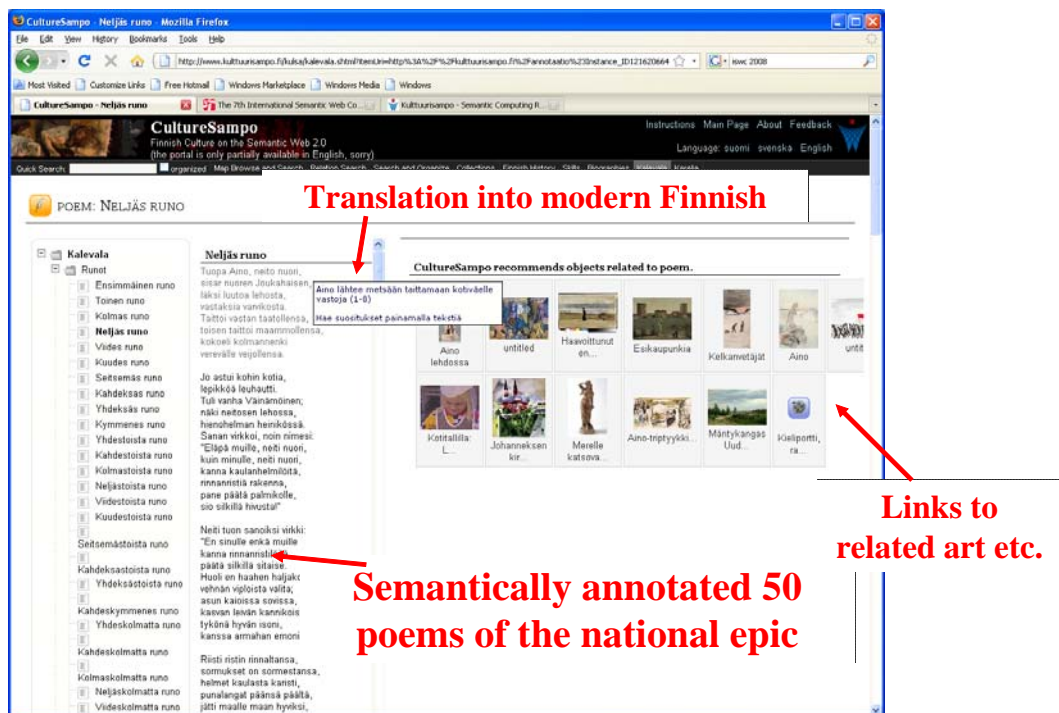


Figure 4.11 Semantic Kalevala epic.

## Perspective 8: Semantic Kalevala

This view 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 also has interesting links to old Finnish folklore, on which it is actually based, and to folk music lyrics. Thousands of runes and pieces of folk music are available in the portal. When reading Kalevala, annotations related to its subsequences can be viewed to help reading, and semantic recommendations to related materials in CultureSampo are automatically produced [17].

For example, in figure 4.11 the user is reading a part of the fourth rune in Kalevala. A part of the rune is selected, and a modern Finnish explanation of the part can be viewed. Furthermore, the system is suggesting semantic links to objects related to the part selected, in this case e.g. to sculpture depicting the fictive person Aino mentioned in the selected poem part.

## Perspective 9: Karelia

This last thematic perspective contains Wikipedia articles about the Karelia area in Finland that has been influential to the Finnish culture. For example, lots of folklore has been collected into national archives from this area, and the Kalevala epic is strongly associated there. Like in the biographies view, the POKA system is used for extracting concepts from the texts (here web pages). Based on the extracted concepts, semantic recommendations to related contents are created for additional information (cf. figure 4.12).

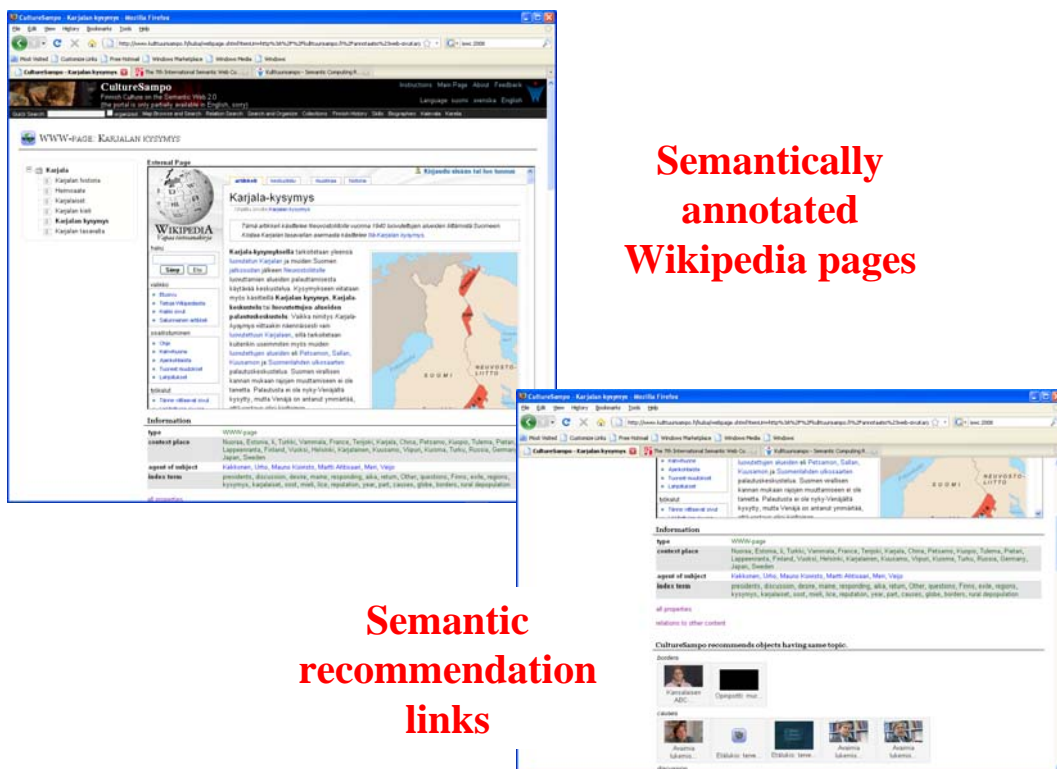


Figure 4.12 Karelia perspective.

In addition to human end-users, the system can be used by machines via AJAX interface and widgets. The application use case here is that collaboratively aggregated and semantically enriched national knowledge base can be used not only by the CultureSampo portal but also by other portals and systems on the web. To facilitate this we introduce the cost-effective idea of utilizing ready to use Web 2.0 mash-up services in the same spirit as Google Ads or Maps are used on external web pages and applications. In this way, museums, libraries, tourism portals, news papers, individual citizens, and other users can include CultureSampo materials, such as semantic search results and recommendation links, on their web pages using mash-ups. This clearly benefits everybody: the materials of the CultureSampo collaborative network get more visi-



bility and the external users can enrich their own materials for "free": only 1–2 lines of Javascript code is needed.

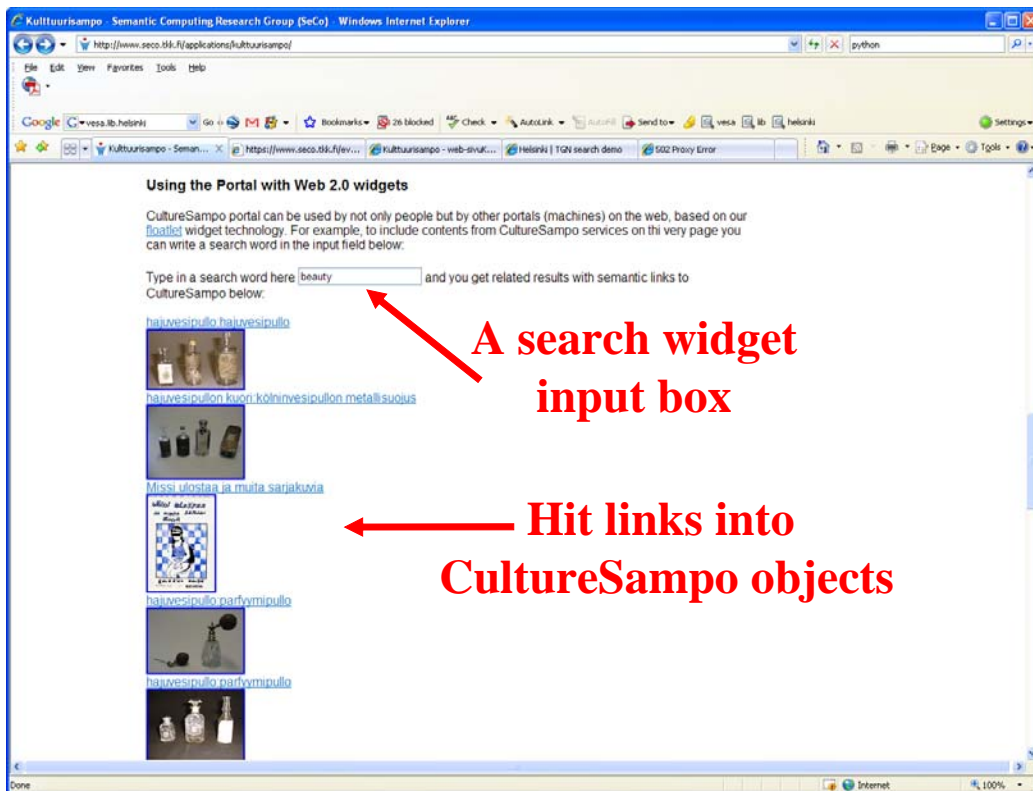


Figure 4.13 Semantic CultureSampo Widget.

An example mash-up is depicted in figure 4.13. Here the user types in a search word in the input field. After each character input, the CultureSampo service is queried and the matching results are displayed below the search field. In this case, the user is typing "beauty" in English, and bathroom equipments and some paintings related to the concept are returned. The actual contents are available only in Finnish. However, search can be performed successfully because parts of the underlying ontologies are available in English and Finnish.

## 5. Discussion

The vision and design of CultureSampo goes beyond current semantic web portals for cultural heritage (Hyvönen, 2009). The contributions of the work underlying the CultureSampo system have many facets and have been made during several years. The contribution of this paper is to give an overview of the whole system, its philosophy, and the new thematic perspective based end user interface, with pointers to papers and web sites explaining the approaches and software developed in more detail.

To summarize our work, CultureSampo contributes to research and development of semantic portals for cultural heritage especially in the following ways:

1. **Cross-domain content, ontologies, and metadata.** The system is highly cross-domain with dozens of content types and metadata schemas. Usually only one schema such as Dublin Core or VRA Core (<http://www.vraweb.org/projects/vracore4/>) is used in cultural portals (cf. e.g. (Schreiber et al., 2006; Wang et al., 2008)).
2. **Event-based narrative semantic models.** CultureSampo makes use of sophisticated semantic annotation models including events and processes.

3. **Semantic search and recommending.** The system uses new kinds of semantic search and recommendation techniques.
4. **Semantic visualizations.** The system has an exceptionally versatile selection of semantic visualizations available, such as different map views, timelines, graphs, process visualization, and semantic video viewing.
5. **Collaborative ontology development.** The system is based on a large nationwide collaboratively maintained infrastructure of ontologies and ontology services.
6. **Collaborative metadata creation.** The system includes a model of and tools for collaborative semantic content creation.
7. **Machine semantics and services.** The services can be made available for machines.
8. **Large collaboration network.** The system has been developed on a national level with contents from over 20 memory organizations, and includes content from many international sources.
9. **Multilinguality.** Although the contents are in Finnish, the system can be used in two other languages.

Two user evaluation studies concerning semantic recommendations of an earlier version of CultureSampo have been performed with some promising results [11]. However, end user evaluation of the interfaces of the new prototype with its various features has not yet been done.

The portal is scalable in terms of its Web 2.0 content creation model and different types of content. An explicit concern in implementing the portal has been the computational efficiency in terms of speed and memory consumption, since a national level cultural heritage portal, if successful, will have lots of content inside and a large number of simultaneous users. For this goal, our earlier memory- and Prolog-based search engine used in MuseumFinland was replaced by conventional search engine technology, Apache Lucene, that was configured to do semantic search using semantic indexing. The CultureSampo search engine has been tested with a knowledge base of 20 million objects resulting in response times of less than 2 seconds on ordinary PC hardware.

## Acknowledgements

Several researchers, including, Airi Hortling, Jouni Hyvönen, Ellen Karhulampi, Suvi Kettula, Helena Mäkimattila, and Jari Väätäinen, have contributed to developing CultureSampo contents and the related infrastructural components. This work is a part of the national FinnONTO research project 2003–2007, 2008–2010, funded mainly by Tekes and a consortium of 38 companies and public organizations. The work is also partly funded by the EU FP7 SmartMuseum project 2008–2010, and the Finnish Cultural Foundation (2008–2010).

## References

- Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening ontologies with DOLCE. In: Proc. of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, Springer–Verlag (2002)
- Hyvönen, E., Ruotsalo, T., Häggström, T., Salminen, M., Junnila, M., Virkkilä, M., Haaramo, M., Kauppinen, T., Mäkelä, E., Viljanen, K.: CultureSampo—Finnish culture on the semantic web. The vision and first results. In: Klaus Robering (Ed.), Information Technology for the Virtual Museum. LIT Verlag. (2008a)
- Hyvönen, E., Viljanen, K., Tuominen, J., Seppälä, K.: Building a national semantic web ontology and ontology service infrastructure—the FinnONTO approach. In: Proceedings of the ESWC 2008, Tenerife, Spain, Springer–Verlag (2008b)

- Hyvönen, E., Alm, O., Kuittinen, H.: Using an ontology of historical events in semantic portals for cultural heritage. In: Proceedings of the Cultural Heritage on the Semantic Web Workshop at the 6th International Semantic Web Conference (ISWC 2007). (November 12 2007)
- Hyvönen, E., Takala, J., Alm, O., Ruotsalo, T., Mäkelä, E.: Semantic Kalevala—accessing cultural contents through semantically annotated stories. In: Proceedings of the Cultural Heritage on the Semantic Web Workshop at the 6th International Semantic Web Conference (ISWC 2007), Busan, Korea. (2007)
- Hyvönen, E.: Semantic portals for cultural heritage. In: Handbook of Ontologies, 2. edition, Springer-Verlag (March 2009).
- Hyvönen, E., Mäkelä, E.: Semantic autocompletion. In: Proceedings of the First Asia Semantic Web Conference (ASWC 2006), Beijing, Springer-Verlag (2006)
- Hyvönen, E., Mäkelä, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., Junnila, M., Kettula, S.: Museum-Finland—Finnish museums on the semantic web. *Journal of Web Semantics* 3(2) (2005) 224–241
- Hyvönen, E., Junnila, J., Kettula, S., Mäkelä, E., Saarela, S., Salminen, M., Syreeni, A., Valo A., and Viljanen, K.: Finnish Museums on the Semantic Web. User's Perspective on MuseumFinland. Proceedings of Museums and the Web 2004 (MW2004), Archives & Museum Informatics (2004).
- Junnila, M., Hyvönen, E., Salminen, M.: Describing and linking cultural semantic content by using situations and actions. In: Klaus Robering (Ed.), *Information Technology for the Virtual Museum*. LIT Verlag. (2008)
- Kauppinen, T., Väättänen, J., Hyvönen, E.: Creating and using geospatial ontology time series in a semantic cultural heritage portal. In: Proceedings of the ESWC 2008, Tenerife, Spain, Springer-Verlag (2008)
- Kurki, J., Hyvönen E.: Relational semantic search: Searching social paths on the semantic web. In: Poster Proceedings of the ISWC + ASWC 2007, Busan, Korea. (2007)
- Mäkelä, E., Viljanen, K., Alm, O., Tuominen, J., Valkeapää, O., Kauppinen, T., Kurki, J., Sinkkilä, R., Käsälä, T., Lindroos, R., Suominen, O., Ruotsalo, T., Hyvönen, E.: Enabling the Semantic Web with Ready-to-Use Web Widgets. In Nixon, L., Cuel, R., Bergamini, C., eds.: *First Industrial Results of Semantic Technologies*, proceedings, co-located with ISWC 2007 + ASWC 2007, Busan, Korea. (2007) CEUR Workshop Proceedings, Vol 293.
- 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)
- Ruotsalo, T., Hyvönen, E.: An event-based approach for semantic metadata interoperability. In: Proceedings of the ISWC 2007 + ASWC 2007, Busan, Korea, Springer-Verlag (2007b)
- 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: *The Semantic Web – Proceedings of the 5th International Semantic Web Conference 2006*. (2006) 951–958
- Staab S., Studer R. (eds.) *Handbook on Ontologies*. Springer-Verlag, March 2009 (2. edition).
- Sheth, A., Aleman-Meza, B., Arpinar, I.B., Bertram, C., Warke, Y., Ramakrishnan, C., Halaschek, C., Anyanwu, K., Avant, D., Arpinar, F.S., Kochut, K.: Semantic association identification and knowledge discovery for national security applications. *Journal of Database Management on Database Technology* 16(1) (Jan–March 2005) 33–53
- Sowa, J.: *Knowledge Representation. Logical, Philosophical, and Computational Foundations*. Brooks/Cole (2000)
- Valkeapää, O., Alm, O., Hyvönen, E.: Efficient content creation on the semantic web using metadata schemas with domain ontology services. In: Proceedings of ESWC 2007, Innsbruck, Austria, Springer-Verlag (2007)
- van Assem, M., Malaise, V., Miles, A., Schreiber, G.: A method to convert thesauri to SKOS. In: Proceedings of the Third European Semantic Web Conference (ESWC'06), Springer-Verlag (2006)

- Viljanen, K., Tuominen, J., Hyvönen, E.: Distributed semantic content creation and publication for cultural heritage legacy systems. In: Proceedings of the 2008 IEEE International Conference on Distributed Human- Machine Systems, Athens, Greece, IEEE Press (2008)
- Viljanen, K., Tuominen, J., Käsälä, T., Hyvönen, E.: Distributed semantic content creation and publication for cultural heritage legacy systems. In: Proceedings of the 2008 IEEE International Conference on Distributed Human- Machine Systems, Athens, Greece, IEEE Press (2008)
- Wang, Y., Stash, N., Aroyo, L., Gorgels, P., Rutledge, L., Schreiber, G.: Recommendations based on semantically-enriched museum collection. *Journal of Web Semantics* (2008)