Ontology-based Disambiguation of Spatiotemporal Locations

Tomi Kauppinen[1], Riikka Henriksson[1], Reetta Sinkkilä[1], Robin Lindroos[1], Jari Väätäinen[2], and Eero Hyvönen[1]

[1] Semantic Computing Research Group (SeCo)
Helsinki University of Technology and University of Helsinki,
P.O. Box 5500, 02015 TKK, Finland
first.last@tkk.fi,
http://www.seco.tkk.fi
[2] Geological Survey of Finland,
firstname.lastname@gtk.fi,
http://www.gtk.fi

Abstract. Geographic place names are semantically often highly ambiguous. For example, there are 491 places in Finland sharing the same name "Isosaari" (great island) that are instances of several geographical classes, such as Island, Forest, Peninsula, Inhabited area, etc. Referencing unambiguously to a particular "Isosaari", either when annotating content or during information retrieval, can be quite problematic and requires usage of advanced search methods and maps for semantic disambiguation. Historical places introduce even more challenges, since historical metadata commonly make spatiotemporal references to historical regions and places using names whose meanings are non-existing or different in different times. This paper presents how these problems have been addressed in a large Finnish place ontology SUO and a historical geo-ontology SAPO. A location ontology server ONKI-Geo has been created for publishing the ontologies and utilizing them as mashup services. To demonstrate the usability of our ontologies, two case applications in the cultural heritage domain are presented.

1 Introduction

Geospatial ontologies define classes and individuals for representing e.g. geographic regions, their properties, and mutual relationships [17, 16, 3]. By sharing ontological resources in different collections and application domains interoperability in terms of geographical locations can be obtained, and intelligent end-user services such as semantic search, browsing and visualization be facilitated. For example, in the semantic portal MuseumFinland[3] [7] a location partonomy[4] was used for annotating museum artifacts with metadata about the place of manufacture and place of usage.

[3] http://www.museosuomi.fi

[4] This partonomy is a part-of hierarchy of individuals of the classes Continent, Country, County, City, Village, Farm etc.

A problem when creating a semantic cultural heritage portal is that places, both modern and historical ones are needed but place finding and place name disambiguation is hard as there are millions of places. Moreover, geography changes rapidly, which makes it hard 1) for the content annotator to make correct references to spatiotemporal regions and 2) for the end-user to understand the changes in historical geography and, as a result, to formulate the queries. For example, many artifacts in MuseumFinland originate from regions that no longer exist and/or are not part of Finland but of Russia with new names after the Second World War. Finding the right names for querying, understanding to what regions on the map the names refer to at different times, and understanding how old historical names relate to modern Finnish and Russian geography creates, at the same time, both a semantic challenge for the technology and an important part of useful content to learn when using the portal.

We propose that (at least) three requirements should be taken into account when utilizing location information in a system for end-users to view cultural heritage collections:

- **Precise annotation with locations** Finding relevant location or place concepts and their properties (like place type, WGS84 coordinates, hierarchical position) easily from geo-ontologies for annotating resources is a key need.
- **Ontology-based spatiotemporal search** It is necessary to be able to use both historical and modern locations and regions as search concepts e.g. in a view-based, or multi-facet search.
- **Context visualization of the locations** An end-user needs to see the locations in a right context, for example their neighbors, covered (i.e. overlapped) historical regions, and their position in a partonomical hierarchy or on a map.

To address these requirements, we present in this paper work done for 1) modeling places, their properties and changes they have overcome as two ontologies SUO and SAPO and 2) publishing the ontologies in an location ontology server ONKI-Geo [6]. We also 3) present two case applications implemented as a part of the semantic CultureSampo portal [8] that use place ontologies in a search for items related to places and show how ONKI-Geo is used in CultureSampo as a subservice using mashup techniques.

2 Geo-Ontology SUO

Spatial structures and their mutual relations are a most essential form of geographic knowledge. The principles by which the geographic domain—and geo-ontologies—are primarily structured from the theoretical viewpoint are topology (the theory of boundaries, contact, and separation), mereology (theory of parts and wholes) and geometry [14]. To facilitate spatial reasoning based on partonomy and subsumption hierarchies in semantic portal applications [**?**,8], we have created a Finnish geo-ontology SUO [5] [5] in RDF [2] and OWL [6].

The SUO has been populated with 1) place information from the Geographic Names Register (PNR) provided by the National Land Survey of Finland[7] and with 2) place information from the GEOnet Names Server (GNS)[8] maintained by the National Geospatial-Intelligence Agency (NGA) and the U.S. Board on Geographic Names (US BGN). PNR contains about 800,000 names of natural and man-made features in Finland, including data such as place type or feature type and the coordinates of a place. The GNS register contains similar information of about 4,100,000 places around the world excluding places in the United States. An analysis of PNR and GNS produced the majority of classes of the SUO-ontology: together it contains over 800 different classes of man-made or natural geographical places, organized into a class hierarchy, and their topological and mereological relations, as well as relations defining coordinate-based geometry for points, curves and polygons. SUO includes classes such as *city*, *province*, *graveyard*, *lake*, or *river*.

3 Historical Geo-Ontology SAPO

Names and borders for political geographic regions change over time: they are e.g. split (e.g. former Czechoslovakia) or merged together (e.g. former East and West Germany). Furthermore, also place names change (e.g. Zaire changed its name to Congo). As a result, content in databases may be indexed with names that do not exist anymore or refer to different regions than names in use today. This creates severe problems for information indexing and retrieval.

We have addressed the problem by developing a method and a historical place ontology SAPO for representing geographical changes [11, 12]. SAPO consists of a small set of classes and relationships (*merged*, *split*, *rename*, ...) for representing changes, and over 1100 change instances concerning around 930 different historical Finnish counties, the complete change history of Finnish counties in 1860–2007. SAPO uses the place types from SUO but extends it into the temporal dimension by telling the begin and end times and changes concerning each place.

Figure 1 shows a part of SAPO where two regions, namely *Nuijamaa (1944–1988)* and *Lappeenranta (1967–1988)* have merged into one region called *Lappeenranta (1989–)*. The merge is defined as two RDF triples $< A, mergedTo, C >$ and $< B, mergedTo, C >$, where A and B are regions that are merged to form C.

Based on the change types, logic rules can be applied to enrich the ontology with new useful relations. For example, in figure 1 the RDF statements *covers*

[5] http://www.seco.tkk.fi/ontologies/suo/
[6] http://www.w3.org/2001/sw/
[7] http://www.maanmittauslaitos.fi/en/
[8] http://earth-info.nga.mil/gns/html/

and *neighborOf* can be derived based on the *mergedTo* change. A rule for deriving *covers* statements is given below:

$$A\ mergedTo\ C\ \wedge\ B\ mergedTo\ C \longrightarrow$$
$$A\ covers\ C\ \wedge\ C\ covers\ A\ \wedge\ B\ covers\ C\ \wedge\ C\ covers\ B$$

Another set of rules adds the triples $< A, neighborOf, B >$ and $< B, neighborOf, A >$ by checking if there are exactly two regions to be merged and if the area of the resulting region is made of a simple polygon.

Mutual *covers* relations of regions can be used to compute coverages between regions of different times, without knowing the actual polygons of the regions, and a *geo-ontology time-series* of regions [11]. Our method produces the coverages[9] between all the places in a historical place ontology.

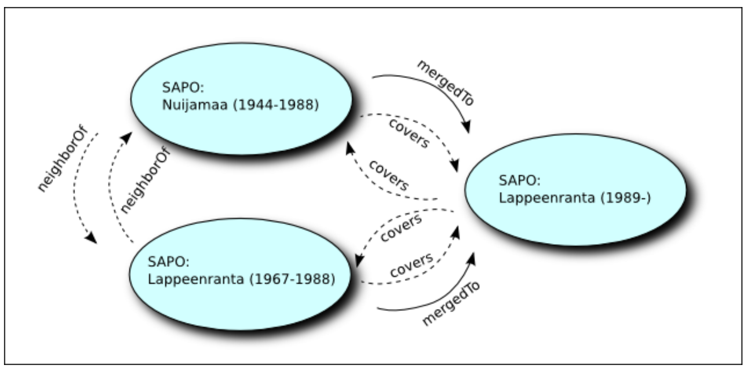

Fig. 1. The base model and infered model of the SAPO historical ontology. The predicates of the infered triples are marked with dashed lines.

The result can be used to show the temporal contextual information about places or in a query expansion. For example, to visualize the result of the inference and to see regions in relation with other regions, we have used the IRMA-Sapo browser[10], depicted in figure 2. With IRMA-Sapo browser, a user can find historical regions beginning with certain letters and also their neighbouring regions, covering regions, and see them in their partonomy.

IRMA-Sapo is based on the intelligent metadata assistant IRMA [13, 15]. IRMA is generic in a sense that it can be used to visualize the context of any concept, i.e. the properties and relationships of that concept.

[9] There are a few exceptions where manual work is needed as discussed in [11].
[10] http://www.seco.tkk.fi/services/irma/

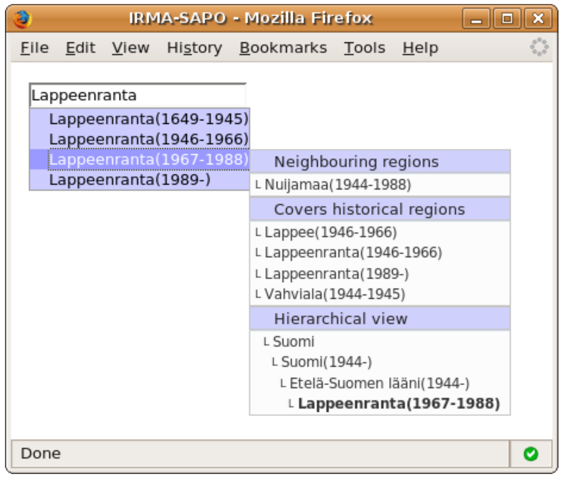

Fig. 2. SAPO historical ontology in an IRMA browser. Search is supported by an autocompletion box which retrieves all the matching historical regions. By clicking a region, the user can further select either from the neighbouring historical regions, covering regions, or from the partonomy hierarchy.

4 Ontology Service ONKI-Geo

ONKI-Geo [6] publishes place ontology SUO and historical place ontology SAPO with online ontology services for humans and machines to use. The user interface utilizes Ajax-techniques[11] for communicating with the ONKI-Geo database server containing the place instances. The place finder contains a simple faceted search engine for narrowing the search along the following dimensions: 1) Place name facet filters matching place instances using string autocompletion for their labels. 2) Place type facet of place types (*City*, *Island*, *Cemetery*, etc.) is used to focus search to places of desired types. 3) Language facet limits search to place names in given languages (Finnish, Swedish, English, and three dialects of Sami). 4) Time facet is used to focus search on historical place names. 5) Map facet allows the user to specify a polygon area in which to search for the place on the map [2]. The polygon functions as a narrowing criteria for the search in the same way as category selections in the other facets. 6) Area facet makes it possible to focus search on continents, countries, and their smaller regions.

The ONKI-Geo Browser is depicted in Figure 3. By using ONKI-Geo one can e.g. find easily on the map all places of 1) some type with 2) names beginning with some letters, and that are 3) inside a polygon, out of millions of places[12]. ONKI-Geo uses Google Maps API for visualization, and can be connected to

[11] http://ajax.org

[12] A related service, Geonames.org also includes a web service to return place name information and supports search e.g. by type and name but not using e.g. a polygon.

and utilized in legacy systems using Ajax and mash-up techniques. In a museum cataloging system, for example, places can be found by using the ONKI-Geo user interface and by pushing the "Select" button, the corresponding URI or coordinates are transferred from the centralized national service into the lecagy application. A technical challenge with the ONKI-Geo service is the large size of the SUO. The response times of the faceted search were several minutes when using a straight-forward Jena implementation [13]. To solve the efficiency problem, only the SUO ontology classes, without the instances, are stored in a Jena model, and a separate indexing layer was created for the millions of instances. The indexing of the instances is created in a relational database (MySQL) making complex combinational searches possible and relatively fast.

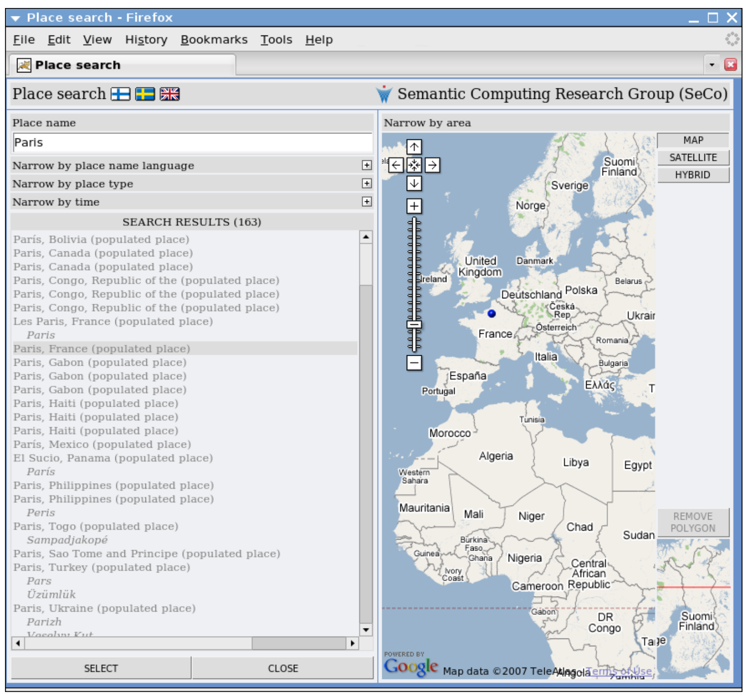

Fig. 3. ONKI-Geo Browser. Search can be constrained by using the facets on the left or by drawing a polygon on the map. By pushing the "Select" button in the left bottom corner, the found concept or selected coordinate information is transferred into the mash-up widget in a legacy application.

[13] http://jena.sourceforge.net

5 Two Case Applications

We have implemented two case applications as a part of the semantic CULTURESAMPO portal to demonstrate the use of place ontologies in the cultural heritage domain.

Map-search allows for searching with places on a map as illustrated in Figure 4. By clicking a place on a map, the items annotated with that place are retrieved and shown on the right side of the map. Map-search uses simple inference by creating a transitive closure over the part-of hierarchy.

Figure 5 illustrates the second case application. Historical regions of the SAPO ontology can be selected in a drop-down menu on the left. Here a temporal, historical place *SAPO:Viipuri(1403-1943)* is selected. As a result, the polygonal boundaries of *Viipuri* of that time are visualized on a contemporary Google Maps satellite image[14], a map, or on a historical map. In addition, those modern places of ONKI-Geo that are inside the polygonal boundaries of the historical region are retrieved in a mashup fashion and can be used to browse the map. The content related to *SAPO:Viipuri(1403-1943)* is retrieved from an RDF repository and is shown in this view on the right. Furthermore, content from historical regions that cover (i.e. overlap) *SAPO:Viipuri(1403-1943)* are listed as recommendations. The coverages are looked up from the global coverage table [11] containing all the coverages between historical places.

6 Discussion and Related Work

One of the major problems in semantic web ontologies is that organizations commonly mint different URIs for same resources [9]. One possible solution is to build coreference systems that offer mappings between different URIs [9]. On the other hand, if organizations would retrieve URIs directly from commonly used ontology library systems and ontology servers offering a rich set of relationships between the URIs, the integration of data sets would be more straighforward. The idea about these ontology servers and libraries have been presented before [4] and they have also been extensively compared [1]. However, there is little work done for offering ontology services for semantic geospatial ontologies where there are specific needs for disambiguation among large amount of spatiotemporal instances. The SPIRIT spatial search engine [10] provides facilities to find web resources relating to places referred to in a query. The geo-ontology within this search engine is used for the purposes of information retrieval by modeling the geographical terminology and the spatial structure of places. It supports for example query disambiguation by recognizing the variant place names and historical alternatives.

Our approach is to provide a rich set of spatial and temporal properties for the place instances in order to facilitate the process of disambiguating between them and for finding the needed instance (i.e. the URI) to be used e.g. in annotation.

[14] http://maps.google.com/

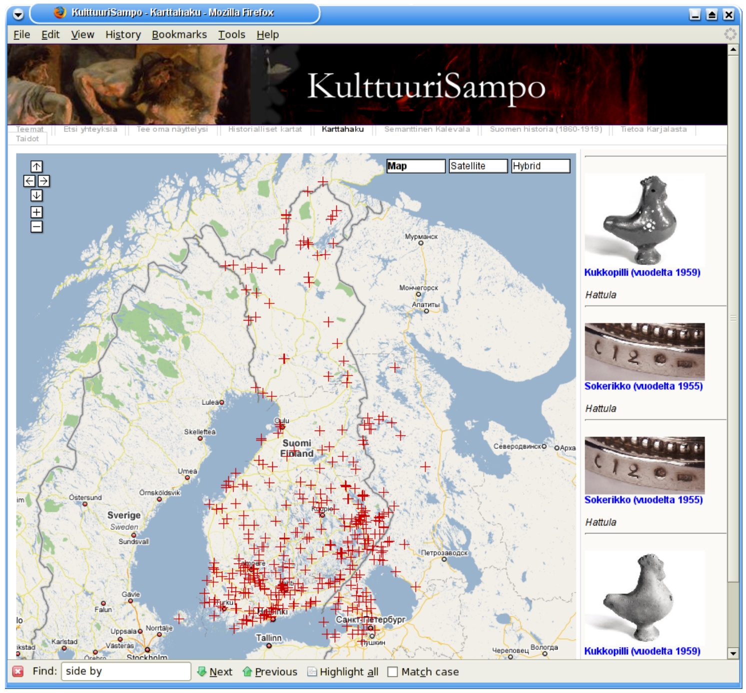

Fig. 4. A search with regions without temporal extensions.

We believe that if there is not enough evidence—i.e. properties and relationships between URIs—that a URI served by a service is the needed one both temporally and spatially then organizations tend to either mint their own URIs for the purpose or to select URIs somewhat randomly from different ad hoc -services.

This paper presented how two ontologies, SUO and SAPO, can be used to represent and visualize different kinds of locations and their changes in time. We also presented the ONKI-Geo ontology server for publishing SUO and SAPO, and for finding the correct places for annotations. Finally, we presented two case applications that demonstrate that the presented ideas work also in practise in the cultural heritage domain.

References

1. M. N. Ahmad and R. M. Colomb. Managing ontologies: a comparative study of ontology servers. In *ADC '07: Proceedings of the eighteenth conference on Australasian database*, pages 13–22, Darlinghurst, Australia, Australia, 2007. Australian Computer Society, Inc.
2. D. Brickley and R. V. Guha. RDF Vocabulary Description Language 1.0: RDF Schema W3C Recommendation 10 February 2004. Recommendation, World Wide Web Consortium, February 2004.

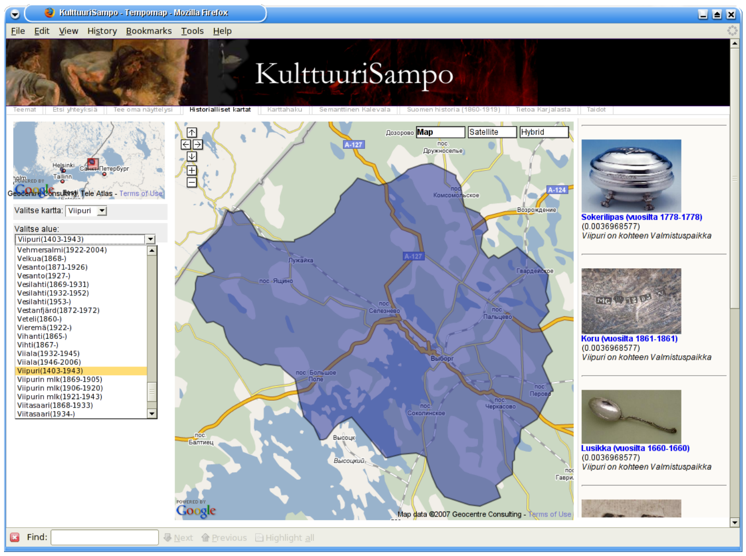

Fig. 5. Temporal parts are used to visualize polygonal boundaries of historical regions in CultureSampo and for searching historical items.

3. J. C.B., A. Abdelmoty, and G. Fu. Maintaining ontologies for geographical information retrieval on the web. In *in Meersman, R.; Tari, Z.; Schmidt, D. C. (Eds.) On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE Ontologies, Databases and Applications of Semantics, ODBASE03*, Catania, Italy, Lecture Notes in Computer Science 2888.
4. Y. Ding and D. Fensel. Ontology library systems: The key to successful ontology reuse. In *Proceedings of SWWS'01, The first Semantic Web Working Symposium, Stanford University, USA*, pages 93–112, 2001.
5. R. Henriksson, T. Kauppinen, and E. Hyvönen. *Core Geographical Concepts: Case Finnish Geo-Ontology.* Location and the Web (LocWeb) 2008 workshop, 17th International World Wide Web Conference WWW 2008, Beijing, China, April 21-25 2008.
6. E. Hyvönen, R. Lindroos, T. Kauppinen, and R. Henriksson. An ontology service for geographical content. In *Poster Proceedings of the International Semantic Web Conference (ISWC 2007), Busan, Korea*, Nov 2007.
7. E. Hyvönen, E. Mäkelä, 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, 2005.
8. E. Hyvönen, T. Ruotsalo, T. Häggstöm, M. Salminen, M. Junnila, M. Virkkilä, M. Haaramo, E. Mäkelä, T. Kauppinen, , and K. Viljanen. Culturesampo–Finnish culture on the semantic web: The vision and first results. In *In: K. Robering (ed.): Information Technology for the Virtual Museum. LIT Verlag, Berlin.*, 2008.
9. A. Jaffri, H. Glaser, and I. Millard. Uri identity management for semantic web data integration and linkage. In *3rd International Workshop On Scalable Semantic Web Knowledge Base Systems.* Springer, 2007.

10. C. B. Jones, A. I. Abdelmoty, D. Finch, G. Fu, and S. Vaid. The spirit spatial search engine: Architecture, ontologies and spatial indexing. In M. J. Egenhofer, C. Freksa, and H. J. Miller, editors, *GIScience*, volume 3234 of *Lecture Notes in Computer Science*, pages 125–139. Springer, 2004.
11. T. Kauppinen and E. Hyvönen. *Modeling and Reasoning about Changes in Ontology Time Series*, chapter 11 in Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems. Rajiv Kishore and Ram Ramesh and Raj Sharman (Eds.). Integrated Series in Information Systems, Volume 14. Springer-Verlag, 2007.
12. T. Kauppinen, J. Väätäinen, and E. Hyvönen. Creating and using geospatial ontology time series in a semantic cultural heritage portal. In *S. Bechhofer et al.(Eds.): Proceedings of the 5th European Semantic Web Conference 2008 ESWC 2008, LNCS 5021, Tenerife, Spain*, pages 110–123, June 1-5 2008.
13. E. Mäkelä, R. Sinkkilä, and E. Hyvönen. Combining cross-ontology navigation with semantic autocompletion. In *Poster+Demo proceedings, ISWC+ASWC-2007, Busan, Korea*, 2007.
14. D. M. Mark, A. Skupin, and B. Smith. Features, objects, and other things: Ontological distinctions in the geographic domain. *In Spatial Information Theory: Foundations of Geographic Information Science (Lecture Notes in Computer Science 2205)*, pages 488–502, 2001.
15. R. Sinkkilä, E. Mäkelä, T. Kauppinen, and E. Hyvönen. Combining context navigation with semantic autocompletion to solve problems in concept selection. In *SEMMA 2008 Workshop, 5th European Semantic Web Conference 2008 ESWC 2008, Tenerife, Spain*, June 1-5 2008.
16. H. Stuckenschmidt and F. V. Harmelen. *Information Sharing on the Semantic Web*. Springer-Verlag, Berlin Heidelberg, New York, 2004.
17. U. Visser. *Intelligent information integration for the Semantic Web*. Springer-Verlag, Berlin Heidelberg, New York, 2004.