ONKI-Paikka: An Ontology Service for Geographical Data

Robin Lindroos, Tomi Kauppinen, Riikka Henriksson, and Eero Hyvönen
Semantic Computing Research Group (SeCo)
Helsinki University of Technology (TKK) and
University of Helsinki
http://www.seco.tkk.fi/
first.last@tkk.fi

Abstract

Geographic place names are widely used in databases, but 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, Field, Peninsula, Inhabited area, etc. Referencing unambiguously to a particular "Isosaari" on the semantic web, either when annotating content or during information retrieval, can be quite problematic and requires usage of advanced search methods and maps for semantic disambiguation. This paper presents an ontology server, ONKI-Paikka, for solving the place finding and place name disambiguation problem. The system currently contains 800,000 place names in Finland and millions of places in other countries. In ONKI-Paikka, places can be found by a faceted search engine, combined with semantic autocompletion and a map service for constraining search and for visualizing results. The service can be connected to applications cost-effectively by using Ajax-technology in the same spirit as Google Maps, used in ONKI-Paikka as a subservice.

1 Providing Geographic Ontology Services

Storing geographic information in an ontological fashion brings many advantages. For instance topological and other spatial relations between places can be described by semantic web technologies like RDF (Resource Description Framework) (RDF) and OWL (Web Ontology Language)[1]. Moreover, places and other geographic features (e.g. man-made structures) can be disambiguated by giving them Universal Resource Identifiers, URIs.

Using literal place names in annotations and search queries is often semantically highly ambiguous because several places of different types may have the same name (e.g., there is a field called "New York" in Finland), places may have multiple names (e.g., different transliterations), different languages may use different names for the same entity (e.g., London vs. Londres). Furthermore, places may have undergone several name changes due to political or other reasons. This makes it very difficult to automatically disambiguate different places based only on their name.

One of the prerequisites of the Semantic Web is a web of unambiguous URI references represented as resources in shared ontologies. In addition, services for finding and referencing to these resources in content annotation and information retrieval are needed. Provision of such ontology services in a cost-effective way is one of the major goals of the ONKI Ontology server concept developed in the National Semantic Web Ontology Project FinnONTO (2003-2007) in Finland (Hyvönen, 2006; Komulainen et al., 2005). ONKI-Paikka is a tool designed for this purpose for the geographic domain: it enables a user to annotate and search resources with precise geographic references and coordinate data.

2 SUO Place Ontology

ONKI-Paikka is based on the Finnish Place Ontology SUO (Suomalainen Paikkaontologia) (Kauppinen et al., 2006) being developed in FinnONTO. The SUO ontology has currently been populated with 1) place information from the Geographic Names Register (GNR) provided by the National Land Survey of Finland[2] and with 2) place information from

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

[2] http://www.maanmittauslaitos.fi/

the GEOnet Names Server (GNS)[3] maintained by the National Geospatial-Intelligence Agency (NGA) and the U.S. Board on Geographic Names (US BGN). GNR 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, accordingly.

ONKI-Paikka can be populated with places from different sources. Every data source forms an isolated set of place information with a unique namespace for all the instances originating from that source.

A major challenge with the ONKI-Paikka service is the large size of the SUO place ontology. For the end-user, finding the intended place reference out of millions of resources becomes complicated. From the implementation view point, straight forward ways of handling the ontology as a list of triplets, with for instance Jena[4], is too slow since there are over ten millions triples to deal with).

3 Finding the Intended Place

The place finder of ONKI-Paikka is a browser-based application which uses AJAX-techniques[5] for communicating with the ONKI-Paikka database server containing place instances. The place finder contains a simple faceted search engine for narrowing the search along following different dimensions: 1) Place name facet filters from the database places using string autocompletion. 2) Place type facet of some 50 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, etc.). 4) Time facet is being implemented in order 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 (Kauppinen et al., 2006). The polygon functions as a narrowing criteria for the search in the same way as category selections in the other facets. For visualization of the places ONKI-Paikka uses Google Maps API [6].

Using ONKI-Paikka one can e.g. find all places of 1) some type with 2) names beginning with some letters and that are 3) inside some area defined as a polygon.

[3] http://earth-info.nga.mil/gns/html/

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

[5] http://ajax.org/

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

4 Dealing with Millions of Place Instances

According to our tests, the response times of the faceted search of ONKI-Paikka were several minutes when using a straight forward Jena implementation on an average computer. To solve the efficiency problem, ONKI-Paikka stores only the data necessary for searches in the database. A separate indexing layer was created for the ontology's millions of instances. Only the SUO ontology classes, without the instances, are stored in a Jena model. This makes ontological reasoning easy and efficient. The indexing of the instances is created in a relational database[7]. The indexing functionality of a database is used to speed up searches making complex combinational searches possible and relatively fast.

Acknowledgments

Our research is a part of the National Semantic Web Ontology Project in Finland[8] (FinnONTO) 2003-2007 funded mainly by the Finnish Funding Agency for Technology and Innovation (Tekes).

References

Eero Hyvönen. Finnonto–building the basis for a national semantic web infrastructure in finland. In *Developments in Artificial Intelligence and the Semantic Web - Proceedings of the 12th Finnish AI Conference STeP 2006*, October 26-27 2006.

Tomi Kauppinen, Riikka Henriksson, Jari Väätäinen, Christine Deichstetter, and Eero Hyvönen. Ontology-based modeling and visualization of cultural spatio-temporal knowledge. In *Developments in Artificial Intelligence and the Semantic Web - Proceedings of the 12th Finnish AI Conference STeP 2006*, October 26-27 2006.

Ville Komulainen, Arttu Valo, and Eero Hyvönen. A tool for collaborative ontology development for the semantic web. In *Proceedings of International Conference on Dublin Core and Metadata Applications (DC 2005)*, Nov 2005.

[7] We currently use MySQL, http://www.mysql.com/.

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

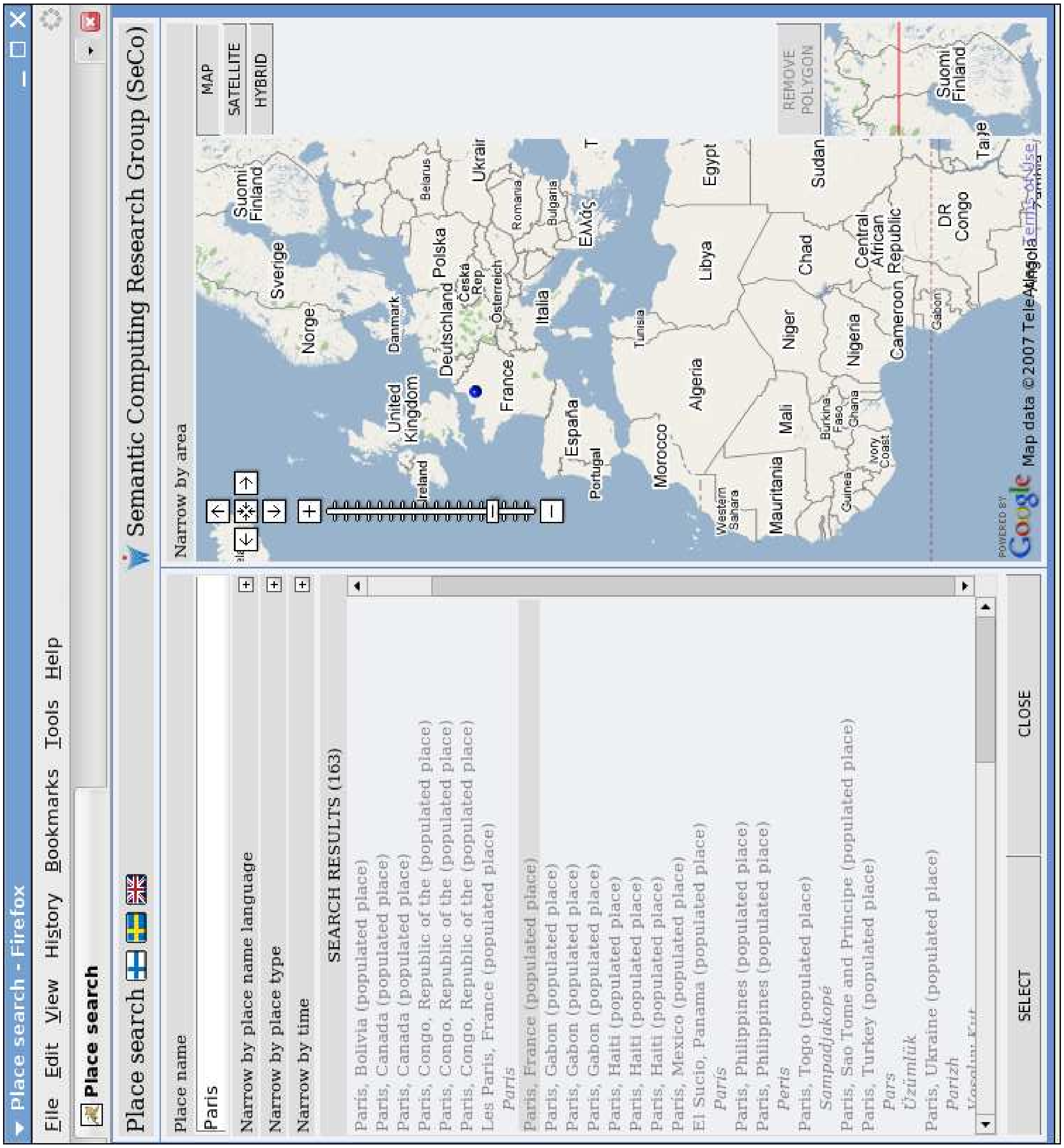

Figure 1: A view of the ONKI-Paikka. The facets for querying places are on the left, and the map for visualizing search results and for constraining search further by a polygon is on the left.