

Ontology Services Based on Crowdsourcing: Case National Gazetteer of Historical Places

Eero Hyvönen, Jouni Tuominen, Esko Ikkala, Eetu Mäkelä

Semantic Computing Research Group (SeCo), Aalto University
<http://www.seco.tkk.fi/>, firstname.lastname@aalto.fi

Abstract. This paper introduces the idea of applying crowdsourcing to evolving ontology services; the goal is to facilitate collaborative maintenance of ontologies in real time as a side effect of annotating contents in legacy cataloging systems. The idea is being implemented in the use case of creating and managing a national level gazetteer of historical places in Finland.

1 Problem: Cataloging with Evolving Shared Ontologies

In our previous work [7], we have shown that using shared ontologies and ontology services [1] is an effective way for creating semantically interoperable annotations in a distributed cataloging environment: The FinnONTO¹ ontologies are now widely used in Finland and the ONKI ontology service has been deployed by the National Library as a national free centralized service, connected to the cataloging systems of tens of organizations via APIs [5].

When annotating data using an ontology service, it has been a challenge to decide what to do when a new concept is needed in a shared ontology. The cataloger needs to make a reference to a concept not present in the shared ontology, say create a new place instance. The traditional approach to maintaining a Knowledge Organization System (KOS) is to contact the committee in charge of maintaining the KOS with a suggestion of a new concept to be added to the system. However, the cataloger cannot wait for the committee's decision for days, weeks, or months. Therefore, a shared mechanism is needed for populating the ontology with the following features: Firstly, it should be possible for the cataloger to create a new concept in real time or she is not able to make the annotation at hand, or has to make a less accurate annotation using only the existing concepts. Secondly, the new concept should be shared in real time with other users. Otherwise they may end up in creating duplicates of the same concept. Thirdly, there should be a mechanism for the maintaining committee of the KOS to edit, approve, or reject the proposed concepts afterwards, in case errors or duplicates arise.

In the following, a system addressing these needs, called Dynamic Ontology Services for evolving ontologies, is proposed. The idea is being implemented in a use case of an ontology service for historical places, with a first demonstration system presented.

¹ <http://www.seco.tkk.fi/projects/finnonto/>

2 Management Process for an Evolving Ontology

We propose an ontology management process involving three groups of people: 1) Ontology Committee (OC) responsible for maintaining and validating the ontology, 2) Developer Users (DU) using the system with the right to add new concept suggestions in the system, 3) Ordinary Users (OU) with the right of using the system as it is. The ontology is divided into three parts: 1) Validated Concepts (VC) constitute the official knowledge graph that has been validated and approved by the OC. 2) Suggested Concepts (SC) constitute a graph that includes concepts proposed by the DUs, but that have not been validated by the OC yet. 3) Corrected Concepts (CC) is a graph of mappings between unaccepted suggestions and accepted concepts in the VC graph.

The ontology evolves by crowdsourcing the DUs as follows:

1. A DU searches for an annotation concept *C* (using, e.g., autocompletion search). The system suggests matching concepts separately from the VC and SC. In this way the DU knows whether the concept is already accepted or was only suggested by someone. Both VC and SC concepts can be used for annotation.
2. If *C* is found and acceptable the DU can use it and the annotation is done.
3. Otherwise, she can create a new concept *C* with mandatory metadata, including a persistent identifier (IRI), labels for human identification, and additional properties, such as (sub)classes, coordinates, etc., depending on the data model of the KOS.
4. *C* is added into the SC, and is immediately available to all users. In particular, the DU is able to use *C* in her annotation at hand immediately.
5. The OC validates new concepts in the SC every now and then.
6. If the new concept *C* is valid, the OC moves it from the SC into the VC. At this point, it is possible to add and edit metadata as long as the IRI (i.e., the meaning of the concept), already used by the community for indexing, is not changed.
7. If the concept is not accepted, i.e., it is removed from the SC, or if the meaning of *C* is changed leading to minting a new IRI, then the OC creates a new correction entry in the CC. A correction entry is a triple (*suggestedIRI*, *mapping*, *existingIRI*) where *suggestedIRI* is the suggested concept, *existingIRI* is a concept in the VC, and *mapping* is a predicate indicating the relation between the suggested and existing concepts. For example, if *C* is a duplicate that already existed in the service, then it is not accepted and *owl:sameAs* can be used as the mapping predicate indicating the existing concept.

The idea of the CC is to give a fallback service to users that have used suggested concepts that were not finally accepted into the VC as they are. Using the correction mapping the data already falsely annotated by a user in her database can be mapped to the correct concepts of the VC later on.

3 Application: Finnish Ontology Service of Historical Places

Benefits The process proposed above is being applied in creating HIPLA, a sustainable, evolving repository and ontology service of historical places [4] in Finland. The idea is that Cultural Heritage organizations connect their legacy cataloging systems to HIPLA

using an API in the same vain as in ONKI [6]. When new places are encountered—a situation quite common with historical places—catalogers are able to suggest and use new concepts in HIPLA and share them with the community in real time. In this way, the ontology can evolve as a side effect of ordinary cataloging work with concepts that are actually used by the community. By crowdsourcing no major investments are needed in developing the ontology.

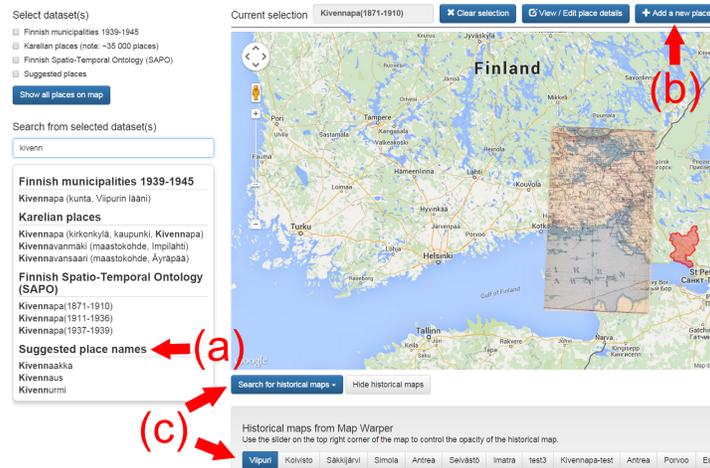


Fig. 1. Searching historical places with HIPLA.

End-user Interface For example, in Fig. 1 the user has typed “kivenn...” and selected one of the historical incarnations of the municipality “Kivennapa” provided by the SAPO ontology [3] from the autocomplete suggestions. In addition, the system shows user suggested places (a), which can be selected. If the desired place does not exist, the user can create a new place suggestion by clicking the button (b).

Historical Maps HIPLA is integrated with another service of historical maps. For map production, i.e., for aligning old maps on top of modern ones, an instance of the Map Warper tool is used². The metadata of the aligned maps is accessed via the Map Warper API that also provides the tiles of the maps, which makes it possible to show the maps on top of Google Maps used in HIPLA. It is possible to search old maps that intersect either the field of view or the selected place, which is essential for a service of historical places whose names often cannot be seen on contemporary maps, only on historical ones. In Fig. 1 the user has selected one old map (c) aligned on Google Maps.

Implementation Demo The prototype depicted in Fig. 1 is under development but can be tested at <http://www.ldf.fi/dev/hipla>. The system is implemented using the Linked Data Finland platform [2], based on Fuseki with a Varnish front end for serving linked data. HIPLA service contains separate graphs for the VC, SC, and CC, with user interfaces for querying, updating, and moving data based on SPARQL.

² <http://mapwarper.onki.fi>

4 Related Work and Discussion

HIPLA is an ontology library service [1] for historical geodata on maps. In contrast to traditional gazetteers, HIPLA not only publishes the data for humans but also for machines (legacy cataloging systems) using the SPARQL endpoint API.

Thesauri of historical places, published as Linked Data, include the Getty TGN³ of some 1.5 million records, 'Pelagios: Enable Linked Ancient Geodata In Open Systems'⁴, and Pleiades⁵. Pelagios and Pleiades are based on crowdsourcing volunteers' work in ontology development. The novelty of HIPLA lays in the idea of crowdsourcing the creation of the ontology to catalogers of Cultural Heritage content, as a side effect of their daily work, using the process presented in this paper. The Historical Gazetteer of England's Place-names⁶ is a service of over 4 million names than can be searched and viewed on modern maps as well as on historical ones. HIPLA has a similar local flavor focusing on places in Finland, but is based on Linked Data. In contrast to the systems above, HIPLA includes a map service for aligning and using historical maps, as in the New York Public Library's Chronology of Place gazetteer⁷.

Our work is supported by the Cultural Foundation of Finland.

References

1. d'Aquin, M., Noy, N.F.: Where to publish and find ontologies? A survey of ontology libraries. *Web Semantics: Science, Services and Agents on the World Wide Web* 11, 96–111 (2012)
2. Hyvönen, E., Tuominen, J., Alonen, M., Mäkelä, E.: Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets. In: *The Semantic Web: ESWC 2014 Satellite Events, Revised Selected Papers*. pp. 226–230. Springer-Verlag (2014)
3. Hyvönen, E., Tuominen, J., Kauppinen, T., Väättäin, J.: Representing and utilizing changing historical places as an ontology time series. In: Ashish, N., Sheth, A. (eds.) *Geospatial Semantics and Semantic Web: Foundations, Algorithms, and Applications*. Springer-Verlag (2011)
4. Southall, H., Mostern, R., Berman, M.L.: On historical gazetteers. *International Journal of Humanities and Arts Computing* 5(2), 127–145 (2011)
5. Suominen, O., Pessala, S., Tuominen, J., Lappalainen, M., Nykyri, S., Ylikotila, H., Frosterus, M., Hyvönen, E.: Deploying national ontology services: From ONKI to Finto. In: *Proceedings of the ISWC 2014, Industry Track. CEUR Workshop Proceedings, Vol-1383* (2014)
6. Tuominen, J., Frosterus, M., Viljanen, K., Hyvönen, E.: ONKI SKOS server for publishing and utilizing SKOS vocabularies and ontologies as services. In: *Proceedings of the 6th European Semantic Web Conference (ESWC 2009)*. pp. 768–780. Springer-Verlag (2009)
7. Viljanen, K., Tuominen, J., Känsälä, 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*. pp. 335–340. IEEE Press (2008)

³ <http://www.getty.edu/research/tools/vocabularies/tgn/>

⁴ <http://pelagios-project.blogspot.fi/p/about-pelagios.html>

⁵ <http://pleiades.stoa.org>

⁶ <http://www.placenames.org.uk>

⁷ <http://nypl.gazetteer.us>