Dynamic Ontology Service for Historical Persons and Places Based on Crowdsourcing

22.1.2016, COST RRL WG2 Workshop

Jouni Tuominen
Department of Computer Science, Aalto University
jouni.tuominen@aalto.fi
Problem: various reference sources

- Established (inter)national registries/ontologies
  - People: VIAF, Getty ULAN, CERL, ...
  - Places: Getty TGN, GeoNames, VIAF, ...
- Internal databases of organizations/systems, e.g., in EMLO
  - Coordination missing, no re-use of others’ work
    - Redundant work being done in different organizations
- Interoperability problems (syntax, semantics)
  - Contents do not get linked automatically
- Differing search user interfaces, APIs, editing tools, etc.

→ No unified access (or “global view”) to all the reference sources and their mutual relations
Requirements for dynamic ontology service

- Use of multiple reference sources simultaneously
- Users may add new people, places, etc.
  - Added instances are made available to other users instantly
- Collaboration of the content producer network
  - Maintain shared instance ontologies instead of internal ones
  - Build ontologies by crowdsourcing the indexers, as part of their daily work
HIPLA prototype

- The idea is prototyped in the Finnish Ontology Service of Historical Places and Maps (HIPLA): http://hipla.fi
HIPLA background: place name issues

- "historians often need specialised gazetteers listing places that no longer exist and names that are no longer used or whose spelling has significantly altered" (Southall et al., 2011)
- Name – place ambiguities (synonymy, homonymy)
  - One place – many different names
  - One place – many names in different languages
  - One name – many places
- Reference ambiguities in time
  - Places change in time » E.g. regional and other changes of Helsinki
  - Names change in time » E.g. place names of the Karelian region
HIPLA background: organizational issues

- Historical places are used in many organizations
  - Museums, libraries, archives, media companies, universities, ...
- Different organizations have their own repositories (if any!)
  - Redundant work is being done in different places
- Registries cannot be re-used easily in applications
  - Ontology services missing
  - Interoperability problems (syntax, semantics)
- Registries are not aligned with international registries
  - Interoperability problems globally
HIPLA solution

- Ontology model for historical places, based on W3C Semantic Web and GIS standards
- Ontology service based on Linked Data: HIPLA
  - Can be used alongside ordinary cataloging work
    - Previous work: ONKI Selector widget [http://onki.fi/widget/selector/](http://onki.fi/widget/selector/)
  - Search user interface with a map view for finding places
  - Multiple data sources, published in distributed SPARQL endpoints
    - Easy to add new data sources
- Crowdsourced process for place data harvesting
  - Catalogers are able to suggest and use new resources in HIPLA and share them with the community in real time
Legacy cataloging systems

**Public/private geographic datasets**
- Suggested places
- Geographic Names Registry
- DBpedia
- Getty TGN

**HIPLA**
no data storage, *but a common access to historical geodata*

Search and select a place from:
- Validated places
- Suggested places

Places from private repositories

Or if place is not found, create a new suggestion - **crowdsourcing**.

**SPARQL Query / Update**

**Map geo-rectifying service**

**Need to make a reference to a historical place**

Place URI
## HIPLA data sources

<table>
<thead>
<tr>
<th>name</th>
<th>source</th>
<th>type</th>
<th>size</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Getty Thesaurus of Geographical Names (TGN)</td>
<td>J. Paul Getty Trust</td>
<td>1800 place types</td>
<td>2 156 896</td>
</tr>
<tr>
<td>Finnish Geographic Names Registry</td>
<td>National Land Survey of Finland</td>
<td>61 place types (point)</td>
<td>797 668</td>
</tr>
<tr>
<td>Karelian places</td>
<td>National Land Survey of Finland</td>
<td>village, house, body of water, etc. (point)</td>
<td>33 938</td>
</tr>
<tr>
<td>Finnish Spatio-Temporal Ontology (SAPO)</td>
<td>SeCo</td>
<td>municipality (polygon, with temporal information)</td>
<td>1 261</td>
</tr>
<tr>
<td>Finnish Municipalities 1939-1945</td>
<td>National Archives of Finland</td>
<td>municipality (polygon)</td>
<td>612</td>
</tr>
<tr>
<td>Senate atlas</td>
<td>National Archives of Finland</td>
<td>georectified map</td>
<td>404</td>
</tr>
<tr>
<td>Karelian maps</td>
<td>National Land Survey of Finland</td>
<td>georectified map</td>
<td>47</td>
</tr>
</tbody>
</table>
Use Case I: Indexing

- During cataloging work the user needs to make a reference (find a URI) to a historical place:
  a) The user knows the place name (or part of it)
     - *Text search with autocompletion*
  b) The user has some idea where the place is located
     - *Browse places on a map*
  c) The place does not exist in the used ontologies
     - *Add a new place suggestion, and use the suggestion immediately*
Use Case II: Place ambiguities
Use Case III: Utilizing historical maps

- Because historical place names can often be seen only in historical maps, HIPLA is integrated with an open source map aligning tool: Map Warper.
- Map Warper makes it possible to view historical maps on top of modern maps, which is especially useful while adding new place suggestions.
HIPLA: future work

- Integration into legacy cataloging systems
- Implementing the crowdsourcing process
  - Easy, efficient tool for suggesting new places
  - Modeling the provenance of place suggestions
  - Validation system for incomplete place metadata
- More search functionalities for the end-user interface
  - Filter search results by place type, search historical maps by year
  - Taking the temporal dimension of places into account
- Model for managing multiple place data sources (cf. VIAF)?
Person search widget prototype
http://www.ldf.fi/dev/people-search-widget/

Finnish historical persons
Aalto, Alvar (1898-1976)
Aalto, Arvo (1932-)
Aalto, Erkki (1904-1984)
Aalto, Ilmari (1891-1934)
Aalto, Juho Lauri (1894-1965)
Aalto, Pentti (1917-1998)
Aaltola, Heikki (1905-1992)
Aaltonen, Aimo (1906-1987)
Aaltonen, Aleksi (1892-1956)
Dynamic ontology service widget

- Integration on the user interface level into cataloging systems, e.g., EMLO (a possible idea in the CofK project)
  - Currently EMLO Webform has an autocompletion search to the people and places in the internal database
  - In EMLO context, various complementary external reference sources are used separately (in web browser tabs), and the data is copied manually from them into the EMLO internal database
    - But some of them are just web pages, not published as data
- API integration would be another possible approach
  - Common (meta)search API to multiple reference sources
EMLO-Collect Webform: places

<table>
<thead>
<tr>
<th>Origin as marked</th>
<th>Is:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin as marked</td>
<td><img src="false" alt="Uncertain" /> <img src="false" alt="Inferred" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Notes on origin</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Destination</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Destination</td>
</tr>
</tbody>
</table>
EMLO: reference sources

**People**

Oxford Dictionary of National Biography [ODNB]
http://www.oxforddnb.com/

Consortium of European Research Libraries [CERL]
http://thesaurus.cerl.org/cgi-bin/search.pl?start=true

Virtual International Authority File [VIAF]
http://viaf.org/

Deutsche Biographie [NDB/ADB]
http://www.deutsche-biographie.de/search?st=ndb

Digitale bibliotheek voor de Nederlandse letteren [DBNL]
http://dbnl.org/index.php

Nieuw Nederlandsch Biografisch Woordenboek [NNBW]
http://www.historici.nl/retroboeken/nnbw/#page=0&size=800&accessor=accessor_index&source=1

History of Parliament Online
http://www.historyofparliamentonline.org/research/members

Clergy of the Church of England database [CCEd]
http://www.theclergydatabase.org.uk/jsp/locations/index.jsp?locKey=9684

British History Online
More info on HIPLA &
dynamic ontology services

- ISWC 2015 demo paper: main focus on the crowdsourcing process

- Short paper (submitted): vision and functionalities

- Long paper (submitted): combining the previous two + Linked Data services and an application use case (WarSampo portal)