A SEMANTIC GEO-CATALOGUE IMPLEMENTATION FOR A REGIONAL SDI

P. Shvaiko, Informatica Trentina
A. Ivanyukovich, Trient Consulting Group
L. Vaccari, EC DG JRC
V. Maltese, University of Trento
F. Farazi, University of Trento
Outline

- Introduction
- Geo-catalogue implementation
- Semantic extension for the geo-catalogue
- Conclusions
SDI in Trentino
Services for GI access

- Services for citizens
- Services for companies
- Services for research centers
- Services for public administrations
Geo-portal > geo-catalogue home page
Architecture - I

The Geo-catalogue consists of the following main modules:

- Metadata management
  - Add/Import/Modify/Delete/Permissions/State management (workflow)
- System configuration
  - Harvesting/Localization/Thesaurus management
- User & Group management
  - Import users/Role mgmt/Group mgmt/Ownership transfer
- Semantic matching engine
  - Intelligent query expansion
Architecture - II

BEA ALUI

Metadata management portlet
System configuration portlet
User and group management portlet

Web-services

Metadata management web-service
System configuration web-service
User and group management web-service

GeoNetwork

Query analysis
Faceted Ontology
Matching
Background ontology

Semantic matching

Geo-database

Authentication

SSO framework

Oracle

LDAP

Query

Results

INSPRIRE conference - June 24th 2010

Feroz Farazi – University of Trento
Implementation notes

- Workflow model for metadata management:
  - Notion of states – Draft, Review and Published
  - ...

- Metadata versioning

- Metadata validation using configurable rules:
  - Min field length (in words)
  - Stop words
  - ...

- Metadata quality control
  - Semaphore for fast quality check

- RNDT v1.0 - compliant implementation for Vector, Raster and Service metadata templates
Semantic extension: overview

- Problems with current tools for search (e.g. keyword search)
  - Different terminology used ⇒ low recall
  - Poor expressivity ⇒ low precision

- Intelligent query expansion to improve user experience
  - based on Semantic Matching techniques: the S-match system
Semantic matching and S-Match

- Semantic matching is a technique for identifying semantic correspondences between ontologies
- A set of matches is called alignment

S-Match is a tool for semantic matching developed at the University of Trento

Integration of S-Match into the geo-catalogue

- S-Match as web-service - advantages:
  - Separate VM usage for:
    - Performance optimization
    - Flexible memory management
  - Simplified updates in production environment:
    - Ontology updates
The ontology of Trentino: methodology

- Objective: Create an ontology that reflects the specificity of Trentino and respects the canons of the analytico-synthetic approach for the generation of a faceted lightweight ontology
  - Analysis and disambiguation of the classes
  - Hierarchical reorganization: semantic relations

- Available data: The ontological analysis was accomplished on the top 45 categories of place names (numbered from P110 to P650) provided by PAT

- The generated ontology includes 5 facets:
  - Antiquity
  - Body of water (rivers, lakes, …)
  - Geological formation (mountains, valleys, …)
  - Facility (roads, railways, shelters, …)
  - Administrative Division (provinces, municipalities, wards, …)
The ontology of Trentino: Body of water

Categories from P210 to P251

- P210 Corsi dacqua/laghi (1 ord.)
- P220 Corsi dacqua/laghi (2 ord.)
- P230 Corsi d’acqua/Canali/Fosse/Cond. forz./Laghi (3 ord.)
- P240 Corsi d’acqua/Canali/Fosse/Cond. forz./Laghi (>3 ord.-25.000)
- P241 Corsi d’acqua/Canali/Fosse/Cond. forz./Laghi (>3 ord.)
- P250 Sorg-Casc-Font-Presa importanti
- P251 Sorg-Casc-Font-Presa minori

Body of water (Idrografia)

- Lake (Lago)
- Group of lakes (Gruppo di laghi)
- Stream (Corso d’acqua)
  - River (Fiume)
  - Rivulet (Torrente)
- Spring (Sorgente)
- Waterfall (Cascata)
  - Cascade (Cascatina)
- Canal (Canale)
Usage examples of S-Match

<table>
<thead>
<tr>
<th>Query</th>
<th>Query expansion with S-Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>watercourse</td>
<td>Rivulet, Stream, River</td>
</tr>
<tr>
<td>falls</td>
<td>Cascade, Waterfall</td>
</tr>
<tr>
<td>water</td>
<td>Rivulet, Waterfall, Cascade, River, Body of water, Stream, Spring, Canal, Group of lakes, Lake</td>
</tr>
<tr>
<td>elevation</td>
<td>Natural elevation, Mountain, Highland, Glacier, Mountain range, Peak, Hill</td>
</tr>
<tr>
<td>installation</td>
<td>Milestone, Hut, Farm, Highway, Railway, Road, Street, Transportation system, Provincial Road, Facility, Shelter</td>
</tr>
<tr>
<td>ice mass</td>
<td>Glacier</td>
</tr>
<tr>
<td>district</td>
<td>Administrative division, Province, Municipality, Ward, Populated place</td>
</tr>
<tr>
<td>transportation facility</td>
<td>Transportation system, Road, Street, Provincial Road, Milestone, Railway, Highway</td>
</tr>
<tr>
<td>cliff</td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

- We found it useful to realize our geo-catalogue service from the available open-source implementation (GeoNetwork) conforming to the INSPIRE Directive and extend it with semantic capabilities.
- It was adapted based on the available technological infrastructure (the geo-portal of Trentino) and specificities of the implementation rules further constrained at the national level, e.g., by RNDT.

- Future work includes an extensive empirical evaluation of the semantic extension of the geo-catalogue.
- The expected impact here includes:
  - Better automation $\Rightarrow$ time and cost savings
    - I can find more rapidly the data that I am interested in
  - Better accuracy $\Rightarrow$ more precise decisions
    - Exhaustive results for search of data and services based on the available resource
THANK YOU
FOR YOUR ATTENTION AND INTEREST