Building ontology on-the-fly for the VGI and authority road networks to facilitate their integration

E. Abdolmajidi, L. Harrie, A. Mansourian

Department of Physical Geography and Ecosystem Science, Lund University
Corresponding author: ehsan.abdolmajidi@nateko.lu.se
Outline of the presentation:

• Problem statement
• Ontology
• Our proposal
• Related works
• Cases
• (Suggestion for) implementation
• Discussion
• Summary
• Future work
Problem statement

- Instance matching
- Segment-based algorithm
- Node-based algorithm
- Pattern-dedicated matching alg.
- Pattern detection
- Better solution

• Accuracy
• Efficiency
• Difficulties

RICH-VGI preconference Workshop- AGILE 2015, Lisboan, Portugal, 9th June. 2015
**Problem statement**

*Patterns*

**Pattern:**
A set of related instances that makes a particular shape which defines a concept (e.g. roundabout) and has a specific functionality.

**Difficulties of pattern detection:**

- Patterns are not the same through a dataset especially in VGIs, but their concept and functionality are the same.
- Relations between different patterns make the issue more complex.

Compactness index ≈ 1
Explicitly defining the conceptualization used in a domain application.

**Capabilities of ontologies:**

- Introducing the concept-instance relations (patterns)
- Introducing the instance-instance relations (relation between patterns)
- The concept is independent of the representation (concept defines the patterns)
Our proposal

Instance matching

Node-based algorithm

Segment-based algorithm

Pattern-dedicated matching alg.

Pattern detection

Better solution

Instance-level Ontology

Ontology

• Accuracy
• Efficiency
• Difficulties
Related work

- Ontology in spatial data field:
  - Extracting ontology from relational databases: An et al. (2006), Baglioni et al. (2007)
  - Schema-level ontology:
    • LinkedGeodata (Auer et al., 2009), Ballatore and Bertolotto (2011),
    • Towntology (COST Action C21, 2005/2009)
    • Ontology-based GIS (Belhadeif and Kholladi, 2009; Fonseca, et al. 2002)
    • OSMonto (Codescu, et al. 2011)
  - Including spatial dimension to ontology:
    • NeoGeo (GeoVocab.org)
    • GeoSPARQL (OGC)
    • PelletSpatial (Stocker and Sirin, 2009; Begetis, 2012)
    • Choros (Georg Christodoulou, 2011)
  - Road network ontology:
    • OTN (Ontology of Transportation Network) (REWERSE project)
    • Hierarchical ontology on multi-scale road model for cartographical applications (Mang and Weng, 2009)
Cases

Instance-level ontology for a roundabout and street:

RA1 is a Roundabout (instance-concept)
{A1, B1, C1, D1, E1, F1} build RA1 (instance-instance)
A1 is a Segment
...

ST1 is a Street
{LA1, LA2} Build ST1
LA1 is a Lane
{R1, Q1} build LA1
R1 is a Segment
...
Q1 intersects {F1, E1}

Reasoning:

ST1 is connected to RA1 (instance-instance relation -> patterns relation)
Cases

Instance-level ontology for integrating two datasets:

Using SWRL:
- \{LA1, LA2\} build \textit{detailed\_part1}
- \textit{detailed\_part1} is a \textit{detailed\_partition}
- \textit{detailed\_partition} is subclass of \textit{street\_partition}
- \{detailed\_part1\} build ST1

Using SWRL:
- \{LA3, LA4\} build \textit{detailed\_part2}
- \textit{detailed\_part2} is a \textit{generalized\_partition}
- \textit{generalized\_partition} is subclass of \textit{street\_partition}
- \{detailed\_part2, generalized\_part2\} build ST2
Implementation

Instance-level ontology

- The instance-level ontology needs to be handled by the data-producers/users.

- Creating instance-level ontology while the data is being produced (On-the-fly) seems to be the best way.

- Methods for gathering information:
  - Questionnaire (Naïve)
  - Data mining and defining rules (SWRL), still need to ask some questions (Aided)
Discussion

• **Benefits:**
  • The user has the best knowledge of what he is producing.
  • The accuracy of information is higher.
  • Independent from representation
  • The Instance-level ontology can be used:
    – Query answering
    – Generalization
    – Multi-scale databases

• **Difficulties:**
  • What is the best ontology design.
  • The questionnaire imposes the main work load on the data producers/users.
  • How to make an optimize questionnaire for gathering minimum information while enables us to produce required knowledge.
  • Automation using data mining and SWRL needs a complex implementation.
Summary

• Complex structures (patterns) make the data integration difficult.
• Pattern detection seems not efficient as the patterns vary through a dataset; however their concepts and functionalities remain the same.
• Ontology (OWL 2.0) can help us to embed the patterns in the ontology.
• Then concept defines the pattern in ontology, and it is representation independent.
• This introduces instance-level ontology which needs the users to produce it on-the-fly.
• The solution puts the main work load on the users but the results are more accurate and can be used in different applications.
• We believe the data should be collected in a cleverer way.
Future work

• developing an instance-level ontology which can meet our requirements and can help us in matching road networks.
• Integrating the instance-level ontology into our node-based algorithm and comparing the results with the former algorithms.
• Preparing an optimize questionnaire for a preliminary implementation of the ontology on-the-fly.
• Developing mechanisms as web services to aid the users for collecting information.