INTEGRATING AUTHORITATIVE AND VOLUNTEERED GEOGRAPHIC INFORMATION - AN ONTOLOGICAL APPROACH

Crowd Sourcing in National Mapping Internship Funding

ACTIVITY Workshop

Leuven (Belgium)
14th May 2013
Jimena Martínez Ramos
[jmr125@mun.ca]
Table of contents

1. Background
2. Problem
3. Objective
4. Proposed approach
5. Semantics in OSM datasets (ongoing work)
6. Conclusions and future work
Background

The need of integrating Geographic Information from different sources

Levels of heterogeneity

System

Syntactic

Structural

Semantic
(meaning of words)
The Problem

Semantic Heterogeneity in Geographic Information

Different conceptualizations: Semantic Heterogeneities

Words or symbols stand for things through ideas

Ogden and Richards, 1923
The Problem

Semantic Heterogeneity in Geographic Information

ONTOLOGIES!

CONCEPTUALIZATION

STANDARDS!

Reality (road)
The Problem
What are ontologies about?

“Explicit specification of a conceptualization”
Gruber, 1993
They are ways to conceptualize a domain.
Objective

Ontology and Standard approach to semantic integration

- **Objective**
  - **Usual approach** to geodata integration using ontologies
    - **Turnpike** = motorway
    - **Freeway** = motorway
    - **Freeway** + **Turnpike** = motorway

- **Common Standard conceptualization**
  - **VGI(OSM)**
    - **Dataset 1**
      - Freeway
    - **Dataset 2**
      - turnpike
  - **Interoperable datasets**
Proposed Approach

The method according the objectives

Official source 1

“freeway”

R2RML mapping

“turnpike”

R2RML mapping

“motorway”

R2RML mapping

Reality

OpenStreetMap

Common Knowledge
Ongoing work: Semantic Heterogeneity in OSM datasets

More than one tag per real-world phenomenon (synonymy)

Number of tags per phenomenon increases with the scale, and % is important

- <highway=bus_stop>
- <public_transport=stop_position>
Ongoing work: Semantic Heterogeneity in OSM datasets

More than one tag per real-world phenomenon (synonymy)

Number of tags per phenomenon evolve with time, and % is still important

- <highway=bus_stop>
- <public_transport=stop_position>

Decreasing level of agreement
Conclusions and future work

Proposed approach is based in a **domain ontology**, which allows:

- Matching datasets to a **common pivot** (R2RML allows **flexible** and direct mappings)
- **No need to know** how to handle **ontologies**.
- Reusing the **mappings**.

**Semantic Heterogeneity in OSM datasets.**

- **Number of tags and their % of occurrence** per real-world phenomenon
- **Time** and **spatial scale** are factors affecting SH in OSM datasets.

**Future work:**

- Developing more the ontology.
- **User-friendly interface** for making R2RML mappings.
- Deeper study **factors involved in SH** in OSM datasets, trying to model it.
Thank you. Gracias.

Questions?

Acknowledgements

AGILE/EuroSDR
NSERC
TU Delft
IGN Spain
Sinfogeo Ltd.

Dr. Jean Brodeur
Drs. Marian de Vries
Marine and Geomatics Lab colleagues

Jimena Martinez  [jmr125@mun.ca]