Modeling and Managing the Semantics of Geospatial Data and Services

Werner Kuhn, Martin Raubal, Michael Lutz
Institute for Geoinformatics
University of Münster (Germany)
{kuhn, raubal, m.lutz}@uni-muenster.de
Motivating Examples – „Distance“

![Map of North Korea's missile threat](image)

**North Korea’s missile threat**

<table>
<thead>
<tr>
<th>Type</th>
<th>Maximum range</th>
<th>Payload</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodong</td>
<td>1,300 km (810 miles)</td>
<td>700 kg (1,550 pounds)</td>
<td>Currently deployed</td>
</tr>
<tr>
<td>Taepodong-1</td>
<td>Up to 10,000 km</td>
<td>Several hundred kg</td>
<td>Test failed 1998, not yet operational</td>
</tr>
<tr>
<td>Taepodong-2</td>
<td>10,000-15,000 km</td>
<td>Several hundred kg</td>
<td>Not yet tested</td>
</tr>
</tbody>
</table>

Source: Task Force for US Korea Policy, Centre for International Policy
Motivating Examples – „Touch“

(a)  
1.  
2.  
3.  
4.  
5.  
6.  
7.  
8.  

(b)  
1.  
2.  
3.  
4.  
5.  
6.  
7.  
8.  

Geomedia  Oracle
Motivating Example – „Edge“

“Where can I cross the Havel?”

Road data (e.g., GDF)
Motivating Examples – „Wind Direction“

![Image of a fiery scene with black smoke]

Kuhn, Raubal & Lutz.  NCGIA Specialist Meeting „Spatial Webs“, Dec 02, 2004
Plant ID: 457
Emission rate: 29 mg/s
Plant location: R: 3405138, H: 5760997
Airport code: FMO
Dispersion map
Calculation gas dispersion plume
Wind report XML...
 semantic heterogeneities

1. same name, different data type, same domain concept
2. different name, same data type, same domain concept
3. same name, same data type, different or differently restricted domain concept(s)

Application Level

Name: „prevailing_direction“
Data Type: XML-ComplexType

Name: „prevailing_direction“
Data Type: „string“

Name: „wind“
Data Type: „string“

Name: „wind“
Data Type: „string“

„wind is blowing from“
points_to_compulsion → true

„wind is blowing to“
points_to_compulsion → false
Generalized Problem

- Buffers, topology, navigation, and gas plume examples show problems with the **semantics** of geographic information describing
  - entities (e.g., road, ferry)
  - processes (e.g., driving, wind)
  - relations (e.g., distance, touch)

- **Context** is essential for semantics
  - Today’s SDIs separate data from operations (e.g. GML)
  - context contained in operations is lost
  - systems or users misinterpret data

- **Ontologies** are supposed to provide that context
  - Today, they don’t
  - They could do better (see Position Paper Y. Bishr)
  - How?
What is Semantics?

Concept

activates

Symbol

stands for

Referent

relates to

“Tank”

[Ogden and Richards, 1923]
Semantics of what?

- We are *not* trying to capture the meaning of natural language expressions!
- We formalize the semantics of **technical symbols** used in GIS, databases, web services of **information communities**
- **But**: information is from and for humans
  - it derives from data through **human interpretation**
  - it emerges at the **user interfaces** of GI technologies
  - thus, we need to capture **meaning** (through cognitive semantics)
Medium-term research program (3-5 years)

solve 3 kinds of semantic interoperability problems:

- data contents
  - e.g., directions from road data
- service semantics
  - e.g., distance operators
- users
  - e.g., parcels touching a road
Semantic Interoperability

- ...is the only real interoperability
  interoperating components share an understanding of their interfaces

- today: „syntactic interoperability“
  interoperating components share an interface, defined by a type signature

\[
\text{GM\_Object :: distance (geometry : GM\_Object) : Distance}
\]
Today: Annotation

ISO 19107, Spatial Schema

„The operation "distance" shall return the distance between this GM_Object and another GM_Object. This distance is defined to be the greatest lower bound of the set of distances between all pairs of points that include one each from each of the two GM_Objects. A "distance" value shall be a non-negative number associated to a distance unit such as meter or standard foot. If necessary, the second geometric object shall be transformed into the same coordinate reference system as the first before the distance is calculated. (…)”

⇒ Natural Language Description
Goal: Axiomatization

Service interfaces, requests, and responses contain *symbols* with undefined semantics

\[ \text{GM\_Object} :: \text{distance} \ (\text{geometry} : \text{GM\_Object}) : \text{Distance} \]

\[ \text{Alexandria} :: \text{distance}_{\text{Stades}} \ (\text{Syene}) = 5040 \]

1. type symbols (standing for classes of objects and literals)
2. values (standing for individual objects and literals)
3. operators (standing for methods)
Ontology in Philosophy

a philosophical discipline—a branch of philosophy that deals with the nature and the organisation of reality

Science of Being (Aristotle, Metaphysics, IV, 1), asks the questions:

What exists?

What characterizes being?

Eventually, what is being?
Ontology in Computer Science

- An ontology is an engineering artifact. It is constituted by
  - a specific vocabulary used to describe a domain
  - assertions on the intended meaning of the vocabulary.
- Thus, an ontology describes a formal specification of a certain domain:
  - shared understanding of a domain of interest
  - formal and machine manipulable model of a domain

“An explicit specification of a conceptualization”
[Gruber93]
### (Informal) Ontologies

<table>
<thead>
<tr>
<th>Nr. Objektbereich</th>
<th>Nr. Objektgruppe</th>
</tr>
</thead>
<tbody>
<tr>
<td>6000 RELIEF</td>
<td>6200 Besondere Geländeoberflächenformen</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nr.  Objektart</th>
<th>Nr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>6201 Damm, Wall, Deich</td>
<td>6201</td>
</tr>
</tbody>
</table>

#### CAP landuse code | CAP landuse label
--- | ---
1 | Urban
1.A | Urban; Residential
1.A.0.a | Urban; Residential; Single family

**Legend**

- Buildings
- FeatureLabels
- Landmarks
- ParcelBoundaries
- RecreationalAreas
- ResortLabel
- Streets
- Trails

**CAP landuse code**

<table>
<thead>
<tr>
<th>CAP landuse code</th>
<th>CAP landuse label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Urban</td>
</tr>
<tr>
<td>1.A</td>
<td>Urban; Residential</td>
</tr>
<tr>
<td>1.A.0.a</td>
<td>Urban; Residential; Single family</td>
</tr>
</tbody>
</table>

**Systema naturae**

Naturae regna tria, secunda, secundum systematicam...

**Notes**

- Kuhn, Raubal & Lutz, NCGIA Specialist Meeting „Spatial Webs“, Dec 02, 2004
Are ontologies enough?

- They contain relationships between terms
  - superclass and subclass, is_a, part_of, has_property, synonym
- Their axioms are
  - absent or
  - in obscure and untested logical axioms
- They fail to capture context
  - activities determine semantics (e.g., driving across a river)
- They are not grounded
  - what does „artifact“ or „entity“ or „process“ mean?
The underlying problem

geoinformation = < x, z >

Spatial (& temporal) Reference Systems ???
Vision

Users of geographic information should be able to refer thematic data to semantic reference systems, just as they refer geometric data to spatial reference systems.

Software should support the
- referencing and grounding process
- projections to simpler semantic spaces
- semantic translation among different reference systems.
A GML file in 2010

... <PropertyName>SourcePoint</PropertyName>
<gml:Point srsName="EPSG:4326">
  <gml:coordinates>-76.9466745614984,38.58321344095582</gml:coordinates>
</gml:Point>

<PropertyName>EmissionRate</PropertyName>
<gml:Flow srsName="ECSS:2408">
  <gml:EmissionRate>29</gml:EmissionRate>
</gml:Flow>

...
Semantic Reference Systems

Spatial Reference System [ISO 19112]
„system for identifying positions in the real world“
e.g., geodetic reference systems

Temporal Reference System [ISO 19108]
„basis against which time is measured“
e.g., calendars

Semantic Reference System [t.b.d.]
basis on which thematic data are interpreted:
‘forest‘, ‘wetland‘, ‘road width‘ etc.
e.g., ontologies
Long-term research program (5-7 years)

- create **methods** and **tools** to design and use semantic reference systems for
  - *grounding* (e.g. “move”)  
  - *projecting* (e.g. roads and ferries to edges)  
  - *translating* (e.g. cadastre to navigation)

- based on real-world **case studies**
  transportation, emergency management, planning
Ferry example: *Grounding*
Grounding in Image Schemas

- Sensory-motor patterns of cognition (Johnson / Lakoff)
  - container, surface, path, link, center-periphery, force...
- developed through bodily experience
- have internal structure
- enable the perception and cognition of meaningful information from the environment
Ferry example: Semantic Projection

Edge
+navigate()

Road
+drive()

Ferry
+cross()

projection

Edge
+navigate()
Semantic Translation

Application Level

1. same name, different data type, same domain concept
2. different name, same data type, same domain concept
3. same name, same data type, different or differently restricted domain concept(s)

Name: "prevailing_direction"
Data Type: "string"
XML-ComplexType

Name: "prevailing_direction"
Data Type: "string"

Name: "wind"
Data Type: "string"

Name: "wind"
Data Type: "string"

"wind is blowing from" points_to_compulsion \( \rightarrow \) true
"wind is blowing to" points_to_compulsion \( \rightarrow \) false
The Way Forward

- Need to identify achievable small steps
- Solve them through experimental research (*not* big system implementations!)
- goal: support reasoning about semantics of GI
- Some steps:
  1. Annotate and discover data
  2. Annotate and discover services
  3. Translate between database schemata
  4. Translate between services
  5. Negotiate meaning
  6. Capture vagueness
  7. Acknowledge role of human perception
Semantic Heterogeneity in SDIs

Which terms can I use for an efficient keyword-based search?

• **Synonyms**: one concept – different terms
  water level, gauge, tide scale
  level, height
  measuring gauge, control point, name

  ⇒ The returned results are not exhaustive!

• **Homonyms**: one term – different concepts
  „water level“ – groundwater vs. surface water level

  ⇒ Not all returned results are relevant!
Semantic Heterogeneity in SDIs

**Retrieval**

*How do I interpret the terms of a data schema and formulate a query?*

<table>
<thead>
<tr>
<th>Feature</th>
<th>BafG</th>
<th>ELWIS</th>
<th>CHMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Pegelmessung</td>
<td>WasserstandMessung</td>
<td>StavVody</td>
</tr>
<tr>
<td>Water level in cm</td>
<td>wasserstand_cm</td>
<td>Pegelmessung</td>
<td>wasserstand_cm</td>
</tr>
<tr>
<td>Name of the control point</td>
<td>name</td>
<td>name</td>
<td>name</td>
</tr>
<tr>
<td>Feature type</td>
<td>CHMI</td>
<td>ELWIS</td>
<td>BafG</td>
</tr>
<tr>
<td>Date &amp; time of measurement</td>
<td>zeitpunkt</td>
<td>datum</td>
<td>datum</td>
</tr>
<tr>
<td>Date of measurement</td>
<td>datum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time of measurement</td>
<td>uhrzeit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geometry as point</td>
<td>gml:pointProperty</td>
<td>standort</td>
<td>gml:position</td>
</tr>
</tbody>
</table>

⇒ Meaning of terms used in the database schema often ambiguous
Demos

- Semantic Search for Geographic Data
  http://www.meanings.de
  (→ Use Cases → Tutorial)

- Prototype „Semantic Query Interface“
Conclusions

- **Spatial** reference systems motivate **semantic reference systems**
  - from static ontologies to computational reference systems

- Upper-level ontologies need **meaningful abstractions**
  - image schemas: *processes* in space and time

- Progress step-wise from information search through semantic translation to meaning negotiation
  - toward technically useful theories of meaning

- Ask first **what** needs to be said about space, then **how**.
For more information...

- MUSIL web site  
  (Muenster Semantic Interoperability Lab):  
  http://musil.uni-muenster.de

- e-mail to  
  {kuhn, raubal, m.lutz}@uni-muenster.de