

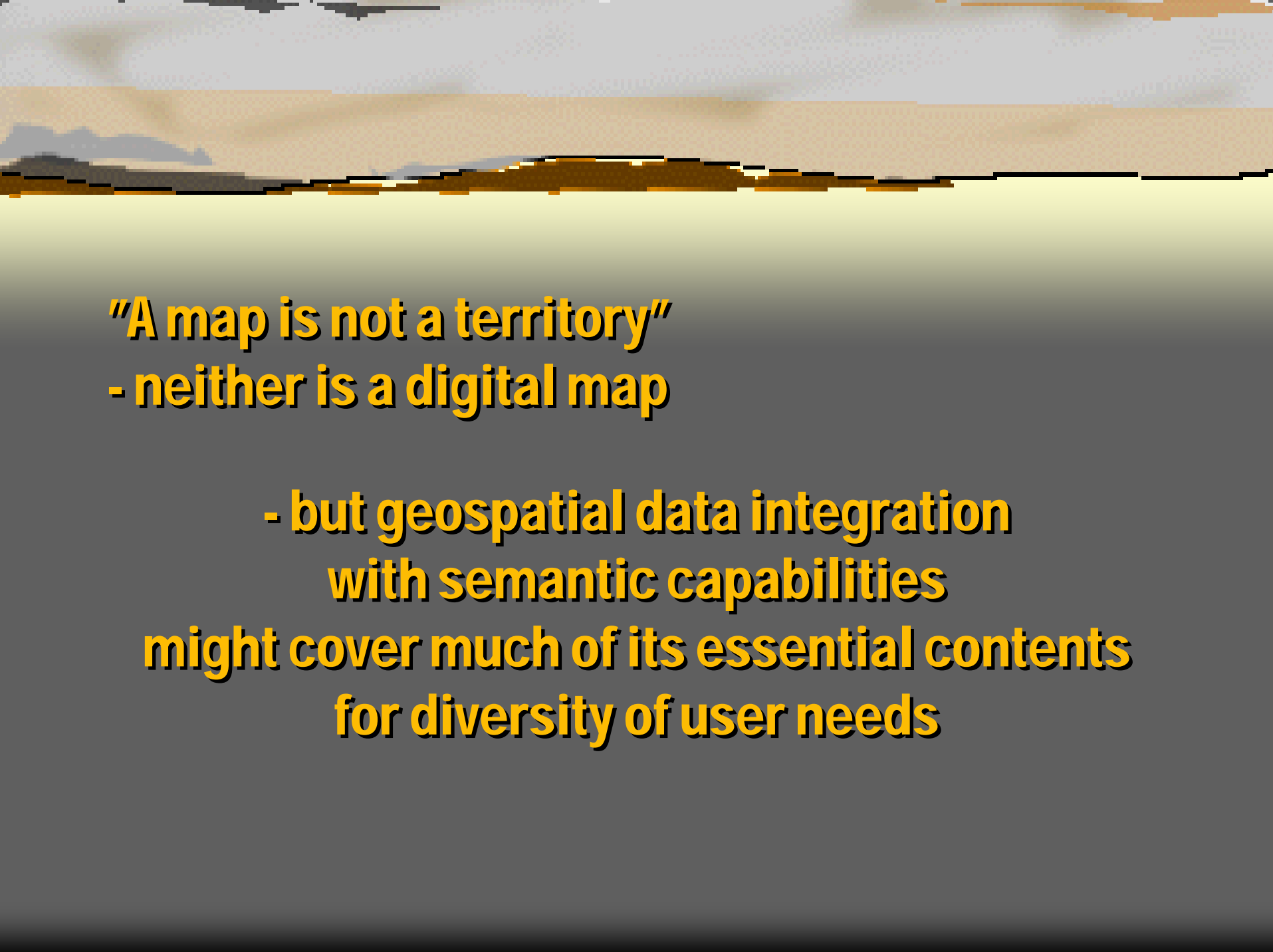
Spatial semantics for geoinformatics

Esa Tiainen

Eero Carlson



The Nordic GIS conference Helsinki 2 - 4 October 2006



**"A map is not a territory"
- neither is a digital map**

**- but geospatial data integration
with semantic capabilities
might cover much of its essential contents
for diversity of user needs**

Contents

- ⇒ Semantics: From implicit to explicit representation
- ⇒ Approaches of explicit representation
 - From geospatial repository to ontologies
- ⇒ Types of semantic heterogeneity problems
- ⇒ Semantic grounding
- ⇒ Search engines vs. rule engines
- ⇒ Conclusions - Rationality

Introduction – Semantics in Information Systems

- ⇒ Information systems present abstractions
 - Information is reduced to relevant for a particular purpose
 - Concepts are reduced to a necessary and adequate set of properties and property values
- ⇒ Data is context-dependent – the semantics are implicit
- ⇒ Data discovery and interoperability require explicit semantics (explicit representations)
 - Explicit semantics describe relations of concepts

Different approaches of explicit representation

Methodology

- ⇒ Semantic similarity
- ⇒ Geosemantic proximity

Applications

- ⇒ Catalogues with shared vocabularies
 - Problems with vocabulary management
- ⇒ Geospatial repository and ontologies
- ⇒ Schema level semantic annotation

Semantic similarity

- ⇒ Distances of selected properties in conceptual space can be measured mathematically
- ⇒ Familiar example:
 - color spaces with specific colors defined in RGB or HSV coordinates illustrates relationships in conceptual space
- ⇒ Applicable to analyze how different geospatial concepts like plateau, lowland, hill and mountain relate to each other
 - for creating concept maps with geometric topology along the selected quality dimensions

Semantic similarity continued

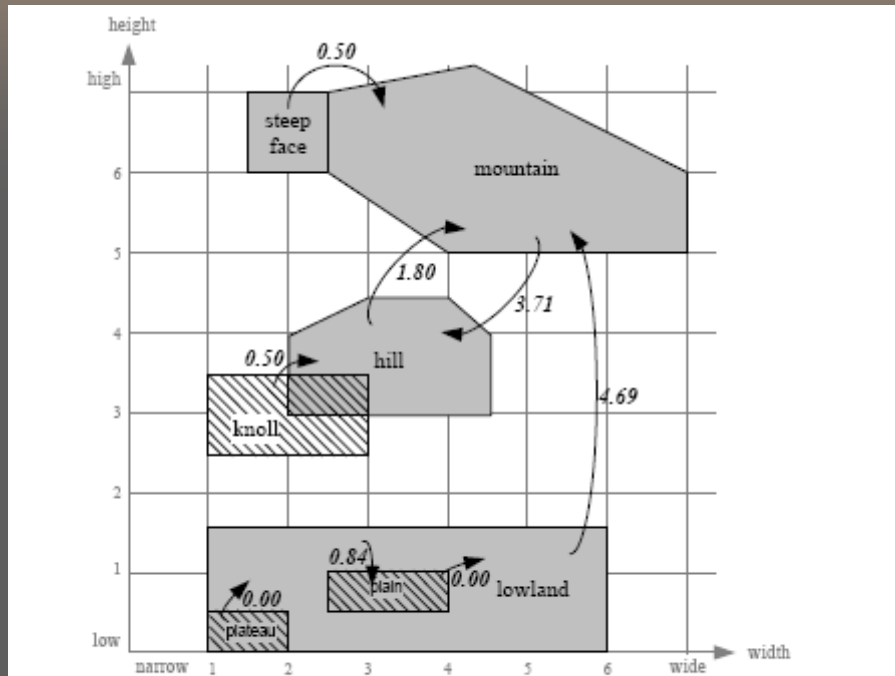


Fig. 13. Similarity values between concepts with different topologic relations in one conceptual space.

(Angela Schwering and Martin Raubal: Measuring Semantic Similarity between Geospatial Conceptual Regions: GeoSpatial Semantics, Lecture Notes in Computer Science 3799/2005)

The example shows that semantic similarity is a way to analyse how different geospatial concepts relate to each other (meet, overlap, cover, include, disjoint) for e.g. creating concept maps, taxonomies, ontologies etc

- ➔ Regions in the conceptual space are represented with prototype vectors
- ➔ However it seems difficult to make a computational reasoning application based on semantic similarity

Geosemantic proximity

- ⇒ The notion of geosemantic proximity suggests that a concept and a conceptual representation are made of
 - intrinsic (implicit) properties providing identity
 - extrinsic (explicit) properties providing its relationships to other concepts - presenting the functionality and restricting the scope of the concept
- ⇒ Thus it includes both aspects of semantics – implicit and explicit
 - Connecting concepts to the context concerned, a context-based approach

Geosemantic proximity continued

- ⇒ Intrinsic properties e.g.(for a concept road)
 - classification type (e.g. highway, main, secondary etc.),
 - surface type (e.g. paved or unpaved),
 - road number or road name, and geometric representation (type)
- ⇒ Extrinsic properties (road)
 - describe its relationships with other features such as built-up areas, railways, bridges, ferry routes,
 - present the boundaries of the concept (like a domain vocabulary or shared vocabulary)
- ⇒ The similarity between concept K's context and its conceptual representation L's context as a four-intersection matrix:

$$GsP(K,L) = \begin{bmatrix} \partial C_K \cap \partial C_L & \partial C_K \cap C_L^\circ \\ C_K^\circ \cap \partial C_L & C_K^\circ \cap C_L^\circ \end{bmatrix}$$

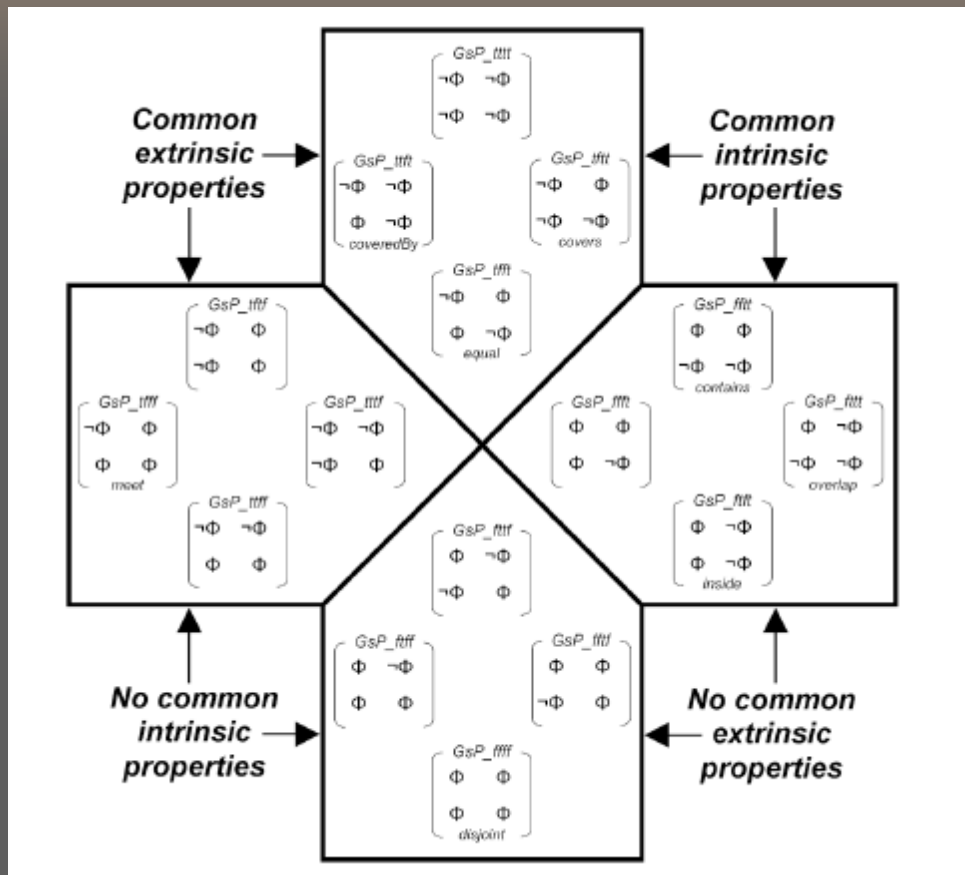
Where:

C_K = Context of concept K

C_K° = Intrinsic properties of C_K

∂C_K = Extrinsic properties of C_K

Geosemantic proximity continued



The four-intersection matrix of intrinsic and extrinsic properties unfolding to sixteen cases characterizing geosemantic proximity relationships like equal, meet, overlap, contains, include, covers, covered by or disjoint.

Geosemantic proximity, e.g. Brodeur, Bédard, Moulin: Geosemantic Proximity to Improve Geospatial Information Discovery in a Wireless Environment, *Geomatica*, vol. 57, no 1, 2003, pp. 49-57

⇒ By adding the geosemantic proximity functionalities e.g. upon a geospatial (data) repository these are extended to facilitate geospatial semantic interoperability

Catalogues with shared vocabularies

- ⇒ Search engines may use OGC-compliant catalogues with terminological reasoning
 - by annotation (referencing the data) to shared vocabularies or thesauri
 - more automatically by using ontology based reasoning components
- ⇒ may be effective with naming heterogeneity (synonyms)
- ⇒ but on property level the required property might be included or excluded – resulting diffuse information

Problems with vocabulary management

- ⇒ The shared vocabulary
 - General enough to cover all information sources,
 - - but specific enough for make meaningful definitions
- ⇒ Leads to need to use domain vocabularies
 - but boundaries of concept domains may be unclear
 - as well as for concepts and terms for social constructions that often have cross-domain nature
- ⇒ The semantic grounding remains obscure
 - Frequent updating of annotations becomes crucial to manage
- ⇒ The task of defining the correspondences, mapping or annotation lies in the responsibility of the information source (provider)

From geospatial repository to ontologies

- ⇒ Geospatial repository with metadata structured to provide the semantics enables geospatial data integration e.g. in a web query
 - "structured metadata" refers to an ontology of the database
- ⇒ Ontology - formulation of structure-preserving mappings
 - **Mapping** – specification of which concepts and relations correspond to which other concepts and relations
 - **Ontology translation** – translation of one ontology into the other
 - **Ontology integration** - composition of ontologies to build new ones (whose respective vocabulary are usually not interpreted in the original domain)

Schema level semantic annotation

- ⇒ Extending the annotation to application schema enables to cover the properties included
 - allows use of domain-independent ontologies, and self-defined query concepts (from familiar shared vocabularies)
- ⇒ A terminological reasoner may allow the classification of data into another context
 - e.g. by equality and subsumption
- ⇒ Components needed
 - creating and incorporating application ontologies
 - user interface supporting connection to application schema
 - an enhanced “cascading catalogue service” that is “aware” of the application ontologies
 - functionalities for manipulating the filters of metadata
 - support of WFS queries for unknown application schemas

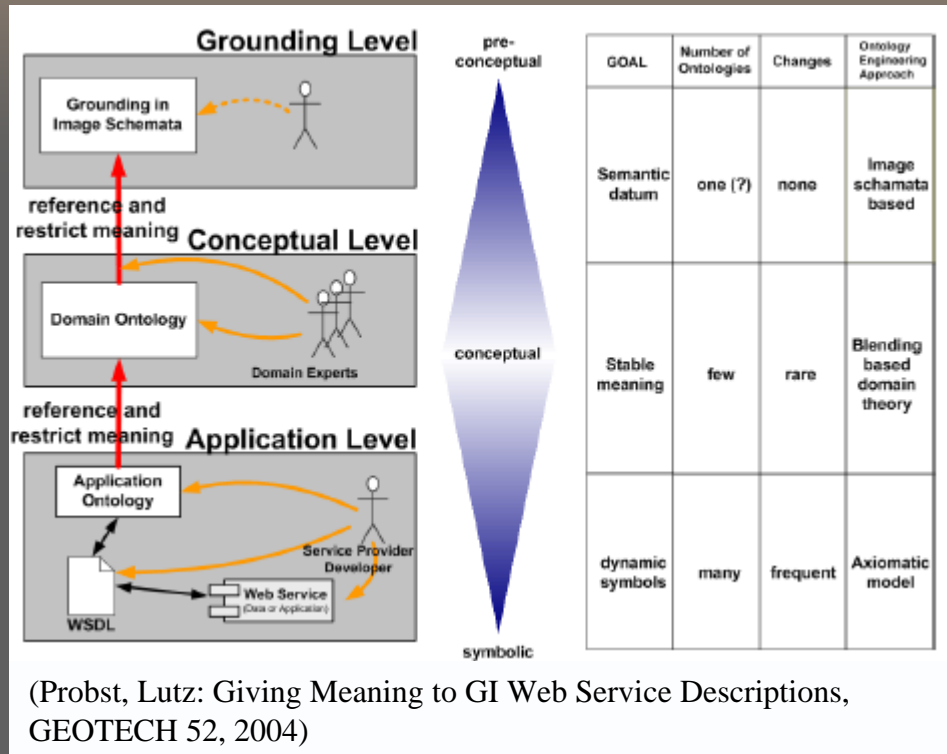
Types of semantic heterogeneity problems

- ⇒ Naming heterogeneity with synonyms
 - Can be solved by annotating the elements used in the WSDL descriptions with concepts from an application ontology
- ⇒ Data type heterogeneity
 - Items represented with different data types; a semantic description of rules is necessary for transformation or extracting the required information
 - How to semantically describe complex types to generate these rules automatically from the descriptions for appropriate parsers
- ⇒ Conceptual heterogeneity with homonyms (when the information refers to different domain concepts)
 - E.g. water level may refer to groundwater or surface water levels
 - Identical descriptions must imply the same meaning e.g. by referencing the application level concepts to specified domain concepts (or in details of property level)

Need for semantic grounding

- ⇒ As noted earlier the boundaries of concept domains maybe unclear or overlapping
- ⇒ Conceptual heterogeneity is solved on schema level semantic annotation
 - by referencing the application level concepts to different domain concepts, or
 - by using different restrictions on the same domain concepts to prevent inconsistency in a service chain
- ⇒ To avoid defining concepts with other concepts on an ever or endlessly higher level of abstraction Kuhn (2003) has introduced idea of semantic reference system

Semantic Reference System



The application and domain levels consist of ontologies, i.e. explicit specifications of a conceptualization.

The application level uses semantic-free symbols of WSDL descriptions to which meaning is already assigned in domain ontology, which in turn grounds its concepts on an image schemata-based grounding level.

- ➔ A user can find a concept of a domain ontology vocabulary and restrictions of its meaning on the application ontology
 - The cognitive task of assigning meaning to a symbol remains with the user
- ➔ Extending concepts with new restrictions or changing existing restrictions can cause consistency problems on the application level

Search vs. Rule engines

- ⇒ Search engines with thesauri or keywords
 - provide diffuse information
- ⇒ Rule engines with ontologies
 - provide meaningful (connotated) information or knowledge
 - ontologies or ontology transformation possible within
 - schema level semantic annotation
 - geosemantic proximity –tool etc.

Conclusions - Rationality

- ⇒ Economic investigations: 26% better return of investment (ROI) with improved interoperability of GI
(Booz Allen Hamilton, 2005: Geospatial Interoperability ROI Study)
 - Calculations include larger investments e.g. on semantics

- ⇒ Benefits with management of semantics
 - Save in **harmonization** efforts by management of semantics
 - Improved **information resource management**
 - in information services, product level (personalised or customized services), generalization issues, updating and distributed data capture . . .
 - A lot of spatial analysis and data integration is within reach of semantics and ontology based methods

Thank you for your attention

- Questions?
- Answers?
- Comments?

Contacts: [esa.tiainen at nls.fi](mailto:esa.tiainen@nls.fi)

