

Ontology-driven geographic information integration

Pablo R. Fillottrani¹

joint work with Agustina Buccella² Alejandra Cechich²

¹Depto. Cs. e Ing. Computación
Universidad Nacional del Sur
Bahía Blanca, Argentina

²Depto. Cs. Computación
Universidad Nacional del Comahue
Neuquén, Argentina

FUB, October 2008



Ontology-driven geographic information integration

- 1 Objectives
- 2 Comparison Criteria
- 3 Geographic information system integration
 - Systems and methodologies
- 4 Evaluation
- 5 Conclusions
- 6 References



Geographic information systems

- **geographic information systems** (GIS), possible distributed on the web, are nowadays widely available
- usually GIS rely on distributed (heterogeneous) databases
- **ontologies** are extensively proposed as a tool to resolve heterogeneity, even in the geographic domain



Need for integration

- the **province of Neuquén**, Argentina through its Servicio de Hidrología, analyses the integration of the geographic databases in the province



- different agencies with different points of view and vocabularies maintain heterogeneous geographic databases
- a one-year project is being funded to evaluate the situation and suggest solutions
- several systems and methodologies have been recently proposed



Basic concepts

- several types of **geographic data heterogeneity**
 - conceptual model
 - spatial model
 - schema
 - semantic
- generally accepted shift from integrating information sources to **integrating ontologies**



Comparison criteria

- three main aspects to evaluate systems have been defined
 - **semantic representation**
 - **integration process**
 - **geographic information**



Semantic representation aspects

- 1 **ontology language**: formal, non-formal
- 2 **expressiveness**: low, medium, high
- 3 **ontological components**: source ontologies, top-level ontology, other ontologies



Integration process and Geographical aspects

- integration process aspects
 - 1 **mapping discovery**: manual, semiautomatic
 - 2 **mapping approach**: LAV, GAV, GLAV
- geographical aspects
 - 1 **recognition of geographic information**: yes, no
 - 2 **geographic information model**: topological relations, qualitative model, other



Selection

- systems and methodologies were selected for comparison because of their
 - relevance
 - novelty
 - availability



Kavouras et al.

- Kavouras et al. (2003) presents a methodology to explore and identify semantic information provided in geographic ontologies
- NLP techniques are used to extract semantic relations from specific geographic data sources: CORINE LC, GDDD and WordNet
- these relations are fixed, like *is-part-of*, *adjacent-to*
- **similarity measures** are used to determine equivalent terms, and their relations



BUSTER

- the **BUSTER** (Bremen University Semantic Translator for Enhanced Retrieval) system is a middleware for semantic geographic data integration (Visser et al. 2004)
- consists of two parts: a tool for intelligent information retrieval, and a tool for semantic data integration
- **wrappers** are used in the syntactic level to uniform source heterogeneity, and **mediators** are used to combine, integrate and abstract information from the wrappers
- a third level of integration is present: **rules** and **reclassification** of terms transform data from source context to goal context
- uses **place name structures** model for reasoning with spatial data



Schwering and Raubal

- Schwering and Raubal (2005) introduce a query method based on spatial relations to integrate information sources
- a **shared vocabulary** is described, queries must be formulated using its terms and relations
- there is a mapping between this shared vocabulary, and different source ontologies (fixed)
- **dimensions** (MDS) are used to relate concepts in the shared vocabulary
- also, similarities are calculated by using distance functions



Sotnykova et al.

- Sotnykova et al. (2005) propose a methodology for the integration of spatio-temporal schemas
- source ontology must be represented in the **MADS model**
- it is based on the specification of **interschema** model by an expert user, in three steps
 - 1 translation of MADS conceptual models to OWL DL
 - 2 definition of semantic correspondences (SC) between terms, properties and instances, by an expert designer
 - 3 four different patterns can be used to integrate SC
- it uses RACER for validating the three steps of the integration process



Hakimpour et al.

- Hakimpour et al (2003) introduce an architecture and a methodology for geographical schema integration
- based on DL reasoning, finds similarities between semantically related terms in different schemas
- the **architecture** contain **source ontologies**, a **global schema**, and a **reasoning system for merging ontologies**
- mappings are such that source terms are defined as views over the global schema
- mappings consist of four **similarity relations**: equality, disjoint, overlapping, specialization
- the **methodology** suggests expert users a set of steps to create the global integrated schema



Hess and lochpe

- Hess and lochpe (2004) define a methodology for semantic integration of geographic conceptual schemas
- two steps: translation and integration, two levels: syntactic and semantic
- **translation** builds the global ontology, first translating schemas to GML and then building the global ontology
- different similarities functions are used both in the syntactic and semantic **integration**
- global ontology acts as a mediator between local ontologies, with taxonomic, aggregation and composition relationships
- conflicts and heterogeneities not solved are managed by expert users



MDSM methodology

- Rodriguez and Egenhofer (2004) present the **Matching Distance Similarity Measure**, a technique for determining semantic similarity among spatial entities
- they attach three **features** (parts, functions, attributes), and three **semantic relations** (synonym, hypernym, meronym) to each concept
- **similarity assessment** is done by a combination of feature matching and semantic distance functions
- the results are combined in a $\langle value, weight \rangle$ sum
- a global ontology is defined by an expert user using terms from local ontologies, thesaurus and the similarity results



ODGIS system

- the **Ontology-Driven Geographic Information System** (Fonseca, 2001) is a framework for the integration of geographic databases
- two steps: **knowledge generation** (ontology building) and **knowledge use** (query answering)
- a multilevel ontology is built in a top down fashion, using different integration mechanisms (is-a, role playing, MSDM similarity) at each level
- lower level ontologies represent task-specific ontologies, source terms are generally mapped to middle-level ontologies



MDSM+TR methodology

- Janowicz (2005) proposes an extension of MDSM with **thematic roles** (TR) within the functional feature
- a TR is a semantic relation between a predicate and an argument: like **agent**, **theme/patient**, **goal**, **location**, etc
- predicates and terms are classified according patterns of TRs
- similarity assessment is modified with new functions that consider TRs



GeoNis system

- Stoimenov et al. (2006) propose a semantic mediator architecture in the **GeoNis** system
- an **offline methodology** discovers mappings between concepts from local geographic ontologies, based on a common reference model and building a global ontology
- similar to Hakimpur et al., mappings are defined by predefined relations
- mappings are classified as: **direct**, **indirect** and **across**
- a reasoner is used for discovering mappings, that can be extended by expert users with the help of a rule engine



Aerts et al.

- Aerts et al. (2006) describe a methodology for integrating heterogeneous topographic databases
- local ontologies are described in OWL using terms from a top-level ontology
- **Formal Concepts Analysis** (FCA) provides heuristics for characterizing necessary and sufficient conditions in each definition
- properties are classified as **classifying** and **non-classifying**
- all this information is used by a reasoner to extract hierarchical and equivalent relations



Ontological aspects

	Ontologies		
	Formal	Expressive	Components
BUSTER	yes	medium	source ontologies
Kavouras et al.	no	medium	source ontologies
Schwering et al.	no	medium	source ontologies
Sotnykova et al.	yes	high	source ontologies
Hakimpour et al.	yes	low	source & a top-level
Hess and lochpe	no	medium	source & a top-level
MDSM	no	medium	source ontologies
ODGIS	no	low	different levels
MDSM + TR	no	medium	source ontologies
GeoNis	yes	medium	source & a top-level
Aerts et al.	yes	low	source & a top-level



Geographic information aspects

	Geographic model
BUSTER	qualitative model
Kavouras et al.	topological rel.
Schwering et al.	topological rel.
Sotnykova et al.	MADS model
Hakimpour et al.	no
Hess and lochpe	no
MDSM	no
ODGIS	no
MDSM + TR	no
GeoNis	topological rel.
Aerts et al.	topological rel.



Integration process aspects

	Process		Mapping
	manual	inferences	
BUSTER	yes	yes	GAV
Kavouras et al.	no	no	GAV
Schwering et al.	no	no	GAV
Sotnykova et al.	yes	no	GAV
Hakimpour et al.	no	yes	LAV
Hess and lochpe	no	no	LAV
MDSM	no	no	GAV
ODGIS	no	no	GLAV
MDSM + TR	no	no	GAV
GeoNis	no	yes	GAV
Aerts et al.	no	yes	LAV



Conclusions

- ontologies is a general accepted tool for geographic data integration
- ad-hoc mapping discovery is common
- few approaches take advantage of formal ontologies for discovering or validating mappings; a top-level ontology is needed in this case
- expert users are always needed in all solutions



Future trends

- geographic entities are ontologically different from non-geographic ones. Ontologies must define coherent ways to include both types of concepts
- formal reasoning over mixed ontologies
- multi-level integration process



Thank you!



References

- “*Ontology-driven geographic information integration*”, A. Buccella, A. Gechich, P. Fillotrani. Computers & Geosciences. To be published
- “*Improving geographic information integration through semantic enrichment*”, A. Buccella, A. Gechich, P. Fillotrani. CACIC 2008 Argentinian Congress on Computer Sciences. Chilecito, Argentina.
- “*Integración de Sistemas de Información Geográfica*”, A. Buccella, A. Gechich, P. Fillotrani. WICC 2007. 9no. Workshop de Investigadores en Ciencias de la Computación. Trelew, Argentina.

