Draft version from the International Journal of Geographical Information Science, 15(7): 679-687, 2001.

FUSION OF TOP-LEVEL AND GEOGRAPHIC DOMAIN ONTOLOGIES BASED ON CONTEXT FORMATION AND COMPLEMENTARITY

MARGARITA KOKLA and MARINOS KAVOURAS

National Technical University of Athens,

15780 Zografos Campus, Athens, Greece,

Tel: 30+1+772-{2637, 2731} Fax: 30+1+772-2734

e-mail: {mkokla, mkav}@survey.ntua.gr

Keywords: geographic ontologies, principle of complementarity, fusion of contexts, concept lattices

Short title: Fusion of top-level and geographic domain ontologies

For correspondence:

MARGARITA KOKLA

National Technical University of Athens,

15780 Zografos Campus, Athens, Greece,

Tel: 30+1+772-2637 Fax: 30+1+772-2734

e-mail: mkokla@survey.ntua.gr

FUSION OF TOP-LEVEL AND GEOGRAPHIC DOMAIN ONTOLOGIES BASED ON CONTEXT FORMATION AND COMPLEMENTARITY

Abstract

This paper proposes a methodology for the fusion of different geographic domain ontologies with top-level ontologies, in order to provide a solid base for information exchange. The proposed methodology discusses context formation and the formalization of geographic concepts using essential properties.

Keywords: geographic ontologies, principle of complementarity, fusion of contexts, concept lattices.

1. Introduction

In order to achieve information exchange between different geographic databases, it is necessary to develop suitable methods for formally defining and representing geographic knowledge. However, the plethora and diversity of data standards and terminologies representing different geographic concepts further complicate the problem of geographic information sharing and reuse. Semantic differences occur between heterogeneous geographic data standards and raise problems during the integration process.

The study of ontology, both from the philosophical and the computer science perspective, may contribute to the unification of different conceptualizations of geographic space into an ultimate geographic ontology. However, this integration can be accomplished only if these ontologies are embedded within a more general, top-level ontology, which provides a solid framework for more specialized applications (Guarino 1998; Sowa 2000)

The methodology presented in this paper, focuses on fusing different geographic ontologies with more general top-level ontologies. The methodology uses Concept Lattices as a tool for the formalization and integration of geographic concepts and relationships encoded in different ontologies, in order to reveal their association and interaction.

The remainder of the paper is organized as follows. Section 2 introduces the principle of complementarity, as well as the concept of context construed for the geographic domain. Section 3 analyzes some of the most important components of geographic contexts, especially properties and how they can contribute to the semantic definition of geographic concepts. Section 4 describes the proposed methodology for the formation and integration of geographic contexts. Finally, an overall evaluation of the proposed methodology is presented in Section 5.

2. Principle of complementarity and contexts

Ontology, from the philosophical, as well as from the computer science perspective, is considered as an important step towards information exchange. For the geographic domain, the plethora and diversity of categorizations is highly dependent on human partition and conceptualization of geographic space. However, geographic categories are intrinsically related within space, and thus form categorical systems that interact (Smith and Mark 1998). This important characteristic of geographic categorizations can be paralleled to the principle of complementarity, introduced by N. Bohr (1934). According to this principle, different views represent different aspects of the world, each of which may be appropriate in some context. Therefore, different categorizations lead to complementary ontologies and this complementarity is desirable, since it accentuates different but interrelated aspects of geographic space. Complementary ontologies may be associated, but only when the contexts are explicitly distinguished (Sowa 2000).

The term "context" is used differently in various sciences, such as linguistics, philosophy, cognitive science, and artificial intelligence. Distinctions between "objective" or "metaphysical" context, and "subjective" or "cognitive" context, as well as between "linguistic" and "semantic context" can be found in the literature (Penco 1999; Sowa 2000). We follow the latter distinction, in which the term "context" acquires two senses. It is used to denote, either the discourse that surrounds a word or passage and helps to determine its interpretation (linguistic, syntactic function of context), or the setting, environment, domain in which an entity or topic of interest exists or occurs (semantic function of context). In the geographic domain, we are mainly interested in the semantic, non-linguistic function of context.

3. Components of geographic contexts

Geographic contexts include information about geographic concept types, characteristics, relations, operations, etc. Concept types can be defined as abstract specifications of entities that exist or may exist in some domain. Properties are the attributes, features, or characteristics of entities. Properties distinguish the concept types, which they characterize. According to the Stanford Encyclopedia of Philosophy, "properties have an important role to play in explaining our ability to recognize and categorize things in the world around us". Although philosophy has a great interest in the nature and existence of properties, there are controversial examples of the sorts of properties that may exist. "Properties" as used in the philosophical sense do not necessarily coincide with "attributes", but rather refer to the property of being something, e.g. the "property" of being "a lake".

Some distinctions of property kinds (Stanford Encyclopedia of Philosophy; Bigelow 2000; Guarino and Welty 2000a; Guarino and Welty 2000b) can contribute to the definition of geographic concepts. Particularizing properties (sortal predicates), for instance, designate the identity of the object and determine the principles by which it can be distinguished from other objects, and thus allow us to count objects (e.g., buildings). They are contrasted with characterizing properties (e.g., high), which characterize individuals, as well as with mass properties (e.g., cement, water), which apply to stuff. Another distinction, which however is a relative one, is that between determinables and determinates. Determinables are general properties (e.g., color, shape), while determinates are more specific versions of these (e.g., red, triangular). Natural kind properties (e.g., river) are important properties that carve nature at its natural joints and are contrasted with artificial properties (e.g., road). Essential or rigid properties are those properties that individuals have in any case in which they exist. For example, a riverbed may be considered as an essential property of a river. Qualities (e.g., distance, volume, area, speed, and temperature), as well as roles (e.g., land use vs. land cover, hotel vs. building) are also considered as sorts of properties.

In the geographic domain, importance mainly lies in the essential properties of geographic categories, which can ensure their semantic and univocal definition. Nevertheless, it is not always easy to determine essential properties for geographic concepts. Thus, the question arises whether rigid properties are the only way to create identity, or a certain combination of non-rigid properties can still provide identity to geographic concepts. For example, the properties of being a "wetland", an "arable land", "flat", "flooded" and containing "irrigation channels", although individually are non-essential, may in combination quite confidently (rigidly) identify "rice field". According to Elder (1998), "any distinctive or peculiar property, essential to an individual object must go together with yet other properties, which likewise distinguish that object with other similar ones".

Semantic relationships (Miller et al. 1993) between two or more concepts are another important component of geographic contexts and can be grouped to three major categories. *Synonymy* refers to similarity in meaning, as for example between categories "stream" and "watercourse". *Hyponymy*, *subtype/supertype*, or *kind of* relation is the subordination/superordination relation. The hyponym inherits all the characteristics of the more generic concept and adds at least one characteristic that distinguishes it from it and from the other hyponyms of that superordinate, e.g., a "river" is a kind of "stream". *Meronymy/holonymy* is the part-whole relation, e.g., "midstream", "ford" and "meander" are parts of "stream".

4. Proposed methodology

Geographic ontologies exhibit heterogeneities, such as differences in the level of detail and the relationships defined between concept types, inconsistencies in definitions, and incomplete specifications of characteristics. In order to resolve these heterogeneities and provide a solid base for the fusion of different domain ontologies, the integration should be based on a more general top-level ontology.

The methodology presented in this paper focuses on the formation of geographic contexts using Formal Concept Analysis (Wille 1992, Ganter and Wille 1999), as well as on the integration of existing geographic domain ontologies and their association with a top-level ontology.

The methodology assumes a formalization of geographic concepts consisting of two parts: the *extension* and the *intension*. The extension includes the entities, which belong to the concept, whereas the intension represents its intrinsic meaning and is usually described in terms of discriminating characteristics, properties or criteria. The semantic basis for geographic concept types is usually intensional. However, a solid intensional basis is provided mainly by essential properties.

The methodology achieves the integration of geographic ontologies, exhibiting differences in application context and thematic resolution. Related categories in different ontologies may be associated with subtype/supertype or overlapping relationships. The methodology is demonstrated using an example of integrating the concept type "stream" as defined by three different ontologies (figure 1): CYC top-level ontology, WORDNET and SDTS.

CYC
Stream: natural, flowing body of water, including everything from great rivers to tiny creeks
• canal: artificial waterway created to be paths for boats, or for irrigation
• river: natural stream of water, normally of a large volume
WordNet
Stream, watercourse: natural body of running water flowing on or under the earth
• river: large natural stream of water
• brook, creek: natural stream of water smaller than a river (and often the tributary of a river)
• branch: stream or river connected to a larger one
• rivulet, rill, run, runnel, streamlet: small stream
Channel: a passage for water (or other fluids)
• canal: long and narrow strip of water made for boats or for irrigation
• gutter, trough: a channel along the eaves or on the roof; collects and carries away rainwater
• spillway, spill: a channel that carries excess water over or around a dam or other obstruction
SDTS
Watercourse: a way or course through which water may or does flow
• branch
• brook
• canal
• creek
• ditch
• gutter
• river
• rivulet
• spillway

Figure 1. Category "stream-watercourse" as defined by CYC, WordNet and SDTS.

The process of integrating multiple ontologies is divided in two main steps: *Semantic Factoring* and *Concept Lattices*. Semantic factoring is the process of analyzing-decomposing the categories of the original ontologies into a set of fundamental categories, which are called semantic factors (figure 2). The case of an overlap between categories is resolved by splitting them into disjoint classes. Their common part forms a new class. Thus, complex concepts are decomposed into the simpler concepts out of which they are constructed. For the running example (figure 2), categories "stream" or "watercourse", as defined by the three ontologies, represent different, overlapping concepts. Semantic factoring decomposes them into simpler, unambiguous concepts (semantic factors), such as "canal" and "river". More specifically, categories "stream", as defined by CYC and WordNet are not equivalent, but overlap. Their overlapping part corresponds to the semantic factor "river" (s_2). Semantic factors, such as "canal" (s_1), "brook" (s_3) and "creek" (s_4) are those that differentiate the two superordinate categories.

		Semantic Factors								
	Original Categories	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9
CYC	**Stream**	x	x							
	canal	x								
	river		x							
	Path artifact	x								
	canal	x								
WORDNET	**Stream, watercourse**		x	x	x	x	x			
	river		x							
	brook			x						
	creek				x					
	branch					x				
	rivulet						x			
	Channel	x						x	x	
	canal	x								
	gutter							x		
	spillway								x	
SDTS	**watercourse**	x	x	x	x	x	x	x	x	x
	branch					x				
	brook			x						
	canal	x								
	creek				x					
	ditch									x
	gutter							x		
	river		x							
	rivulet						x			
	spillway								x	

Figure 2. Semantic Factoring

Similarly, in the property attribution matrix (figure 3), original categories are assigned properties that distinguish them from other categories. In this particular example, properties were derived from CYC and SDTS ontologies, since WordNet does not include properties to the definition of categories. However, essential properties, in case they can be defined, are the most eligible to be used in the methodology, since they unambiguously define categories and distinguish them from other categories.

	Properties																	
		in-ContOpen	waterage	hasAsTributary	path type	shipping	direction of flow	controlling depth	wetted perimeter	form_ratio	branch/parent	commercial_shipping	force_of_flow	flooded	volume	navigable	seasonal_depth	discharge
	Original Categories	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	a_{17}
CYC	**Stream**	x																
	canal	x	x															
	river	x		x														
	Path artifact				x													
	canal				x	x												
WORDNET	**Stream, watercourse**						x											
	river						x											
	brook						x											
	creek						x											
	branch						x											
	rivulet						x											
	Channel							x										
	canal							x										
	gutter							x										
	spillway							x										
SDTS	**watercourse**								x									
	branch								x	x								
	brook								x		x							
	canal								x			x						
	creek								x				x					
	ditch								x					x				
	gutter								x						x			
	river								x							x		
	rivulet								x								x	
	spillway								x									x

Figure 3. Property Attribution Matrix

The cross-table of the integrated context (figure 4) associates the semantic factors with the properties assigned to the categories out of which semantic factors were derived.

Semantic Factors	Properties																
	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	a_{17}
s_1	x	x		x	x		x	x			x						
s_2	x		x			x		x							x		
s_3						x		x		x							
s_4						x		x				x					
s_5						x		x	x								
s_6						x		x								x	
s_7							x	x						x			
s_8							x	x									x
s_9								x					x				

Figure 4. Cross-table of the integrated context

Then, Formal Concept Analysis (Wille 1992; Ganter and Wille 1999) is applied, in order to combine the semantic factors and their properties and generate what is called a Concept Lattice. The basic concepts of Formal Concept Analysis are:

- A *Formal Context* (G, M, I) is a set of objects G, a set of attributes M and a binary incidence relation I.
- An *Incidence Relation* I, or *gIm* is the connection between objects and attributes
- A *Formal Concept*, *Conceptual Class* or *Category* is a collection of entities or objects exhibiting one or more common properties or characteristics:

  A pair (A, B) is a Formal Concept of the context (G, M, I), where A is called the *extent* and B the *intent* of the formal concept.
- A *Superconcept/subconcept relation* is the order proceeding top-down from more generalized concepts with larger extent and smaller intent to more specialized concepts with smaller extent and larger intent:

  $(A_1, B_1) \leq (A_2, B_2)$ if $A_1 \subseteq A_2$.

- A *Concept Lattice* {B (G, M, I); ≤} is the ordered set of all formal concepts of a formal context.

The algorithm for producing a concept lattice from a cross-table, such as that of the integrated context shown in figure 4, includes several steps. Firstly, for $X \subseteq G$ and $Y \subseteq M$, formal concepts of the context (G, M, I) are derived using the following operations:

$X' = \{m \in M \mid gIm \text{ for all } g \in X\}$ (1)

$Y' = \{g \in G \mid gIm \text{ for all } m \in Y\}$ (2)

$C_i = (X'', X') \cup (Y', Y'')$ (3)

X' is the set of the attributes describing each object in X, whereas Y' is the set of objects possessing attributes in Y. For the running example, X' (1) and Y' (2) take the following form:

$X' = \{\{s_1\}', \{s_2\}', \{s_3\}', \ldots\ldots, \{s_9\}'\}$,

$Y' = \{\{a_1\}', \{a_2\}', \{a_3\}', \ldots\ldots, \{a_{17}\}'\}$,

where $\{s_1\}'$ is the set of attributes assigned to the semantic factor s_1 ("canal"), i.e., $\{a_1, a_2, a_4, a_5, a_7, a_8, a_{11}\}$, whereas $\{a_6\}'$ is the set of semantic factors described by the attribute a_6 ("direction of flow"), i.e., $\{s_2, s_3, s_4, s_5, s_6\}$.

From X' and Y', (X", X') and (Y', Y") are derived, associating correspondingly sets of attributes with common sets of semantic factors ($\{s_1\}$, $\{a_1, a_2, a_4, a_5, a_7, a_8, a_{11}\}$), and semantic factors with common sets of attributes ($\{s_2, s_3, s_4, s_5, s_6\}$, $\{a_6\}$). Final classes, denoted by C_i, are the union of sets (X″, X′) and (Y′, Y″).

For the integrated context of the example represented in figure 4, the operations result in the following final categories:

$C_1 = (\{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9\}, \{a_8\})$ "watercourse" (SDTS)

$C_2 = (\{s_2, s_3, s_4, s_5, s_6\}, \{a_6\})$ "stream, watercourse" (WordNet)

$C_3 = (\{s_1, s_2\}, \{a_1, a_8\})$ "stream" (CYC)

$C_4 = (\{s_1, s_7, s_8\}, \{a_7\})$ "channel" (WordNet)

$C_5 = (\{s_1\}, \{a_1, a_2, a_4, a_5, a_7, a_8, a_{11}\})$ "canal"

$C_6 = (\{s_2\}, \{a_1, a_3, a_6, a_8, a_{15}\})$ "river"

$C_7 = (\{s_3\}, \{a_6, a_8, a_{10}\})$ "brook"

$C_8 = (\{s_4\}, \{a_6, a_8, a_{12}\})$ "creek"

$C_9 = (\{s_5\}, \{a_6, a_8, a_9\})$ "branch"

$C_{10} = \{(s_6\}, \{a_6, a_8, a_{16}\})$ "rivulet"

$C_{11} = (\{s_7\}, \{a_7, a_8, a_{14}\})$ "gutter"

$C_{12} = (\{s_8\}, \{a_7, a_8, a_{17}\})$ "spillway"

$C_{13} = (\{s_9\}, \{a_8, a_3\})$ "ditch"

Secondly, by examining the extents and intents of the formal concepts, it is possible to specify subconcept/superconcept relations between them, since more general formal concepts have larger extents and smaller intents than more specific ones.

The concept lattice of the running example, derived by the formal concepts associated with each other with the subconcept/superconcept relation is shown in figure 5. "Watercourse", as defined by SDTS is the most general category and includes categories "stream, watercourse" and "channel" as defined by WordNet and "stream" as defined by CYC, as well as others, such as "ditch". The integrated concept lattice reveals the differences in the conceptualization of category "stream, watercourse" by the three ontologies. Despite their differences, these homonymous categories comprise a common, equally perceived part corresponding to the categories "river" and "canal". These categories are included in the three ontologies, but are associated with different superordinate categories. The integrated concept lattice reveals the importance of those commonly conceived categories in the association of different, overlapping ontologies. Furthermore, the disclosure of the relations between different categories contributes to their interpretation in case of information exchange.

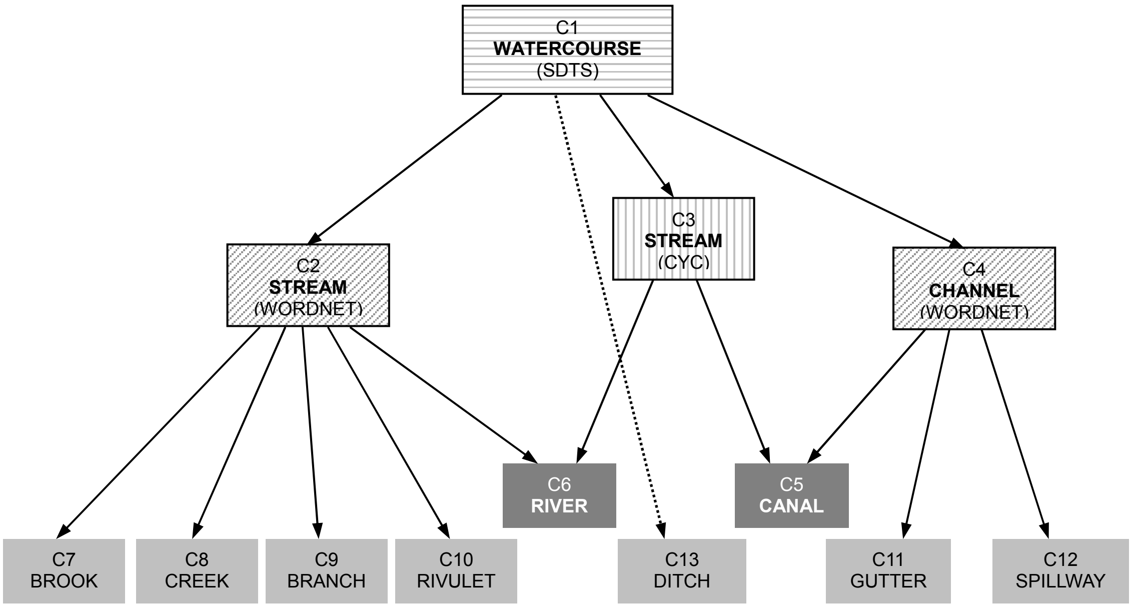

Figure 5. Integrated Concept Lattice

5. Conclusion

Semantic Factoring and Concept Lattices prove to be powerful tools in the formation and integration of geographic contexts. More specifically, the proposed methodology allows the detection of possible implicit relations between concepts, which are not pre-defined, i.e., the detection of hierarchical relationships, which were not initially obvious. Furthermore, the integrated concept lattice includes new classes derived from the fusion or division of originally overlapping ones, which increase its semantic completeness. Thus, lattices, in contrast to trees and partially ordered sets, are richer structures, which conform to fundamental characteristics of geographic categories, such as multidimensionality and the existence of overlapping relationships between them.

Moreover, the original ontologies are retained, since the resulting concept lattice reveals their interaction without altering their categories, nor prohibiting their independent use. Thus, it incorporates different, complementary conceptualizations of geographic space, each of which being suitable for some context and level of detail. The alternation of conceptualizations, whose association is preserved by the integrated concept lattice, as well as the selection of the appropriate categories according to the context and level of detail of specific applications, facilitates information exchange and reuse.

Based on these findings, our research continues towards a closer determination of essential properties for identifying geographic concepts and accounting for the empirical evidence on cognition of geographic categories and the pertaining conceptual fuzziness.

References

BIGELOW J., 2000, Semantic Properties, Monash University, wysiwyg://body.11/http:www.arts.m…il/department/bigelow/semantic.html

BOHR N., 1934, *Atomic Theory and the Description of Nature*. (Cambridge: Cambridge University Press).

CYCORP, Inc., Upper Cyc Ontology, http://www.cyc.com/

ELDER C., 1998, Essential Properties and Coinciding Objects*, Philosophy and Phenomenological Research*, **58**, 317-331.

GANTER B., and WILLE R., 1999, *Formal Concept Analysis, Mathematical Foundations* (Berlin Heidelberg: Springer-Verlag).

GUARINO N., 1998, Formal Ontology and Information Systems. In *Formal Ontology in Information Systems*, edited by N. Guarino (Amsterdam: IOS Press), pp. 3-15.

GUARINO N., and WELTY C., 2000, A Formal Ontology of Properties (Preliminary Version). In *Proceedings of the ECAI-2000 Workshop on Applications of Ontologies and Problem-Solving Methods*, edited by R.Benjamins, A.Gomez-Perez, N.Guarino, and M.Uschold, http://www.ladseb.pd.cnr.it/infor/ontology/ontology.html

GUARINO N., and WELTY C., 2000, Ontological Analysis of Taxonomic Relationships. In *Proceedings of ER 2000: The International Conference on Conceptual Modeling* (Springer-Verlag LNCS), http://www.ladseb.pd.cnr.it/infor/ontology/ontology.html

MILLER G.A., BECKWITH R., FELLBAUM C., GROSS D., and MILLER K., 1993, Introduction to WordNet: An On-line Lexical Database, ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.pdf

PENCO C., 1999. Objective and Cognitive Context. In *Proceedings of the Second International and Interdisciplinary Conference, CONTEXT'99*: *Modeling and Using Context*, edited by P. Bouquet, L. Serafini, P. Brézillon, M. Benerecetti, F. Castellani (Berlin Heidelberg: Springer-Verlag), pp. 270-283.

SMITH B., and MARK D.M., 1998, Ontology and Geographic Kinds. In *Proceedings of the International Symposium of Spatial Data Handling (SDH '98)*, Vancouver, Canada.

SOWA, J. F., 2000, *Knowledge Representation: Logical, Philosophical and Computational Foundations*, (USA: Brooks/Cole).

STANFORD ENCYCLOPEDIA OF PHILOSOPHY, edited by E. N. Zalta, The Metaphysics Research Lab at the Center for the Study of Language and Information, Stanford University, Stanford, CA, http://plato.stanford.edu/

UNITED STATES GEOLOGICAL SURVEY, Spatial Data Transfer Standard, http://mcmcweb.er.usgs.gov/sdts/

WILLE, R., 1992, Concept Lattices and Conceptual Knowledge Systems. *Computers and Mathematics with Applications*, **23-6-9**, 493-515.

WORDNET - a Lexical Database for English, Cognitive Science Laboratory, Princeton University, http://www.cogsci.princeton.edu/~wn/