A Foundational Framework for Structuring Geographical Categories

Boyan Brodaric, Geological Survey of Canada

Introduction

Geographical categories are often represented in machine-readable ontologies to aid information interoperability, retrieval, fusion, analysis, workflow execution and image classification. Within these activities categories are routinely merged, split, and compared. Such operations are particularly dependent on the representation choices made at the syntactic, structural and semantic levels: e.g. category similarity can be determined according to syntactic resemblances of names, structural metrics such as distance in a graph, or semantic metrics such as possessing common relations. To enable these operations categories are regularly transformed to a common representation framework, suggesting a need for optimal strategies at each representation level. This paper builds on advances in foundational ontologies and geospatial semantics to develop a representation framework for geographical categories at the semantic level, using the *Lake* category in a running example.

A geographical category is considered here to be an abstraction used to classify geographical individuals, such as *Lake* classifying *Lake Ontario*, an instance is a classified individual, and instantiation refers to the relation between a category and one of its instances. Category structuring refers to the central aspects of the meaning of a category such as the properties involved in a cognitive prototype, or the properties involved in necessary or sufficient conditions. Understanding the structuring of a category is then a significant precondition for advanced category processing, but existing schematic and semantic approaches are limited by representation frameworks with narrow ontologic range and thus restricted category structuring:

- Structural approaches: operate on category constructs such as graphs, schemas, or flows, and are relatively agnostic about the inherent semantics. For example, while generalization relations (*Lake isa Physical-Object*) are often differentiated, categories structured around qualities (*Deep-Lake*) or processes (*Volcanic-Lake*) are often not well distinguished and are consequently processed similarly.

- Semantic approaches: operate on category constructs oriented around a limited suite of ontological distinctions. For example, cognitive semantic approaches focus on categories structured around qualities (*Deep-Lake*), while other cognitive approaches emphasize categories structured around goals or roles (*Lake-used-for-Fishing*). Although often overlooked, more complex categories could be structured around geographical relations, such as *Lake-connected-to-Rivers-crossing-Cities* which is perhaps useful in pollution analysis.

Such narrow ontologic range can promote inconsistencies and inabilities in the machine processing of categories. The ontologic framework developed in this paper addresses this by delineating several types of categories and their organizing primitives, with an aim to increase category generality and granularity in processing mechanisms. For example, under this framework a category structured around qualities could be more similar to other categories structured around qualities than to those organized around processes or relations—intuitively, *Deep-Lake* is more ontologically similar to *Shallow-Lake* than to *Lake-used-for-Fishing*, even if all are schematically equal because they share a common parent node (*Lake*) in some graph.

Approach

The proposed framework is organized as a matrix in which the columns build on top-level primitives from the DOLCE foundational ontology (Masolo *et. al*, 2003), and the rows denote different levels of category organization as identified by the OntoClean approach (Guarino & Welty, 2002). This work refines these existing distinctions and adds to them, resulting in an expanded matrix in which each cell denotes a category type with unique structuring criteria, as shown in Table 1. For immediate purposes, only categories about geographical objects are considered (roads, lakes, mountains, rock bodies, countries), using *Lake* as an example, leaving open the possibility that the matrix can be applied to other entities such as processes and qualities. Also, although the rows of the matrix might be connected by specific relations, these are not considered in this paper.

The columns of the matrix, and its top row, consist of the top-most DOLCE primitives: *Endurant* (*Physical-Object*, *Feature*), *Perdurant* (*Process*, *Event*, *State*), *Quality* (e.g. *Color*), *Abstract* (e.g. *Blue*), as well the foundational relations that bind them together (e.g. a quality is *inherent-in* an endurant, an endurant is a *participate-in* a perdurant). An individual can then be structured according to its involvement in these foundational relations. For example, *Lake Ontario* might be classified according to inherent quality values (*Deep-Lake*) or the processes in which it participates (*Lake-used-for-Fishing*). The foundational relations can also be specialized to achieve greater ontologic granularity: e.g. the *participates* relation between an endurant and perdurant can be specialized to account for its geo-pragmatics, to indicate how a geographical object might be used by an agent, produced by a process or event, or how it caused a process, event, or state (Brodaric, 2007). This helps explain goal, role, and ad-hoc categories, such as *Lake-used-for-Fishing*, which can then be seen as organized around the use relation, including past, current and future use. The matrix also reifies *Relation* as a column, for categories organized around a relation to a relation such as *Lakes-connected-to-Rivers-crossing-Cities*.

The top row of the matrix, consisting of the top-most DOLCE primitives, corresponds to OntoClean non-sortals. These are the most general categories, which can typically be instantiated in any domain. The second row specializes the foundational primitives into the most general categories that are instantiated in a single domain, such as *Water-Body* (*Physical-Object constituted-by Water*) or *Water-Container* (*Feature hosts Water-Body*). These correspond to OntoClean sortal types in that they are rigid and supply identity (Welty & Guarino, 2001): an instance of them is always such, in that a water body instance never stops being a water body, even if further classified as *Lake*, *Pond*, or *River*, and water body instances possess common identity criteria such as unique geospatial location. Each domain individual must also instantiate one of these categories. The notion of being schematic is further added here as a constraint to these categories: such categories can define a quality space for structuring more specific categories, e.g. water bodies possess qualities such as *Size*, *Shape*, and *Depth*, but the *Water-Body* category is not structured around specific quality values such as *Size* = *Large*.

The third row contains domain types that are classified, such as *Lake* (*Water-Body* with *Size* = *Large* and *constituted-by Freshwater).* Formally these are OntoClean non-type sortals that carry identity without supplying it, are not necessarily rigid, and can be structured around quality values: *Lake* and *Ocean* are identified in the same way inheriting identity criteria from *Water-Body*, water body instances produced by volcanoes are always such, but instances of *Lake-used-*

for-Fishing might not always be such due to changes in regulation, transportation, or fish stock, and *Deep-Lake* is structured around a depth value (e.g. *Depth = Deep* or *Depth > 240m*). The fourth row adds to the previous OntoClean distinctions the notion of a situated category, which is a category structured around a relation to one or more individuals where the relation to the individual is either direct (*Lake-near-Toronto*, *Lake-produced-by-the-Keli-Mutu-Volcano*) or indirect via another situated category (*Lake-in-a-Western-Country*). This dependence on individuals contextualizes situated categories, tying them to certain geospatial or historical locales, and broadens previous definitions for situated categories focussed mainly on perdurant individuals (Brodaric, 2007; Brodaric & Gahegan, 2007). Formally, situated categories are not necessarily rigid, e.g. instances of categories structured around perdurants such as *Lake-produced-by-the-Keli-Mutu-Volcano* or *Lake-that-caused-Flood-X* have their meaning tied to a historical occurrence and are hence always such, but instances of *Lake-near-Toronto* might not be such if the geographical area occupied by Toronto contracts or the proximity criteria change. Situated categories might also help explain some prototypical effects, e.g. instances with a common process might exhibit diverse quality values. Finally, the fifth row consists of individuals (*Lake Ontario*).

Table 1: a framework for structuring the *Lake* category.

Foundational	**Perdurant: Process, Event, State**			**Endurant: Physical-Object**	**Quality**	**Abstract: Region**	**Relation**
Schematic	Fishing Forest Fire Fighting	Volcanism Tectonic Rifting	Flooding	**Water Body**	max-depth color size	max-depth-value color-value size-value	crossing
Classification	***Use*** Lake-used-for-fishing Lake-used-for-fighting-forest-fires	***Origin*** Lake-produced-by-volcanism (Volcanic Lake) Lake-produced-by-rifting (Rift Lake)	***Effect*** Lake-that-caused-flooding	**Lake** Lake-near-city Lake-part-of-transport-route	Lake-with-max-depth Lake-with-color	size = large Deep-lake (max-depth = deep, > 240m) Blue-lake (color = blue)	Lake-connected-to-rivers-crossing-cities
Situated	Lake-used-for-fishing-competition-X Lake-used-for-fighting-forest-fire-Y	Lake-produced-by-volcanism-X (Keli Mutu Lake) Lake-produced-by-rifting-Y (African Rift Valley Lake)	Lake-that-caused-flood-X	Lake-near-Toronto Lake-part-of-the-St.-Lawrence-Seaway	Lake-with-Lake-Ontario-max-depth Lake-with-Lake-Ontario-color	Lake-with-Lake-Ontario-blue-color	Lake-connected –to-rivers-crossing-Montreal
Individual	Lake Ontario, Lake Erie,…	Tiwu Ata Polo (Indonesia),… Lake Tanganyika (Rift Valley),..	Cerknica Lake (Slovenia) ,...	Lake Ontario,...	Lake Ontario,...	Lake Ontario,…	Lake Ontario, Lake Temiscaming

Conclusions

This paper suggests category processing would benefit from an enhanced suite of category types whose structuring basis can be clarified with ontologic primitives, in this case from DOLCE and OntoClean, supplemented by the situated category type. As the main point here is to advance category processing through enhanced ontologic distinctions, alternative or supplemental ontologic approaches might also be used to attain this overall goal. In particular, the incorporation of geospatial primitives into such frameworks would be a promising extension. Also note that while this framework at first appears to be oriented towards categories structured around logic-based definitions, it can also apply to categories structured around instances that are variably typical of the category, insofar as the latter categories also make ontologic distinctions about the properties of the category that are involved in determining instance typicality.

References

Brodaric, B. (2007). Geo-pragmatics for the Geospatial Semantic Web. Transactions in GIS, 11(3):453-477.

Brodaric, B. Gahegan, M. (2007). Experiments to examine the situated nature of geoscientific concepts. Spatial Cognition and Computation, 7(1):61-95.

Welty, C., Guarino, N., (2001). Supporting ontological analysis of taxonomic relations. Data and Knowledge Engineering, 39(2001): 51-74.

Guarino, N., Welty, C. (2002). Evaluating ontological decisions with OntoClean. Communications of the ACM, 45(2):61-65.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A. (2003). WonderWeb DeliverableD18, Ontology Library (final). Laboratory For Applied Ontology, http://www.loa-cnr.it/Papers/D18.pdf.