REPRESENTING GEO-SCIENTIFIC DOMAIN CONCEPTS

Boyan Brodaric

Penn State Geography *and* Geological Survey of Canada
brodaric@NRCan.gc.ca

1. Introduction

The geo-sciences, including geology, ecology, soil science, etc., exemplify domains in which discovering the identity and structure of geographic occurrences and related concepts is a primary goal. In such domains both occurrences and concepts are interactively discovered and evolved: they (1) originate from empirical evidence, (2) evolve as evidence and theory accumulate and change, and (3) are characterized by degrees of uncertainty and granularity. This suggests representation of such concepts includes not only a concept's intension, in terms of essential and non-essential properties (e.g. attributes, relations, functions, constraints), but also includes representation of intension development in terms of affecting evidence, processes, inferences and qualifying uncertainties; conversely, it also suggests occurrences classified as belonging to the concept—its extension—be augmented with similar factors. In short, supplementing property-based occurrence and concept representation with explanatory intension and extension development will enhance the meaning of occurrences and concepts. Such representation has scientific merit as it enables re-examination and testing to support general scientific inquiry.

2. The Case for Intension and Extension Development

Intension/extension development generally plays a minor role in ontology structure: a concept's intension typically utilizes a classification function (concept specific or generic), to place occurrences into the concept's extension, and has an instantiation (instance-of) relation between intension and extension. This approach is insufficient for (1) concept origination, in that it does not distinguish between occurrences in the extension and those for intension development; (2) concept evolution, in that it does not account for the factors causing versioning; and (3) concept uncertainty, as the classification function and instantiation relation are unqualified. Support for a more sophisticated view of at least concept origination is provided in several research fields:

- **Scientific Method**: cognitive and philosophic studies of scientific method distinguish the mechanisms and contents in knowledge discovery (i.e. intension development) from those in knowledge evaluation (i.e. expansion to a larger population—extension development) (e.g. Zimmerman, 2002; Sowa, 2001).
- **Machine Learning**: likewise, the training set used to build an intension is typically distinguished from the larger population of occurrences that are classified using the intension.
- **Cognitive Science**: conceptual functions for managing intensions and extensions are being viewed as increasingly distinct (Laurence & Margolis, 2003), implying distinct contents: intensions can arise from prototypes, theories and situations, while extensions may derive from related simulations (Barsalou, 2002), definitions, or other means. The functions may be grounded in cognitive hardware and built up from experience (Lakoff, 1987).

This position argues that augmenting the representation of occurrences and concepts with explanatory intension and extension development will enhance "*material*" ontology, i.e. ontologies based on conceptualizations of domains rather than *a prior* entities (Smith, 1998), as well as "*prototype-based*" ontology, i.e. ontologies derived from evidence of varying typicality (Sowa, 2000, 495-6). It assumes the identity and state of occurrences and concepts can originate and evolve, and that intension and extension content, and related functions, can be distinct.

3. Characterizing Intension and Extension Development for Concept Origination

In previous work, the set of occurrences contributing to intension development is unconventionally denoted as the concept's *category*, and the set of occurrences that comprise its extension is conventionally denoted as its *class* (Brodaric & Gahegan, 2002). This distinction does not imply that a concept's *category* is fully a subset of its *class*: e.g. the category may possess both positive and negative examples (a delineation made in rough sets), or variably typical examples (as per fuzzy sets). Furthermore, scientific occurrences and concepts are not always derived from examples but may arise from occurrences that represent properties, e.g. many physical occurrences, such as electrons, are indirectly observable via their properties and possess a theoretical extension; for many geographic occurrences, such as geologic formations, some parts are directly observable and so the extension is partially theoretical. The intension for such entities can be seen as a generalization over property occurrences. Figure 1 depicts this interaction qualitatively as mappings between a set of target concepts $\boldsymbol{C}$, a set of concepts $\boldsymbol{P}$ serving the role of properties, sets of occurrences $\boldsymbol{P_c}$ and $\boldsymbol{P_a}$ for the respective classes and categories of $\boldsymbol{P}$, and sets of occurrences $\boldsymbol{C_c}$ and $\boldsymbol{C_a}$ for the respective classes and categories of $\boldsymbol{C}$:

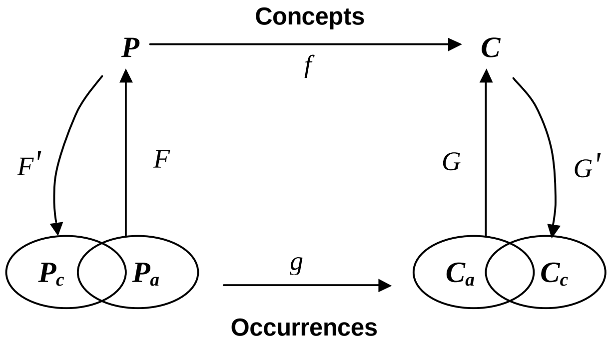

Figure 1: mappings for concepts and occurrences; note the intension development mappings commute.

Viewing concepts from the perspective of intension development suggests a qualitative classification for domain concepts, as shown in Table 1. If $\boldsymbol{C_a}$ and $\boldsymbol{P_a}$ are understood from a scientific realist perspective as sets of observables, subsets of which are cognitive perceivables, and if ø denotes the empty set, then: *abstract* concepts are inferred from property concepts, *theoretical* concepts from observable properties, *prototypical* concepts originate from observable and variously prototypical examples, and *situated* concepts, often geographical, can originate from partial examples and their observable properties. This construction is geared to cognitive conceptualizations, e.g. those thought to underlie domain ontology, particularly geo-scientific, and makes no claims for upper-level ontology. The various mappings are listed in Table 2: *attribution* establishes the identify and content of both concepts and occurrences from properties, *clustering* establishes concept identity and content from examples, *instantiation* assigns a concept's properties to an occurrence, which may involve occurrence creation, and its inverse *classification* places an occurrence into a concept's extension via its intension.

Table 1: domain concept types from intension development

$\boldsymbol{C_a}$	$\boldsymbol{P_a}$	concept type	intension development	extension development
Ø	Ø	abstract	f	G'
Ø	¬Ø	theoretical	$f\ F$	G'
¬Ø	Ø	prototypical	G	G'^{-1}
¬Ø	¬Ø	situated	$g\ G$	g

Table 2: mappings

mapping	description
$f,\ g$	attribution
$F\ ,\ G$	clustering
F', G'	instantiation
F'^{-1}, G'^{-1}	classification

4. Applications

Augmenting concepts with intension and extension development seems particularly applicable to ontologies that are empirically or possibly socially inferred. Consequently, the approach is currently being adapted to concept management in geo-scientific databases and their federations (Brodaric & Gahegan, 2002). In addition, the approach has potential to contribute to the object/field debate inasmuch it may provide an ontological characterization of object development from field properties; it may also provide supplementary characterization of inference in geo-scientific concept formation (e.g. F,G as logical induction). Aspects of these possibilities are currently being explored. Other outstanding issues with respect to intension and extension development include a characterization of change and uncertainty for both the state and identity of occurrences and concepts, as well as grounding concepts and occurrences in atomic properties such as those derived from cognitive research (e.g. Probst, *et al.*, 2003).

5. Conclusions

The paper argues that the structure of domain ontologies, those based on human conceptualizations, can be augmented to include factors contributing to intension and extension development. The augmented structure is mainly motivated from cognitive research into conceptual structure and function, as well as scientific method, and is algebraically described as mappings between sets of concepts and occurrences. An unexpected result of this description is a typology of domain concepts that initially seems aligned with cognitive science research into conceptual structure and function. Test applications involve management of geo-scientific concepts.

6. References

Barsalou, L.W. (2002). Being there conceptually: Simulating categories in preparation for situated action. In: Stein, N.L., Bauer, P.J., Rabinowitz, M. (eds.), Representation, memory, and development: Essays in honor of Jean Mandler. Erlbaum, pp. 1-16.

Brodaric, B., Gahegan, M. (2002). Distinguishing Instances and Evidence of Geographical Concepts for Geospatial Database Design. In: Geographic Information Science-2nd Int'l Conference, GIScience 2002, M. Egenhofer and D. Mark (eds.), LNCS 2478, Springer, Berlin, 22-37.

Lakoff, G. (1987). Women, Fire and Dangerous Things. University of Chicago Press.

Luarence, S., Margolis, E. (2003). Concepts. In: The Blackwell Guide to the Philosophy of Mind. Warfield, T., Stich, S. (eds.). Blackwell.

Probst, F., Raubal, M., Kuhn, W. (2003). An Image-Schema-Based Upper-Level Ontology for the Geospatial Domain. http://musil.uni-muenster.de/documents/research_paper_Probst_Raubal_Kuhn.pdf.

Smith, B. (1998). An introduction to Ontology. In: Peuquet, D., Smith, B., Brogaard, B. (Eds.), The Ontology of Fields. National Center for Geographic Information and Analysis, Santa Barbara, CA, 10-14.

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

Sowa, J.F. (2001). Signs, Processes and Language Games, Foundations for Ontology. http://www.jfsowa.com/pubs/signproc.htm.

Zimmerman, C. (2000). The development of scientific reasoning skills. Developmental Review, 20: 99-149.