[FC-03-001] Foundational Ontologies

Abstract

Foundational ontologies are tools for knowledge representation at a general level. They introduce hierarchies of concepts and relationships between these concepts. Foundational ontologies provide a common base for building more specific domain ontologies, which describe knowledge in expert domains. Foundational ontologies formally define categories and relations, improving the consistency of domain ontologies and facilitating their integration and knowledge exchange. Foundational ontologies structure knowledge in a hierarchy where at a highest level are defined universal entities and individuals, which instantiate the previous ones, and continuants and occurents, depending on their persistence in time. They also propose a set of abstract relationships from which more specific ones can be derived to relate concepts. Different foundational ontologies have been developed, proposing different knowledge organisations. In geospatial science, an important aspect of ontologies is the representation of spatial regions and spatial relationships. Regions can be built as point-sets or from atomic regions that correspond to elementary geospatial entities. The former approach facilitates quantitative reasoning in geometrical spaces while the latter is more appropriate for qualitative reasoning and the definition of high-level relationships. However, the representation must take into account the perception of boundaries and the possible vagueness of the concepts.

Keywords: knowledge, ontology, vagueness

Author & citation

Guilbert, E. (2024). Foundational Ontologies. The Geographic Information Science & Technology Body of Knowledge (2024 Version), John P. Wilson (Ed.) DOI: 10.22224/gistbok/2024.1.17.

Explanation

1. Foundational Ontologies
2. Content of Foundational Ontologies
3. Existing Ontologies
4. Geospatial Ontologies

1. Foundational Ontologies

In information science, the real world is described through concepts and categories that represent objects along with their processes and their relationships. As defined by Studer et al. (1998), an ontology is a “formal, explicit specification of a shared conceptualisation”. It provides a representation of knowledge shared within a given domain or discipline. In comparison to conceptual models used in application development, ontologies provide tools to support such a knowledge representation but are independent from any application. Hence, they provide a common knowledge base on which experts agree upon and that can be shared.

Ontologies can be distinguished according to their level of generality or abstraction. Ontologies that present concepts specific to a domain or a task are simply called domain or task ontologies. They gather the collective knowledge that is shared by experts of the domain. They are also used for data integration and analysis. In geospatial science, ontologies such as landform ontologies or natural disaster ontologies, are considered as domain ontologies, modelling domain classes such as hill, landslide or earthquake.

At the highest level of generality are foundational ontologies, also known as top-level or upper-level ontologies. They are ontologies built independently from any application domain. As such, they provide a higher level of categorisation with distinctions among the entities of the world (physical objects, quantities, events, etc.). They include central notions that can be used in lower-level ontologies.

Distinction between levels of generality is not strict. Some domain ontologies are more specific than others. For example, an application ontology on urban planning formalizes both domain knowledge on the urban environment and task knowledge on planning. Some ontologies can contain more general concepts and have a broader scope. Such ontologies are referred to as mid-level or core ontologies. An example is the Semantic Web for Earth and Environment Technology Ontologies (SWEET, Raskin et al. 2004) which is a set of ontologies for earth system science. It defines top-level concepts such as physical properties and space and includes domain-specific ontologies common to earth science disciplines.

Upper-level ontologies provide a foundation for the development of lower-level domain ontologies. It is still possible to develop domain ontologies without using foundational ontologies and some developers still consider there is no benefit from it. Foundational ontologies can be seen as too abstract and too comprehensive when considering a specific domain, demanding too much effort to understand. As an example, SWEET is not derived from a top-level ontology but includes concepts from which lower domain ontologies are built. For that purpose, SWEET is formed of a base ontology that is extended by a collection of domain ontologies. On the contrary, GFO-Bio is a core ontology based on the foundational ontology General Formal Ontology (Herre et al., 2006).

Nonetheless, foundational ontologies present several advantages and as mentioned by Keet (2020). They provide a controlled vocabulary, including consistent categories and relations that have been formally defined. As such, they provide a set of basic categories and relations from which lower ontologies can be built. This improves the quality of the ontologies by waiving ambiguity. Concepts derived from the same categories are easier to integrate since they share common properties and relationships. Consequently, integration of domain ontologies built on the same foundational ontology is facilitated.

Ontologies, including foundational ontologies, are mainly found on web portals and provide a structure for linking data for the Semantic Web. By providing a common knowledge base reusable for new ontologies, foundational ontologies help achieving FAIR principles (findability, accessibility, interoperability, reusability).

2. Content of Foundational Ontologies

Categories in a foundational ontology define different types of entities based on a set of

core categories that distinguish different types or entities. Hence, categories are mainly organised in a hierarchy but are also related together with properties, providing more complex relationships than in a taxonomy. Several important distinctions are commonly found in foundational ontologies. The first is the distinction between universals and particulars. Particulars (or individuals) are non-repeatable. Each particular is unique and has its own characteristics. Conversely, universals are repeatable and define similarities between particulars. For example, Paris and London are two particulars of the universal City. Usually, foundational ontologies focus mainly on the definition of universals and particulars are left to domain ontologies, at a lower level. However, some ontologies, such as the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE, Masolo et al., 2003), define particulars from their properties only, without referring to universals.

A second distinction is made between continuants (or endurants) and occurents (or perdurants). A continuant is an entity that is wholly present at a given time. For example, a book is a continuant because, at any time of its existence, the whole book can be observed. An occurrent, by contrast, is defined over a period of time and cannot be fully observed at a time instant. Occurents are usually processes or actions identified for a period of time. Hence, a snapshot of a person reading demonstrates only a part of the reading process.

A third distinction is between physical (or concrete) and non-physical (or abstract) entities. New concepts are then obtained by combining more generic concepts. For example, a process can be defined as a physical, occurrent entity.

Categories in ontologies are described by properties and are related to each other by relationships. However, at an upper level, relationships rely on axioms and mathematical logic and must be generic enough to apply across different domains. Similarly, properties remain generic and are not assigned a specific data type. In order to remain generic, properties can be instead defined as classes that are related to categories they describe. For example, the colour of an object (a continuant) is defined by a relationship between the continuant class and the colour class, which itself can be a subclass of the quality class.

Examples of relationships are parthood, participation (e.g. when an individual takes part in a process), constitution, (e.g. a vase is made of clay), and dependency (when the existence of an object depends on the existence of another).

Parthood is one of the most essential relationships in an ontology. It can be defined as a primitive relationship characterised by three properties.

- Reflexivity (an entity is part of itself);
- Antisymmetry, two different entities cannot be part of each other;
- Transitivity: if x is part of y and y is part of z then x is part of z.

However, parthood remains a generic relationship and must be refined. For example, if the hand is part of a person, and a football player, who is a person, is a part of a football team, then, by transitivity, the hand is part of the team. The former relation is a structural parthood while the latter is a membership. Hence, primitive relations in foundational ontologies are specialised in a taxonomy. An example of such specialisation is the spatial parthood, which can be further specialised for 3D-objects and 2D-objects.

3. Existing Ontologies

Many projects to create foundational ontologies began early in the 21st century. Some of the most prominent are Basic Formal Ontology (BFO) (Grenon, 2003), Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) (Masolo et al., 2003) and the General Formal Ontology (GFO) (Herre et al., 2006). As top-level ontologies, BFO, DOLCE and GFO are libraries providing general categories organised in a hierarchy. Their differences lay mainly in how categories are organised. BFO and GFO were developed as upper ontologies for biomedical science while DOLCE tries to design an ontology of natural language and common-sense categories.

BFO is a small and general ontology, with a limited number of concepts. It separates universals and individuals in a way that individuals are instantiated by universals (i.e. with an instance of relationship), which cannot be instantiated. BFO starts by separating them into continuants and occurents before introducing dependency and processes (Figure 1, left).

GFO firstly separates entities between sets and items. considers three categories: universals, concepts and symbols. Concepts represent meanings in someone's mind and can capture universals. Symbols are signs or texts instantiated that can linguistically express a concept. Distinction between continuants and occurents is done at a deeper level in the ontology.

DOLCE is an "ontology of particulars" that does not include universals. Particulars are divided between endurants and perdurants but also qualities and abstracts. Characteristics of particulars are then defined by relating them to their qualities.

Foundational ontologies mentioned above are openly available in OWL (Web Ontology Language). They can be viewed and edited in ontology editors such as Protégé. Thus, they can directly be used to develop new domain or application ontologies. The most common approach to build domain ontologies when working with foundational ontologies is the top-down approach. It consists first of choosing the most appropriate foundational ontology. Upper ontologies are usually organised in modules so that only modules relevant to the domain can be considered. Then, categories are linked with the concepts in the domain knowledge.

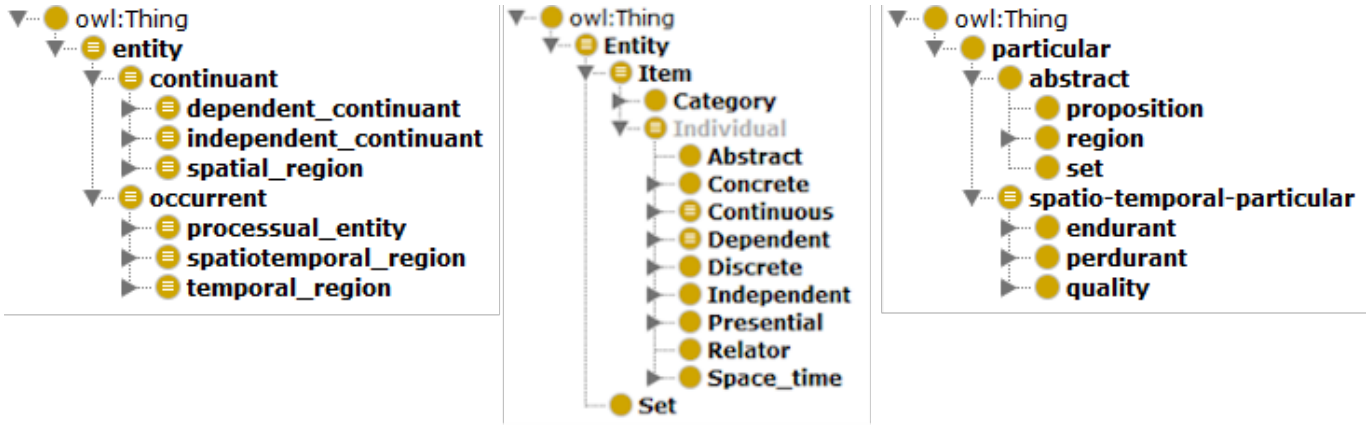


Figure 1. Top categories in BFO (left), GFO (centre) and Dolce (right) viewed in Protégé. Source: author.

4. Geospatial Ontologies

Geospatial ontologies can be addressed from both philosophical and computer science perspectives. The former defines theories in order to describe geospatial entities, processes and their relations as constituents of the geographical reality. The latter is looking at formalising and standardising knowledge for information exchange but also for digital representation and processing. As discussed by Mark et al. (2004), both perspectives are closely interconnected. Claramunt (2020) mentions that geospatial ontologies should provide a taxonomy and a formal vocabulary that can be computerised at the software engineering level.

Space and time are basic categories in foundational ontologies. However, several issues and requirements arise when choosing a foundational ontology in order to address both perspectives. First, geospatial entities can be represented in different continuums. Space and time can be modelled into a single 4D space-time continuum or as a space and a time continuum. The 4D continuum is the representation chosen for perdurant entities while endurant entities are fully defined in a space continuum. Space itself can be represented in different ways depending on its relationship with real-world entities. In BFO and GFO, spatial regions are associated with entities directly through a spatial relation. In DOLCE, the location of a particular is a spatial quality whose value is a region in the geometrical space.

Spatial regions can be made up of two types of elements. Space can be composed of abstract entities that are points. A region defining the shape and position of an entity is thus defined by a set of points. On the other hand, space can be partitioned in atomic regions whose location is in accordance with basic elements that describe geographical entities. As such, atomic regions provide higher-level relations for spatial qualitative reasoning. This is also consistent with the fact that people think mostly about space in terms of regions rather than points and lines (Hobbs et al., 2006). However, in point-set theory, the absolute space can be defined as a Euclidian space and rely on mathematical definitions, bringing computational efficiency.

Spatial entities are bounded in space. Thus, regions where spatial entities are located have boundaries. In point-set theory, boundary definition relies on algebraic topology. The closure of a region A is the smallest closed set that contains A. The interior of A is defined by the largest open set it contains, and its boundary is the difference between its closure and its interior. Thus, if two entities are in contact, they share a common boundary. The boundary of a region is of a lower dimension than the region: the boundary of a surface is a polygonal line, the boundary of a line is the pair of points at its extremities and a point has no boundary (Figure 2).

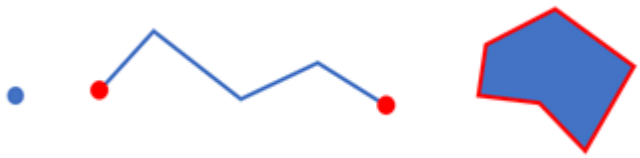


Figure 2. Interior (in blue) and boundary in red of a point, line and polygon. Source: author.

In such consideration, the boundary is a crisp geometry. However, boundary definition can lead to some philosophical issues (Hahmann and Grüninger, 2012). If two regions are

closed, their boundary belongs to both regions. Thus, the boundary belongs to the entities located in these regions. One may question if the properties defined along the boundary belong to one entity or the other. Several solutions are proposed for this. The boundary can be considered as belonging to both regions or can be seen as an object demarcating both regions. For example, the boundary between two countries can be seen as belonging to neither of them.

Depending on the entities, boundaries can be of two kinds. For a physical object such as a book or a table, the boundary clearly marks a separation with other objects. In many geospatial objects, boundaries do not correspond to a natural division, as for an island, but rather to a human decision. This is the case for example for country boundaries. Smith (1996) refers to the first kind as bona fide boundaries and to the second as fiat boundaries. This distinction is reflected in BFO and DOLCE.

GFO instead distinguishes material and space boundaries. An object can have both a material boundary and a space boundary (Baumann et al. 2016). The material boundary demarcates the object from its environment. If this boundary is a natural discontinuity, it is equivalent to a bona fide boundary. Material boundaries occupy a space boundary. While two space boundaries overlap where objects are in contact, the material boundaries are still considered distinct.

Furthermore, in many cases, the boundary is not clearly defined by a proper line. This is specifically the case with landforms such as mountains and hills. The foot of a hill or mountain often does not show a sharp delineation from surrounding landforms. Two main approaches have been designed to account for vagueness in boundary definition. The first one relies on fuzzy sets (Burrough, 1996). The second one relies on rough sets, with the egg-yolk model (Cohn and Gotts, 1996): a region is divided into a yolk where the entity is definitely in and a white where it possibly is. While the egg-yolk model is used for qualitative spatial reasoning, fuzzy sets, where membership functions are defined, allow for computation and a quantification of the fuzziness (Burrough, 1996).

Indeed, vagueness does not come from nature but from human-defined concepts. They require entities such as mountains and hills to be named and further delineated by a boundary (Kavouras and Kokla, 2008, chapter 2). The distinction between a hill and a mountain itself is arbitrary and depends on linguistic and cultural variations. Furthermore, location of most geospatial entities is thus vague but computational data models often require crisp delineations. Indeed, this gap between the high-level, qualitative representation and a quantitative, computationally tractable representation remains a fundamental challenge for the development of geospatial ontologies (Claramunt, 2020).

References

Burrough, P. A. (1996). Natural Objects with Indeterminate Boundaries. In Burrough, P. A., Geographic Objects with Indeterminate Boundaries, 1st Edition. CRC Press.

Claramunt, C. (2021). Ontologies for geospatial information: progress and challenges ahead. Journal of Spatial Information Science, 20, 35–41.

Cohn, A.G., Bennett, B., Gooday, J.M., & Gotts, N.M. (1997). Qualitative Spatial Representation and Reasoning with the Region Connection Calculus. GeoInformatica

1(275–316).

Grenon, P. (2003). The Formal Ontology of Spatio-Temporal Reality and its Formalization. In Foundations and Applications of Spatio-Temporal Reasoning (FATSR). AAAI Spring Symposium, Technical Report SS-03-03. (pp. 27–34). AAAI Press.

Hahmann, T. & Grüniger, M. (2012). Region-Based Theories of Space: Mereotopology and Beyond. In Hazarika, S. M. (ed) Qualitative Spatio-Temporal Representation and Reasoning: Trends and Future Directions (pp. 1-62). IGI Publishing.

Herre, H., Heller, B., Burek, P., Hoehndorf, R., Loebe, F., & Michalek, H. (2006). General Formal Ontology (GFO): A Foundational Ontology Integrating Objects and Processes. Part I: Basic Principles; Research Group Ontologies in Medicine (Onto-Med): Leipzig, Germany.

Hobbs, J., Blythe, J., Chalupsky, H., & Russ, T. A. (2006). A Survey of Geospatial Resources, Representation and Reasoning. Technical Report, Information Sciences Institute, University of Southern California.

Kavouras, M. & Kokla, M. (2008). Theories of Geographic Concepts: Ontological Approaches to Semantic Integration. CRC Press.

Keet, C. M. (2020). An Introduction to Ontology Engineering. v1.5, February 2020. 306p.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N. & Oltramari, A. (2003). WonderWeb Deliverable D18 Ontology Library (final); Technical report No. 2001-33052; WonderWeb Ontology Infrastructure for the Semantic Web.

Raskin, R., Pan, M., & Mattmann, C. (2004, June). Enabling Semantic Interoperability for Earth Science Data. In 4th NASA Earth Science Technology Conference.

Smith, B. (1996). Mereotopology: A Theory of Parts and Boundaries. Data and Knowledge Engineering, 20(287–303)

Studer, R., Benjamins, R., and Fensel, D. (1998). Knowledge Engineering: Principles and Methods. Data & Knowledge Engineering, 25(1-2):161198.