

From top-level to domain ontologies: Ecosystem classifications as a case study

Thomas Bittner

Department of Philosophy, Department of Geography
New York State Center of Excellence in Bioinformatics and Life Sciences
National Center for Geographic Information and Analysis
State University of New York at Buffalo

Abstract. This paper shows how to use a top-level ontology to create robust and logically coherent domain ontology in a way that facilitates computational implementation and interoperability. It uses a domain ontology of ecosystem classification and delineation outlined informally Bailey’s paper on ‘Delineation of Ecoregions’ as a running example. Bailey’s (from an ontological perspective) rather imprecise and ambiguous definitions are made more logically rigorous and precise by (a) restating the informal definitions formally using the top-level terms whose semantics was specified rigorously in a logic-based top-level ontology and (b) by enforcing the clear distinction of types of relations as specified at the top-level and specific relations of a given type as they occur in the ecosystem domain. In this way it becomes possible to formally distinguish a number of relations which logical interrelations are important but which have been confused and been taken to be a single relation before.

1 Introduction

Ontologies are tools for specifying the semantics of terminology systems in a well defined and unambiguous manner [1]. *Domain ontologies* are ontologies that provide the semantics for the terminology used to describe phenomena in a specific discipline or a specific domain. In this paper the domain of ecosystem classification and delineation is used as an example. Other domains include hydrology and environmental science, as well as medicine, biology, and politics.

In contrast to domain ontologies, *top-level ontologies* specify the semantics for very general terms (called here top-level terms) which play important foundational roles in the terminology used in nearly every domain and discipline. Top-level terms that are relevant to this paper are listed in Table 1.

Building a domain ontology is an expensive and complex process [3]. Research has shown that robust domain ontologies must be [1,4]: (i) based on a well designed top-level ontology; (ii) developed rigorously using formal logic. This means that the semantics of the domain vocabulary is specified within a logic-based framework using top-level terms with an already well established semantics. One advantage of this approach is that top-level ontologies need to be developed only once and then can be reused in many different domain ontologies.

first arg.	second arg.	relational top-level terms	Symbolic representation
individual	individual	individual-part-of	IP
individual	universal	instance-of	$Inst$
individual	collection	member-of	ϵ
universal	universal	sub-universal-of (is a)	\sqsubseteq
universal	universal	(up/down) universal-part-of	uUP, dUP
collection	universal	extension-of	Ext
collection	collection	sub-collection-of	\leq
collection	collection	(up/down) partonomically-included-in	uPI, dPI
collection	individual	sums-up-to	Sum
collection	individual	partition-of	Pt

Table 1. Types of top-level relations, their signatures, and their abbreviated top-level terms. (Adopted from [2].)

Another advantage is that a top-level ontology provides semantic links between the domain ontologies which are based on it.

A *logic-based* ontology is a logical theory [5]. The terms of the terminology, whose semantics is to be specified, appear as names, predicate and relation symbols of the formal language. Logical axioms and definitions are then added to express relationships between the entities, classes, and relations denoted by those symbols. Through the axioms and definitions the semantics of the terminology is specified by admitting or rejecting certain interpretations. In [2] a logic-based ontology for the top-level terms listed in Table 1 was presented.

Disciplines in which logic-based domain ontologies are quite common include medicine, biomedicine, and microbiology. Examples of logic-based medical domain ontologies are GALEN [6], SNOMED CT [7], and the NCI Thesaurus [8]. An example of a domain ontology for biomedicine and microbiology is the description logic based version of the Gene Ontology [9]. Currently efforts are being made to create a single suite of interoperable biomedical ontologies with a common top-level ontology as unifying ontological and formal basis [10, 11].

There are still only preliminary attempts to provide logic-based domain ontologies within the geo-spatial domains [12, 13]. Examples are [14, 15] for general ontologies of geographic categories, [16] for a domain ontology for hydrology, and [17, 18] for domain ontologies for ecosystems. The latter provide a starting point for this paper.

This paper uses the example of ecosystem classification and delineation to demonstrate how top-level ontologies can help to enhance the degrees to which information processing tools can be used in the retrieval, management and integration of data by improving the robustness and logical rigor of domain ontologies they are used to structure the data to be processed. In Section 2 a simplified version of a logic-based ontology for the top-level terms in Table 1. Section 3 discusses important distinctions in the use of top-level terms in top-level and domain ontologies. Section 4 demonstrates how to build a robust logic-based domain ontology of ecosystem classification and ecoregion de-

lineation by using the top-level ontology of Section 2 and the informal definitions of domain specific terms presented in [19].

2 A simple top-level ontology

Following [2] three disjoint sorts of entities are distinguished: (i) individual endurants (New York City, New York State, Planet Earth); (ii) endurant universals (*human being, heart, human settlement, socio-economic unit*); and (iii) collections of individual endurants (the collection of grocery items in my shopping bag at this moment in time, the collection of all human beings existing at a given time). In the logical theory this dichotomy of individuals, universals, and collections is reflected by distinguishing different sorts of variables – one sort for each category.

The theory is presented in a sorted first-order predicate logic with identity and use the letters w, x, x_1, y, z, \dots as variables ranging over (endurant) individuals; c, d, e, g as variables ranging over universals; and p, q, r, p_1, \dots as variables ranging over collections. The logical connectors $\neg, =, \wedge, \vee, \rightarrow, \leftrightarrow$ have their usual meanings (not, identical-to, and, or, if ... then, and if and only if (iff), respectively). The symbol \equiv is used for definitions. (x) symbolizes universal quantification (for all $x \dots$) and $(\exists x)$ symbolizes existential quantification (there is at least one $x \dots$). All quantification is restricted to a single sort. Restrictions on quantification will be understood by conventions on variable usage. Leading universal quantifiers are omitted. Labels for axioms begin with ‘A’ and labels for definitions begin with ‘D’.

Please note that the aim of this section is to give a self-contained and simplified axiomatic theory which is sufficient to demonstrate how to use a top-level ontology to build an atemporal domain ontology of ecosystem classification and delineation. For a more fully developed ontology see [2]. For additional discussions of universals see for example [20].

2.1 Mereology of individuals

Individual-part-of relations hold between individual endurants. For example, my heart is an individual part of my body, the Niagara Falls are individual parts of the Niagara River, Nebraska is an individual part of the United States of America. $IP\ xy$ signifies that individual x is part of individual y .

The individual x *overlaps* the individual y if and only if there exists an individual z such that z is a part of x and z is a part of y (D_O).

$$D_O \quad O\ xy \equiv (\exists z)(IP\ zx \wedge IP\ zy)$$

For example, Yellowstone National Park overlaps Wyoming, Montana, and Idaho.

The standard axioms requiring that individual parthood is reflexive (AM1), antisymmetric (AM2), transitive (AM3) are included in the theory. In addition it is required that if every z that overlaps x also overlaps y then x is part of y (AM4).

$$\begin{array}{ll}
AM1 \ IP \ xx & AM3 \ IP \ xy \wedge IP \ yz \rightarrow IP \ xz \\
AM2 \ IP \ xy \wedge IP \ yx \rightarrow x = y & AM4 \ (z)(O \ zx \rightarrow O \ zy) \rightarrow IP \ xy
\end{array}$$

2.2 Collections, sums, partitions, and partonomic inclusion

Collections are like (finite) sets of individuals with at least one member. Examples of collections include: the collection of Hispanic people in Buffalo's West Side as specified in the 2000 census records, the collection of federal states of the USA, the collection of postal districts in the USA, etc.

The symbol ' ϵ ' stands for the member-of relation between individuals and collections. The notation $\{x_1, \dots, x_n\}$ is used to refer to a finite collection having x_1, \dots, x_n as members. A minimal set of axioms requires: collections comprehend in every case at least one individual (AC1) and that two collections are identical if and only if they have the same members (AC2); for every x there is a collection having x as its only member (AC3); the union of two collection always exists (AC4).

$$\begin{array}{ll}
AC1 \ (\exists x)(x \epsilon p) & AC3 \ (\exists p)(p = \{x\}) \\
AC2 \ p = q \leftrightarrow (x)(x \epsilon p \leftrightarrow x \epsilon q) & AC4 \ (\exists r)(x)(x \epsilon r \leftrightarrow x \epsilon p \vee x \epsilon q)
\end{array}$$

The following definitions are included: collection p is a *sub-collection* of the collection q ($p \leq q$) if and only if every member of p is also a member of q (D_{\leq}); Collection p is *discrete*, $D p$, if and only if the members p do not overlap (D_D); The individual y is the *sum of the members of the collection* p if and only if every individual w overlaps y if and only if y overlaps some member of p D_{Sum} ; Collection p *partitions* the individual y if and only if y is the sum of p and p is discrete (D_{Pt}).

$$\begin{array}{ll}
D_{\leq} & p \leq q \equiv (x)(x \epsilon p \rightarrow x \epsilon q) \\
D_D & D p \equiv (x)(y)(x \epsilon p \wedge y \epsilon p \wedge O \ xy \rightarrow x = y) \\
D_{Sum} & Sum \ py \equiv (x)(O \ xy \leftrightarrow (\exists z)(z \epsilon p \wedge O \ xz)) \\
D_{Pt} & Pt \ pyt \equiv Sum \ pyt \wedge D \ pt
\end{array}$$

For example, the collection which has the federal states of the USA and the District of Columbia as its only members is discrete. The USA is the sum of this collection. Moreover, this collection partitions the USA.

Collection p is *upwards partonomically included* in collection q if and only if every member of p is an individual part of some member of q (uPI). Collection p is *downwads partonomically included* in collection q if and only if every member of q has some member of p as an individual part (dPI).

$$\begin{array}{ll}
D_{uPI} & uPI \ pq \equiv (x)(x \epsilon p \rightarrow (\exists y)(y \epsilon q \wedge IP \ xy)) \\
D_{dPI} & dPI \ pq \equiv (y)(y \epsilon q \rightarrow (\exists x)(x \epsilon p \wedge IP \ xy))
\end{array}$$

For example, let USC be the collection which has all the counties¹ of the USA as its members and let USF be the collection that has all the federal states of the

¹ To keep matters simple I ignore the fact that in Louisiana counties are called 'parish' and in Alaska counties are called 'borough'.

USA as its members. Then *USC* is up- and downwards partitionomically included in *USF*: every member of *USC* (a county) is part of some member of *USF* (a federal state) and every member of *USF* (a federal state) has some member of *USC* (a county) as its part.

2.3 Universals, instantiation, and universal parthood

The variables c, d, e, g are used for universals (classes, types) like (*human being, federal state, mountain, forest, tree, plant*, and so forth). The relation of instantiation holds between individuals and universals. For example New York City is an instance of the universal *city*. ‘*Inst xc*’ signifies that the individual x instantiates the universal c . In terms of *Inst* one can define: c is a sub-universal-of d if and only if the instances of c are also instances of d (D_{\sqsubseteq}); c is a proper sub-universal-of d if and only if c is a sub-universal of d and d is not a sub-universal-of c (D_{\sqsubset}); collection p is the extension of universal c if and only if for all x , x is a member of p if and only if x instantiates c (D_{Ext}).

$$\begin{aligned} D_{\sqsubseteq} & \quad c \sqsubseteq d \equiv (x)(Inst\ xc \rightarrow Inst\ xd) \\ D_{\sqsubset} & \quad c \sqsubset d \equiv c \sqsubseteq d \wedge d \not\sqsubseteq c \\ D_{Ext} & \quad Ext\ pc \equiv (x)(x \in p \leftrightarrow Inst\ xc) \end{aligned}$$

For example, the universal *federal state* is a sub-universal-of the universal *socio-economic unit*. Therefore every instance of *federal state* (e.g., New York State) is also an instance of *socio-economic unit*. The extension of the universal *federal state* is the collection of all federal states. This collection has as members the federal states of the USA, the federal states of Germany, etc.

The following axioms are included: every universal has an instance (AU1); there is maximal universal (AU2); two universals are identical if and only if they have the same instances (AU3);² if two universals share a common instance then one is a sub-universal of the other (AU4); and if c is a proper sub-universal of d then there is a proper sub-universal e of d such that c and d have no instance in common (A5).

$$\begin{array}{ll} AU1 (\exists x)Inst\ xc & AU3 (x)(Inst\ xc \leftrightarrow Inst\ xd) \leftrightarrow c = d \\ AU2 (\exists x)(y)(y \sqsubseteq x) & AU4 (\exists x)(Inst\ xc \wedge Inst\ xd) \rightarrow c \sqsubseteq d \vee d \sqsubseteq c \\ & AU5 c \sqsubset d \rightarrow (\exists e)(e \sqsubset d \wedge \neg(\exists x)(Inst\ xc \wedge Inst\ xe)) \end{array}$$

From these axioms it follows that universals form tree-like hierarchies ordered by the sub-universal relation. In the scientific realm such tree-like structures most closely resemble classification hierarchies established using the Aristotelean method of classification. Using this method classification trees (intended to resemble hierarchies of universals) are built by defining a universal lower down in the hierarchy by specifying the parent universal together with the relevant

² This certainly is an oversimplification but will not cause problems for the limited scope of this paper.

differentia, which specify what marks out instances of the defined universal or species within the wider parent universal or genus, as in: *human* =_{df} *rational animal* where ‘rational’ is the differentia [21, 18]. Differentia need to be such that the immediate sub-universals of a given universal are jointly exhaustive and pairwise disjoint. Thus besides rational animals there are non-rational animals and all animals are either rational or non-rational.

Corresponding to the partonomic inclusion relations *uPI* and *dPI* between collections the relations of upward and downward universal parthood, *uUP* and *dUP*, between universals are introduced: *c* is an *upward-universal-part-of* universal *d* if and only if every instance of *c* is an individual part of some instance of *d* (D_{uUP}); *c* is a *downward-universal-part-of* universal *d* if and only if every instance of *d* has some instance of *c* as an individual part (D_{dUP}).

$$\begin{aligned} D_{uUP} \quad uUP \, cd &\equiv (x)(Inst \, xc \rightarrow (\exists y)(Inst \, yd \wedge IP \, xy)) \\ D_{dUP} \quad dUP \, cd &\equiv (y)(Inst \, yc \rightarrow (\exists x)(Inst \, xc \wedge IP \, xy)) \end{aligned}$$

For example, the universal *waterfall* is an upwards-universal-part of the universal *river*, since every instance of *waterfall* is individual-part-of some instance of *river*.

The formal theory presented in this section is called TLO. A computational representation of this theory can be found at <http://www.buffalo.edu/~bittner3/Theories/Papers/Cosit2007Theory.html>. A more sophisticated version which constitutes the basis of Basic Formal Ontology (BFO) can be accessed via <http://www.ifomis.org/bfo/fol>.

3 From top-level ontologies to domain ontologies

The terms of a top-level ontology refer to *classes* of relations in certain intended domains of interpretation which satisfy the relevant axioms of the top-level ontology. In the remainder I will refer to the relations that satisfy the axioms associated to the top-level term *T* of TLO in the intended domains of interpretation *relations of type T*. For example, relations (in the intended domains of interpretation) which satisfy the axioms (AU1-5, D_{\sqsubseteq} , and D_{\sqsubset}) associated with the term ‘sub-universal-of’ (abbreviated by \sqsubseteq) are called *relations of type sub-universal-of*, or sub-universal-of relations for short. Similarly, relations (in the intended domains of interpretation) which satisfy the axioms (AM1-4 and D_O) associated with the term ‘individual-part-of’ (abbreviated by *IP*) are called *relations of type individual-part-of*, or individual-part-of relations for short.

In domain ontologies top-level terms often refer *specific relations* in a particular domain. For example, in a domain ontology of socio-economic units, the term ‘sub-universal-of’ may refer to a specific relation which holds between socio-economic units, and which satisfies the axioms associated with the term ‘sub-universal-of’ in TLO. Thus in some sense, a domain ontology is a formal representation of one specific *model* (in the model-theoretic sense) of the underlying top-level ontology.

The distinction between types of relations as they are specified in top-level ontologies and particular relations of a given type in a given domain becomes

even more important in domains where there is more than one relation of a given type. In the ecosystem example in Section 4 there will be three distinct relations of type sub-universal-of and three distinct relations of type universal-part-of.

The remainder of this section addresses more formally the distinction between a specific relation on a particular domain and types of relations as specified in a top-level ontology.

3.1 Specific binary relations

A (specific) *binary relation* R with domain of discourse $\mathcal{D}(R)$ is a set of ordered pairs of members of the set $\mathcal{D}(R)$, i.e., $R \subseteq \mathcal{D}(R) \times \mathcal{D}(R)$.³ If R is a binary relation with domain $\mathcal{D}(R)$ then I will also say that R is a binary relation on $\mathcal{D}(R)$. I write $R(x, y)$ to say that R holds between $x, y \in \mathcal{D}(R)$, i.e., $R(x, y)$ if and only if $(x, y) \in R$. If R is a binary relation on $\mathcal{D}(R)$ then one can define the relations uR and dR on the powerset (the set of all subsets) of $\mathcal{D}(R)$, i.e., $\mathcal{D}(uR) = \mathcal{D}(dR) = \mathcal{P}(\mathcal{D}(R))$, as follows:

$$\begin{aligned} uR(X, Y) &=_{df} \forall x \in X : \exists y \in Y : R(x, y) \\ dR(X, Y) &=_{df} \forall y \in Y : \exists x \in X : R(x, y) \end{aligned} \quad (1)$$

R is called an *individual-level* relation on $\mathcal{D}(R)$ while uR and dR are *class-level* relations on $\mathcal{P}(\mathcal{D}(R))$ [11, 22]. For example, if R is the individual-part-of relation between human body parts (e.g., my left arm is an individual-part-of my body) then the class-level relation uR is the relation upwards-universal-part-of between body part universals (the universal *left arm* is a upwards-(and downwards)-universal-part-of the universal *human body*.)

For any binary (individual-level or class-level) relation R , one can define the *immediate- R -relation*, R_i on $\mathcal{D}(R)$ in terms of R : The immediate- R -relation, R_i , holds between x and y if and only if the relation R holds between x and y and there is not member z of the domain of R such that $R(x, z)$ and $R(z, y)$.⁴

$$R_i(x, y) =_{df} R(x, y) \text{ and } \neg(\exists z)(z \in \mathcal{D}(R) \text{ and } R(x, z) \text{ and } R(z, y)) \quad (2)$$

Consider Fig. 1 which depicts the graph of the relation immediate-sub-universal-of between ecosystem universals such that the set of nodes of the graph is the domain of this relation. The relation immediate-sub-universal-of is a specific example of an R_i -relation in the sense of Definition 2. The corresponding R -relation is the sub-universal-of relation between the ecosystem universals represented by the nodes of the graph.

In general, let $\Gamma = (N, E)$ be the graph of the relation R_i . Then the set of nodes of Γ is the domain of R_i , i.e., $N = \mathcal{D}(R_i) = \mathcal{D}(R)$, and $(x, y) \in R_i$

³ In the meta-language the language of set theory is used to talk about specific relations and their properties in a general but precise way.

⁴ There may be situations where R_i is the empty relation – See [23] for details. For most non-empty R_i , R_i is intransitive – See [23] and Table 2.

if and only if there is an edge from node x to node y in the graph, i.e., $V = \{(x, y) | R_i(x, y)\}$. In the remainder of this paper it is always assumed that the set of nodes, N , is finite, and it will always be the case that if Γ is the graph of the relation R_i then R is the reflexive and transitive closure of R_i . R_i -relations will be used in Section 4 to draw graphs of specific binary relations between individual ecoregions and ecosystem classes. These graphs serve as graphic representations of a ‘formal’ domain ontology. Under the assumptions made here one can always obtain R from the graph of R_i and vice versa.

Any given binary (individual-level or class-level) relation either has or lacks each of the logical properties listed in Table 2 (and of course others). The properties of reflexivity, antisymmetry, transitivity, as well as the root property are standard and do not need further discussion. Relation R has the *no-single-immediate-predecessor* property (NSIP) if and only if for all $x, y \in \mathcal{D}(R)$, if $R_i(x, y)$ then there is a $z \in \mathcal{D}(R)$ such that $R_i(z, y)$ and $x \neq z$. For example, the relation immediate-sub-universal-of in Fig. 1, has the NSIP property: every non-leaf node of the graph has at least two children. The relation which graph is depicted in Fig. 2(b) lacks the NSIP property.

Relation R has the *single-immediate-successor* property (SIS) if and only if some members of $\mathcal{D}(R)$ stand in the R_i relation and no $x \in \mathcal{D}(R)$ stands in the R_i relation to distinct members of $\mathcal{D}(R)$. For example, the relation immediate-sub-universal-of in Fig. 1, has the SIS property: every node has at most one parent. Graphs of relations that lack the SIS property form lattice structures, in which there may be nodes with more than one parent, rather than trees. (See [23] for extended discussions.)

property	description
reflexive	$\forall x \in \mathcal{D}(R) : R(x, x)$
antisymmetric	$\forall x, y \in \mathcal{D}(R) : \text{if } R(x, y) \text{ and } R(y, x) \text{ then } x = y$
transitive	$\forall x, y, z \in \mathcal{D}(R) : \text{if } R(x, y) \text{ and } R(y, z) \text{ then } R(x, z)$
intransitive	$\forall x, y \in \mathcal{D}(R) : \text{if } R(x, y) \text{ and } R(y, z) \text{ then not } R(x, z)$
root	$\exists x \in \mathcal{D}(R) : [\exists y \in \mathcal{D}(R) : R(y, x) \text{ and } \forall y \in \mathcal{D}(R) : R(y, x) \text{ or } y = x]$
NSIP	$\forall x, y \in \mathcal{D}(R) : (\text{if } R_i(x, y) \text{ then } \exists z \in \mathcal{D}(R) : (R_i(z, y) \text{ and } x \neq z))$
SIS	$\exists x, y \in \mathcal{D}(R) : R_i(x, y) \text{ and}$ $\forall x, y, z \in \mathcal{D}(R) : (\text{if } R_i(x, y) \text{ and } R_i(x, z) \text{ then } y = z)$

Table 2. Properties which a binary relation R either has or lacks. [23].

3.2 Properties and types of relations

One can classify binary relations according to their logical properties. Relations of type *partial ordering* are relations which are reflexive, antisymmetric, and transitive. Relations of type *individual-part-of* are relations that hold between individual entities and that are reflexive, antisymmetric, transitive, and in addition satisfy axiom AM4. Hence relations of type *individual-part-of* are also of

type *partial ordering*. Relations of type *rooted partial ordering* are relations of type *partial ordering* that also have the root property. Since relations satisfy the axioms of the top-level ontology in virtue of the properties they possess or lack, the types of relations defined in terms of satisfaction of axioms of the top-level ontology corresponds to the typing of relations according to their logical properties.

Relations of type *finite list-or-tree forming* are relations of type *rooted partial ordering* which are such that if R is a relation of type *rooted partial ordering* then R_i (the immediate R -relation in the sense of Def. 2) has the SIS property. All the relations whose R_i -graphs are depicted in this paper are of type *finite list-or-tree forming*. Intended interpretations of the top-level term universal-part-of are relations of type *finite tree forming* that hold between between universals and which are related to the individual-part-of relation between their instantiating individuals in the ways specified in D_{uUP} and D_{dUP} .

Relations of type *finite tree forming* are relations of type *rooted partial ordering* which are such that if R is a relation of type *rooted partial ordering* then R_i has the NSIP and the SIS properties. The relations whose R_i -graphs are depicted in Figures 1, 2(a), 3, 4, 5(a) and 5(b) all are of type *finite tree forming*. Intended interpretations of the top-level term sub-universal-of of TLO are relations of type *finite tree forming* that hold between between universals and are related in the appropriate ways to the instance-of relation between individuals and universals, i.e., satisfy the axioms AU1-5, D_{\square} , and D_{\square} .

In the remainder of this paper I use SMALL CAPITAL LETTERS to signify top-level terms referring to types of entities such as INDIVIDUAL and UNIVERSAL, and for top-level terms referring to types of relations such as INDIVIDUAL-PART-OF, UNIVERSAL-PART-OF, SUB-UNIVERSAL-OF, etc. These terms correspond to the symbols used in the formal theory, TLO, as summarized in Table 1. I use **typewriter font** and superscripts to distinguish specific relations in the domain ontology from relation-types denoted by top-level terms in the top-level ontology. I.e., I write **sub-universal-of**¹ and **sub-universal-of**² to refer to distinct specific relation among ecosystem universals, both of which satisfy the axioms of the SUB-UNIVERSAL-OF relation in the top-level ontology in virtue of their logical properties.

4 Ecosystem classification and delineation

In this section a domain ontology of ecosystem classification and delineation is developed. The domain ontology is based on Bailey's influential paper "Delineation of ecosystem regions" [19]. Bailey's specification of his domain ontology is from a formal-ontological perspective rather imprecise and ambiguous. It will be made more rigorous and precise by (a) using the semantically well-defined top-level terms listed in Table 1 and (b) by enforcing the clear distinction of types of relations as specified in the top-level ontology and specific relations of a given type as they occur in the ecosystem domain. This will make it possible to explicitly distinguish a number of relations which have been confused and been taken

to be a single relation before. Graphic rather than symbolic representations for those relations will be used in this section. The symbolic representations can be obtained from the graphs as described in Section 3.1.

4.1 Classification of geographic ecosystems with respect to broad climatic similarity, and definite vegetational affinities

Consider the following sentence:

(BL1) "Ecoregions are large ecosystems of regional extent that contain a number of smaller ecosystems. They are geographical zones that represent geographical groups of similarly functioning ecosystems" [19, p. 365]

Using the terms of the top-level ontology one can rephrase (BL1) as follows: Ecoregions are INDIVIDUALS that are INDIVIDUAL-PARTS-OF the biosphere on the surface of the Earth (an INDIVIDUAL). Ecosystems are UNIVERSALS and are INSTANTIATED BY ecoregions⁵ with similar functional characteristics. Geographic ecosystems are ecosystem UNIVERSALS which instantiating ecoregions are of geographic scale or larger. Every ecoregion has smaller ecoregions as INDIVIDUAL-PARTS all of which are INSTANCES-OF the UNIVERSAL *ecosystem*.

According to this interpretation of (BL1), 'are' (in BL1) is intended to mean INSTANTNCE-OF, 'contains' is intended to mean HAS-INDIVIDUAL-PART, 'represent' is intended to mean INSTANCE-OF, 'geographical groups of similarly functioning ecosystems' is intended to mean UNIVERSALS that are INSTANTIATED BY similarly functioning ecoregions (INDIVIDUALS).

(BL2) "Regional boundaries may be delineated ... by analysis of the environmental factors that most probably acted as selective forces in creating variation in ecosystems" [19, p. 366]

(BL2) indicates that the *differentia* used for distinguishing ecosystem UNIVERSALS are environmental factors that create the variations between the ecoregions that INSTANTIATE distinct ecosystem UNIVERSALS. Environmental factors used as differentia fall into the two major groups of climate and vegetation [19]. The climate categorization is based on the annual and monthly averages of temperature and precipitation [24].

Thus, geographic ecosystems are UNIVERSALS which INSTANCES, ecoregions, are characterized by broad climatic similarity, definite vegetational affinities, etc. The resulting classification hierarchy is depicted in Fig. 1 where the nodes of the depicted graph are ecosystem UNIVERSALS that form the domain of the relation **sub-universal-of**¹ and the directed edges represent the relation **sub-universal-of**_{*i*}¹ (the immediate sub-universal-of relation in the sense of equation 2

⁵ Ecoregions here are individual ecosystem, i.e., 'ecoregion' and 'ecosystem individual' treated as synonyms. Following Bailey no distinction between an individual ecosystem and the region it occupies at a given point in time is made.

which domain corresponds to the nodes of the graph). The root of the tree is the ecosystem UNIVERSAL *geographic ecosystem*. The relevant differentia which indicate what marks out INSTANCES of the immediate SUB-UNIVERSALS of the ROOT are roughly Koeppen’s climate groups [19]⁶. For example,

Humid Temperate Ecosystem =_{df} *Geographic Ecosystem* with humid temperate climate.

The relevant differentia which tell us what marks out INSTANCES of the immediate SUB-UNIVERSALS of the ecosystem UNIVERSALS that are differentiated by climate groups, are roughly Koeppen’s climate types [19]. For example,

Prairie Ecosystem =_{df} *Humid Temperate Ecosystem* with prairie climate.

The relevant differentia which tell us what marks out INSTANCES of the immediate SUB-UNIVERSALS of the ecosystem UNIVERSALS that are differentiated by climate groups and climate types are climax plant formations [19]. For example,

Prairie Bushland Ecosystem =_{df} *Prairie Ecosystem* with climax vegetation type Bushland.

It follows that all ecoregions that INSTANTIATE the UNIVERSAL *Humid Temperate Ecosystem* are characterized by the humid temperate climate group. Similarly, ecoregions that INSTANTIATE the UNIVERSAL *Prairie Bushland Ecosystem* are characterized by the humid temperate climate group, by the prairie climate type, and by bushland vegetation. In other words: *Prairie Bushland Ecosystem* is a sub-universal-of¹ *Prairie Ecosystem* which in turn is a sub-universal-of¹ *Humid temperate domain* and thus, according to definition (D_{\square}), every INSTANCE-OF *Prairie Bushland Ecosystem* is an INSTANCE-OF *Prairie Ecosystem* and is also an INSTANCE-OF *Humid Temperate Ecosystem*.

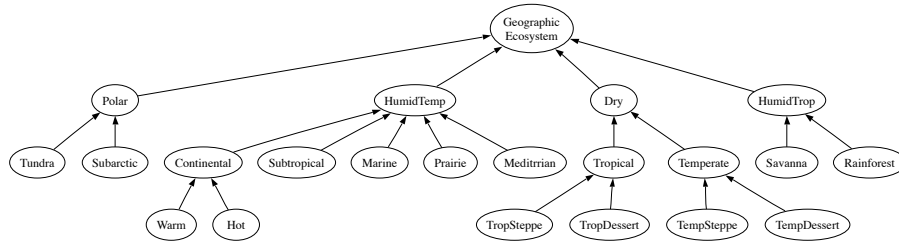


Fig. 1. Graphs of the relations sub-universal-of¹ and upwards-universal-part-of¹. Notice that in Bailey’s classification all the leaf nodes in the depicted graph have further sub-universals which are omitted here. See [19, p. 369] for a more complete tree.

⁶ Bailey collapses Koeppen’s subtropical and temperate climate groups into ‘Humid temperate’.

4.2 Universal parthood relations between ecosystem universals

One can make (BL1) even more precise by using the top-level relation UNIVERSAL-PART-OF to specify more precisely what is meant by "Ecoregions are large ecosystems of regional extent that contain a number of smaller ecosystems". This portion of (BL1) indicates that, corresponding to the relation **sub-universal-of**¹ between ecosystem UNIVERSALS, there is a hierarchical order in which smaller ecosystem UNIVERSALS are parts of larger ecosystem UNIVERSALS.

Let ecosystem UNIVERSAL E_1 be a **sub-universal-of**¹ ecosystem UNIVERSAL E_2 as depicted in Fig. 1. One can verify that every INSTANCE-OF UNIVERSAL E_1 is an INDIVIDUAL-PART-OF some INSTANCE-OF UNIVERSAL E_2 . Thus, in addition to the relation **sub-universal-of**¹, the relation **upwards-universal-part-of**¹ holds between the ecosystem UNIVERSALS depicted in Fig. 1. For example, every ecoregion which is an INSTANCE-OF the UNIVERSAL *Prairie Bushland Ecosystem* is INDIVIDUAL-PART-OF an ecoregion that is an INSTANTCE-OF the UNIVERSAL *Prairie Ecosystem*. Similarly, every ecoregion that is an INSTANCE-OF the UNIVERSAL *Prairie Ecosystem* is in turn INDIVIDUAL-PART-OF some ecoregion that is an INSTANCE-OF the UNIVERSAL *Humid Temperate Ecosystem*.

One can see that there is a correspondence between the relation **sub-universal-of**¹ and **upwards-universal-part-of**¹ in the sense that for all nodes E_1 and E_2 in the graph of Fig. 1: E_1 is a **sub-universal-of**¹ E_2 if and only if E_1 is a **upwards-universal-part-of**¹ E_2 . However, from the top-level ontology it is clear that these two relations are very different: **upwards-universal-part-of**¹ is a class-level relation corresponding to an individual-level relation of type INDIVIDUAL-PART-OF (Definitions (1) and (D_{uUP})). By contrast, the relation **sub-universal-of**¹ is NOT a class-level version of an individual-level relation.⁷

4.3 Classification of ecosystems according to kinds of climatic and vegetation characteristics

(BL3) "A hierarchical order is established by defining successively smaller ecosystems within larger ecosystems ... subcontinental areas, termed domains, are identified on the basis of broad climatic similarity ... domains ... are further subdivided, again on the basis of climatic criteria, into divisions ... divisions correspond to areas having definite vegetational affinities" [19, p. 366]

(BL3) indicates that, in addition to the classification of ecosystems according to *particular* climatic and vegetational affinities among the INSTANTIATING ecoregions (Fig. 1), Bailey also classifies ecosystems according to the *kinds* of climatic and vegetation characteristics that characterize the INSTANTIATING ecoregions. Bailey [19] distinguishes the following (additional) geographic ecosystem UNIVERSALS: *domains* are ecosystem UNIVERSALS, which INSTANTIATING ecoregions

⁷ This indicates some serious limitations of the extensional conception of relations as introduced in Section 3.1, which identifies the relations **sub-universal-of**¹ and **upwards-universal-part-of**¹.

are characterized *only* by the climatic group; *divisions*, which INSTANTIATING ecoregions are characterized by climatic group *and* climatic type but *not* by plant formations; and *provinces*, which INSTANTIATING ecoregions are characterized by climatic group *and* climatic type *and* the climax plant formations [19, Fig. 1, p. 367].⁸ The graph of the resulting relation **sub-universal-of_i²** is depicted in Fig. 2(a), which shows ecosystem UNIVERSALS as nodes and the relation **sub-universal-of_i²** as directed edges.

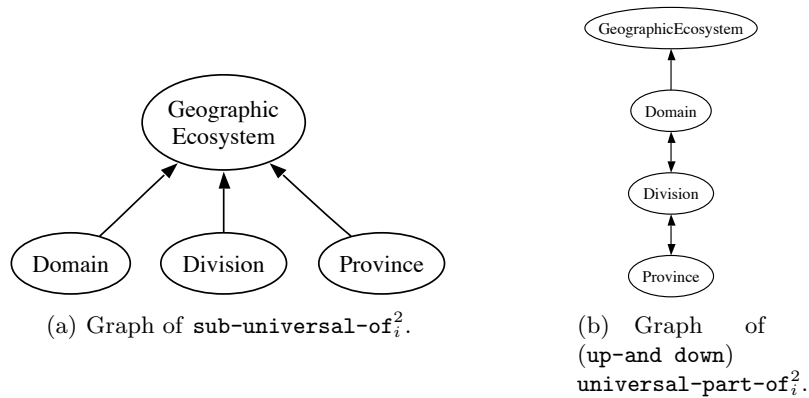


Fig. 2. Relations among ecosystem universals whose instances are characterized by the same kinds of climatic and vegetation characteristics.

Under the assumption that ecoregions that are homogeneous in climate group, type, etc. are also spatially maximal⁹, there is an interrelationship between the size of an ecoregion and the kinds of climatic, vegetation, etc. characteristics that characterize that ecoregion in the sense described in the previous paragraph. As pointed out in (BL3) the universal *domain* is INSTANTIATED BY ecoregions of subcontinental scale. Ecoregions that INSTANTIATE the universals *division* and *province* are of successively smaller scales.

The hierarchical ‘nesting’ of provinces into divisions into domains is captured by relations of type UNIVERSAL-PART-OF. Let **upwards-universal-part-of²** be the relation which holds between the universals *domain*, *division*, *province* and *Geographic ecosystem* such that every INSTANCE-OF the UNIVERSAL *division* is INDIVIDUAL-PART-OF some INSTANCE-OF the UNIVERSAL *domain* and similarly for *province* and *division*, and *domain* and *Geographic Ecosystem*. The graph of **upwards-universal-part-of_i²** is represented by the upwards pointing arrows in Fig. 2(b).

⁸ Sections are omitted in this paper.

⁹ Bailey seems to mean maximal but not necessarily singly-connected parts of a given continent.

Between the universals *domain*, *division*, and *province* in addition the relation **downwards-universal-part-of**² holds: every INSTANCE-OF the UNIVERSAL *domain* has some INSTANCE-OF the UNIVERSAL *division* as an INDIVIDUAL-PART and every INSTANCE-OF the UNIVERSAL *division* has some INSTANCE-OF the UNIVERSAL *province* as an INDIVIDUAL-PART. Notice, that NOT every INSTANCE-OF *Geographic Ecosystem* has some INSTANCE-OF *domain* as an INDIVIDUAL PART. (The Subtropical Division is an INSTANCE OF *geographic ecosystem* but there is no INSTANCE OF *domain* that is as an INDIVIDUAL PART OF the Subtropical Division.) The graph of the relation **downwards-universal-part-of**_{*i*}² is represented by the downwards arrows in Fig. 2(b).

One can see that, in contrast to the graphs of **sub-universal-of**_{*i*}¹ and **upwards-universal-part-of**_{*i*}¹, the graphs of **sub-universal-of**_{*i*}² and **upwards-universal-part-of**_{*i*}² are quite different. This is another reason why it is important to distinguish between relations of type SUB-UNIVERSAL-OF and relations of type (UPWARDS-) UNIVERSAL-PART-OF.

4.4 ‘Intersecting’ both classifications

There are ecoregions that are INSTANCES-OF both, the UNIVERSAL *domain* and the UNIVERSAL *Humid Temperate Ecosystem*. Bailey [19] calls the UNIVERSAL that has as INSTANCES all ecoregions that are INSTANCES-OF both, *domain* and *Humid Temperate Ecosystem*, *Humid Temperate Domain*. *Dry domain* is the UNIVERSAL that has as INSTANCES all ecoregions that INSTANTIATE *Dry ecosystem* and *domain*. Similarly for *Polar domain*, *Tundra division*, etc. (See also [17] for a similar approach or ‘intersecting’ classification trees.)

Let **sub-universal-of**³ be the SUB-UNIVERSAL-OF relation which has as its domain the set which has as its members UNIVERSALS that are constructed in the way described in the previous paragraph and, in addition, the UNIVERSAL *Geographic ecosystem*. An important feature of the relation **sub-universal-of**³ is, that *Geographic ecosystem* is the only UNIVERSAL that a has proper SUB-UNIVERSAL. Thus, the graph of **sub-universal-of**_{*i*}³ is a flat but rather broad tree as indicated in Fig. 3.

The graph of the relation **sub-universal-of**_{*i*}³ is a *refinement* of the graph of the relation **sub-universal-of**_{*i*}² in Fig. 2(a) in the sense that each of the nodes *domain*, *division*, and *province* in the graph of the relation **sub-universal-of**_{*i*}² is replaced by a set of jointly exhaustive and pairwise disjoint UNIVERSALS. For example, the node *Domain* in Fig. 2(a) is replaced by the nodes *PolarDomain*, *Humid Temperate Domain*, *Dry domain*, and *Humid Tropical domain* in Fig. 3. Similarly for the other nodes in Fig. 2(a).

Let **upwards-universal-part-of**³ be the UPWARDS-UNIVERSAL-PART-OF relation on the domain of **sub-universal-of**³. The graph of **upwards-universal-part-of**_{*i*}³ (Fig. 4) is a refinement of the graph of **upwards-universal-part-of**_{*i*}² (Fig. 2(b)) in the sense that the nodes *domain*, *division*, and *province* of the graph of **upwards-universal-part-of**_{*i*}² correspond to layers or *cuts* [18] in the tree formed by the graph of **sub-universal-of**_{*i*}³.

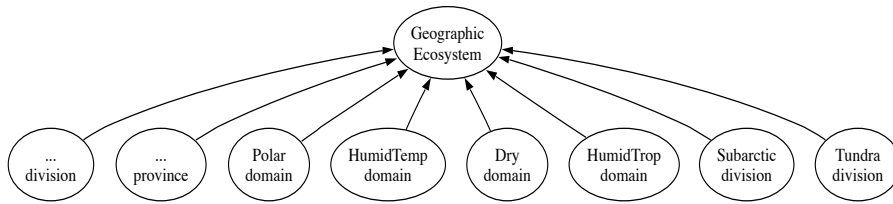


Fig. 3. Graph of the relation **sub-universal-of**_i³.

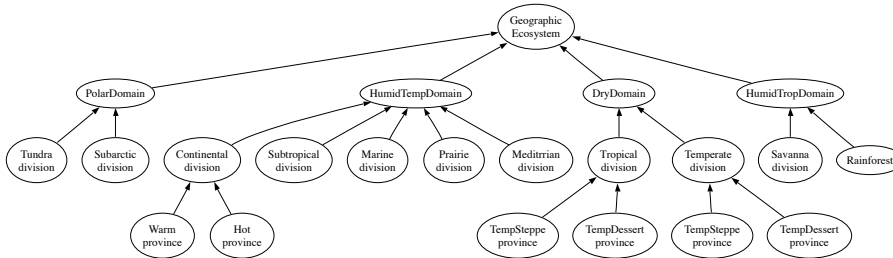


Fig. 4. Graph of the relation **universal-part-of**_i³.

Notice, that there is a graph-isomorphism between the graphs of the relations **sub-universal-of**_i¹, **upwards-universal-part-of**_i¹, and **upwards-universal-part-of**_i³, i.e., the graphs of the relations are structurally identical. Notice, however, that the *domains* of the relations **upwards-universal-part-of**¹ and **upwards-universal-part-of**³ are quite different, since the **SUB-UNIVERSAL-OF** relations, **sub-universal-of**_i¹ and **sub-universal-of**_i³ are very different as easily recognizable in Figures 1 and 3.

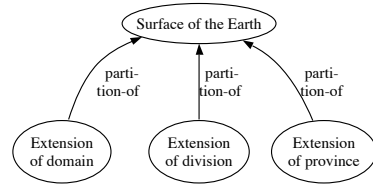
4.5 Ecosystem delineation

The discussion so far has focused on ecosystem classifications (**SUB-UNIVERSAL-OF** relations) and on the hierarchical spatial nestings that are induced by these classifications through the corresponding **UNIVERSAL-PART-OF** relations. However Bailey [19] also emphasizes the delineation of ecoregions. Delineation here refers to the establishing of fiat boundaries [25]¹⁰ that separate ecoregions which **INSTANTIATE** ecosystem **UNIVERSALS** that are *differentiated* in the **SUB-UNIVERSAL** hierarchy. That is, *delineation* at the level of ecoregions (**INDIVIDUALS**) corresponds to establishing *differentia* between ecosystem **UNIVERSALS**.

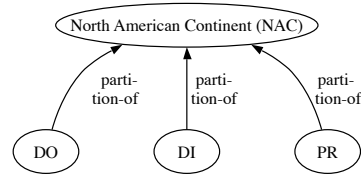
Consider the classification of geographic ecosystem universals into *domain*, *division*, *province*, etc. (formally represented by the relation **sub-universal-of**² as depicted via **sub-universal-of**_i² in Fig. 2(a)). This classification is such that the **EXTENSION** of the **UNIVERSAL domain** (i.e., the **COLLECTION** of ecoregions that **INSTANTIATE** the **UNIVERSAL domain**) **PARTITIONS** the (biosphere

¹⁰ Of course, those boundaries are subject to vagueness, as Bailey himself points out.

on) the surface of the Earth in the sense that that (i) no distinct MEMBER-OF the EXTENSION-OF *domain* have a common INDIVIDUAL-PART and (ii) jointly the MEMBERS-OF the EXTENSION-OF *domain* SUM-UP-TO the INDIVIDUAL ‘(Bio-sphere on) the surface of Earth’. That is, the ecoregions that are MEMBERS OF the EXTENSION OF *domain* are jointly exhaustive and pair-wise disjoint. Similarly, the EXTENSIONS of *division* and *province* all PARTITION the INDIVIDUAL *Surface of Earth*. (Fig. 5(a).)



(a) The extensions of the universals *domain*, *division*, and *province* partition the surface of Earth. (**partition-of^E**)



(b) Collections of geographic ecoregions that partition *NAC*. **partition-of^{NAC}**

Fig. 5. Partitions formed by collections of ecoregions.

Consider the UNIVERSAL *Dry Domain*. Obviously, not all ecoregions that are INSTANCES OF this UNIVERSAL are INDIVIDUAL-PARTS-OF the North American continent (*NAC*). In fact, there is only a single INSTANCE of *Dry Domain* that is INDIVIDUAL-PART-OF *NAC*. The other INSTANCES-OF *Dry Domain* are INDIVIDUAL-PARTS-OF other continents.

For many practical purposes it is useful to refer not to all INSTANCES of UNIVERSALS like *domain*, *division*, and *province* but only to those INSTANCES that are INDIVIDUAL-PART-OF the North American Continent (*NAC*). For this purpose the notion of COLLECTION is used. Let *DO* be the COLLECTION of ecoregions that are INSTANCES-OF the UNIVERSAL *domain* and that are INDIVIDUAL-PART-OF *NAC*; let *DI* be the COLLECTION of ecoregions that are INSTANCES-OF the UNIVERSAL *division* and that are INDIVIDUAL-PART-OF *NAC*; and let *PR* be the COLLECTION of ecoregions that are INSTANCES-OF the universal *province* and that are INDIVIDUAL-PART-OF *NAC*. The COLLECTION *DO* is a PARTITION-OF the INDIVIDUAL North American Continent (*NAC*) in the sense that (i) no distinct MEMBERS-OF *DO* have a common INDIVIDUAL-PART and (ii) jointly the MEMBERS OF *DO* SUM-UP-TO *NAC*. Similarly, the COLLECTIONS *DI* and *PR* are PARTITIONS-OF *NAC*. The graph of the relation **partition-of^{NAC}** is depicted in Fig. 5(b).

The COLLECTIONS *DO*, *DI*, and *PR* not only PARTITION *NAC* they are also (up-and downward) PARTONOMICALLY-INCLUDED in one another in the sense that they are hierarchically structured such that every MEMBER-OF *PR* is an INDIVIDUAL-PART-OF some MEMBER-OF *DI* and every MEMBER-OF *DI* has some

MEMBER-OF *PR* as INDIVIDUAL-PART. Similarly for *DI* and *DO* and for *PR* and *DO*. Of course this mirrors the UNIVERSAL-PART-OF relations **upwards** and **downwards-universal-part-of**² between the universals *province*, *division*, and *domain* (Fig. 2(b)).

5 Conclusions

In this paper a logic-based top-level ontology was used to create a domain ontology of ecosystem classification and delineation. The aim was to express the domain ontology underlying Bailey's paper 'Delineation of Ecoregions' [19] in a logically rigorous form that is accessible not only to human domain specialists but also to computers.

Notice that the claim of the paper is not that Bailey's definitions must be interpreted in the ways suggested here. In fact it is a weakness of Bailey's definitions that they are imprecise (at least from a logical and computational perspective) and can be interpreted in different ways, and thus leave (at least from the perspective of a non-domain specialist) the exact nature of the relationships between the classification of ecosystems into different kinds and the spatial nesting of ecoregions that instantiate those kinds implicit.

It was the aim of this paper to make one possible interpretation of Bailey's work as precise as possible by using notions like INDIVIDUAL, UNIVERSAL, INSTANCE-OF, PART-OF, UNIVERSAL-PART-OF, PARTITION-OF, etc., which exact meaning was specified using an axiomatic theory. This makes it easier for other researchers to understand and to criticize this particular interpretations.

The analysis of this paper also showed is that it is important to distinguish types of relations as specified at the top-level from specific relations of a given type as they occur in a specific domain. By taken this distinction into account it became possible to explicitly distinguish a number of relations among ecosystems and ecoregions which have been confused and been taken to be a single relation before.

References

1. Guarino, N.: Formal ontology and information systems. In Guarino, N., ed.: Formal Ontology and Information Systems, (FOIS'98), IOS Press (1998) 3–15
2. Bittner, T., Donnelly, M., Smith, B.: Individuals, universals, collections: On the foundational relations of ontology. In Varzi, A., Vieu, L., eds.: Proceedings of the third International Conference on Formal Ontology in Information Systems, FOIS04, IOS Press (2004) 37–48
3. Rector, R.: Modularization of domain ontologies implemented in description logics and related formalisms including OWL. In: Proceedings of the international conference on Knowledge capture. (2003) 121–128
4. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening ontologies with DOLCE. *AI Magazine* **23**(3) (2003) 13–24
5. Copi, I.: *Symbolic Logic*. Prentice Hall, Upper Saddle River, NJ 07458 (1979)

6. Rector, A., Rogers, J.: Ontological issues in using a description logic to represent medical concepts: Experience from GALEN: Part 1 - principles. *Methods of Information in Medicine* (2002)
7. Spackman, K., Campbell, K., Cote, R.: SNOMED RT: A reference terminology for health care. In: *Proceedings of the AMIA Annual Fall Symposium*. (1997) 640–4
8. Sioutos, N., de Coronado, S., Haber, M.W., Hartel, F.W., Shaiu, W.L., Wright, L.W.: Nci thesaurus: a semantic model integrating cancer-related clinical and molecular information. *J Biomed Inform* **40**(1) (2007) 30–43
9. The Gene Ontology Consortium: Creating the gene ontology resource: Design and implementation. *Genome Res* **11** (2001) 1425–1433
10. OBO: Open biomedical ontologies [<http://obofoundry.org>]. (2006)
11. Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A., Rosse, C.: Relations in biomedical ontologies. *Genome Biology* **6**(5) (2005) r46
12. Abdelmoty, A.I., Smart, P.D., Jones, C.B., Fu, G., Finch, D.: A critical evaluation of ontology languages for geographic information retrieval on the internet. *Journal of Visual Languages & Computing* **16**(4) (2005) 331–358
13. Agarwal, P.: Ontological considerations in giscience. *International Journal of Geographical Information Science* **19**(5) (2005) 501–536
14. Grenon, P., Smith, B.: SNAP and SPAN: Towards dynamic spatial ontology. *Spatial Cognition and Computation* **4**(1) (2004) 69–103
15. Mark, D., Smith, B., Tversky, B.: Ontology and geographic objects: An empirical study of cognitive categorization. In Freksa, C., Mark, D., eds.: *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science*. Number 1661 in LNCS, Springer Verlag (1999) 283–298
16. Feng, C.C., Bittner, T., Flewelling, D.: Modeling surface hydrology concepts with endurance and perdurance. In Egenhofer, M.J., Freksa, C., Miller, H.J., eds.: *Proceedings of GI-Science 2004*. LNCS 3234, Springer-Verlag, Berlin, Heidelberg (2004) 67–80
17. Sorokine, A., Bittner, T.: Understanding taxonomies of ecosystems: a case study. In Fisher, P., ed.: *Developments in Spatial Data Handling*, Springer Verlag, Berlin (2005) 559–572
18. Sorokine, A., Bittner, T., Renscher, C.: Ontological investigation of ecosystem hierarchies and formal theory for multiscale ecosystem classifications. *geoinformatica* **10**(3) (2006) 313–335
19. Bailey, R.G.: Delineation of ecosystem regions. *Environmental Management* **7** (1983) 365–373
20. Lowe, E.J.: Recent advances in metaphysics (keynote). In Welty, C., Smith, B., eds.: *International Conference on Formal Ontology in Information Systems*. (2001)
21. Smith, B., Koehler, J., Kumar, A.: On the application or formal principles to life science data: a case study in the gene ontology. In Rahm, E., ed.: *Data Integration in the Life Sciences*. Volume 2994 of LNBI., Springer Verlag (2004) 79–94
22. Donnelly, M., Bittner, T., Rosse, C.: A formal theory for spatial representation and reasoning in bio-medical ontologies. *Artificial Intelligence in Medicine* **36**(1) (2006) 1–27
23. Bittner, T., Donnelly, M.: Logical properties of foundational relations in bio-ontologies. *Artificial Intelligence in Medicine* **39** (2007) 197–216
24. Koepfen, W.: *Grundriss der Klimakunde*. W. de Gruyter, Berlin (1931)
25. Smith, B.: Fiat objects. *Topoi* **20**(2) (2001) 131–48