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. We present a methodology of how to use a top-level ontology to create a domain ontology from existing scientific texts by (1) identifying informal definitions of domain-specific terms, (2) substituting terms refering to top-level relations by terms of the top-level ontology, and (3) refining the definitions of the domain-specific terms by taking into account additional vocabulary provided by the top-level ontology. We demonstrate this methodology by applying it to Bailey's paper on 'Delineation of Ecoregions'.

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 we use as an example the domain of ecosystem classification and delineation. Other domains include hydrology and environmental science, as well as in 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 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]. Recent research has shown that robust domain ontologies must be [1, 4]:

1. based on a well designed *top-level ontology*;
2. developed rigorously using formal logic.

Robust domain ontologies use top-level ontologies as their foundation. This means that the semantics of the domain vocabulary is specified 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 used in many different domains. 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

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	universal-part-of	uUP, dUP
collection	universal	class-extension-of	Ext
collection	collection	sub-collection-of	$\subseteq$
collection	collection	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].)

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, 6] 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 medical domain ontologies are GALEN [7], SNOMED(CT) [8], and the UMLS [9]. An example of a domain ontology for biomedicine and microbiology is the description logic based version of the GeneOntology [10]. Currently efforts are made to provide a top-level ontology as the unifying ontological and formal basis for all those bio-medical ontologies [11].

Unfortunately there are only preliminary attempts to provide logic-based domain ontologies within the geo-domains [12, 13]. Examples are in [14, 15] for general ontologies of geographic categories, in [16, 17] for domain ontologies for ecosystems, and in [18] for a domain ontology for hydrology.

This paper builds upon [16, 17]. It uses the example of ecosystem classification and delineation to demonstrate how the consistent usage of the top-level terms in Table 1 according to the semantics formally specified in a logic-based top-level ontology helps (a) to improve the preciseness of definitions in scientific discourses and (b) to build robust domain ontologies. The reminder of this paper is structured as follows. In Section 2 we give a simplified version of a logic-based ontology for the top-level terms in Table 1. In Section 3 present our methodology of how to build domain ontologies based on a top-level ontology and existing scientific texts. In Section 4 we apply this methodology to the domain of ecosystem classification and ecoregion delineation by using the top-level ontology to improve the definitions of domain specific terms presented in [19].

2 A simple top-level ontology

Following [2, 6], we distinguish three disjoint sorts of entities: (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.

We present the theory in a sorted first-order predicate logic with identity. We use the letters $w, x, x_1, y, z, \ldots$ as variables ranging over (endurant) individuals; c, d, e, g as variables ranging over universals; and $p, q, r, p_1, \ldots$ 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). We use the symbol $\equiv$ for definitions. We write (x) to symbolise universal quantification (for all x ...) and $(\exists x)$ to symbolise existential quantification (there is at least one x ...). 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 [6].

2.1 Mereology of individuals

Individual-part-of is a relation that holds 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. We write $IP\ xy$ to signify that individual x is part of individual y.

We define that 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.

We add the standard axioms requiring that individual parthood is reflexive (AM1), antisymmetric (AM2), transitive (AM3). We also require that if every z that overlaps x also overlaps y then x is part of y (AM4).

$AM1\ IP\ xx$

$AM2\ IP\ xy \wedge IP\ yx \rightarrow x = y$

$AM3\ IP\ xy \wedge IP\ yz \rightarrow IP\ xz$

$AM4\ (z)(O\ zx \rightarrow O\ zy) \rightarrow IP\ xy$

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.

We use 'ϵ' to stand for the member-of relation between individuals and collections. We use the notation $\{x_1, \ldots, x_n\}$ to refer to a finite collection having $x_1, \ldots, x_n$ as members. We require that 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). In addition we require that the following collections exist: for every x there is a collection having x as its only member (AC3); the union of two collection always exists (AC4).

$$AC1\ (\exists x)(x \,\epsilon\, p) \qquad AC3\ (\exists p)(p = \{x\})$$
$$AC2\ p = q \leftrightarrow (x)(x \,\epsilon\, p \leftrightarrow x \,\epsilon\, q) \qquad AC4\ (\exists r)(x)(x \,\epsilon\, r \leftrightarrow x \,\epsilon\, p \vee x \,\epsilon\, q)$$

We define: 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 as its only members is discrete. The USA is the sum of the collection which has the federal states of the USA as members. The collection of federal states also partitions the USA.

Collection p is *upwards partionomically 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 partionomically 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 p be the collection which has all the counties[1] of the USA as its members and let q be the collection that has all the federal states of the USA as its members. Then p is up- and downwards partionomically included in q: every member of p (a county) is part of some member of q (a federal state) and every member of q (a federal state) has some member of p (a county) as its part.

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

2.3 Universals, instantiation, and universal parthood

We use variables c, d, e, g ranging over 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*. We write *Inst xc* to signify that the individual x instantiates the universal c. We 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}).

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

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.

We require that every universal has an instance (AU1). (From this then immediately follows that c is a sub-universal of d if and only the extension of c is a sub-collection of the extension of d.) We also require: there is maximal universal (AU2); if two universals share a common instance then one is a sub-universal of the other (AU3); and if d has a proper sub-universal then all instances of d are also instances of proper sub-universals of d (A4).

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

Universals are here assumed to form tree-like hierarchies ordered by the sub-universal relation. In the scientific realm this tree-like structure is most closely resembled by 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 differentia, which tells us 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 [20, 21]. Differentia need to be such that the immediate sub-universals of a given universal are jointly exhaustive and pair-wise 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 we introduce the relations of upward and downward universal parthood, *uUP* and *dUP*, between universals: 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}).

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

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*.

We call the formal theory presented in this section TLO.

3 From top-level ontologies to domain ontologies

In the remainder of this paper we show how the *top-level* ontology, TLO, presented in Section 2 can be used to develop (a portion of) a *domain ontology* for ecosystem classification and ecoregion delineation. A domain ontology in most scientific domains will be based on existing scientific texts. In our example we develop a domain ontology based on Bailey's influential paper "Deliniation of ecosystem regions" [19]. We apply the following methodology:

1. *Substitute* terms used in established scientific definitions of domain-specific by terms of the top-level ontology
2. *Refine* the definitions of the domain-specific terms by taking into account additional vocabulary of the top-ontology
3. Create a formal representation of the resulting domain ontology

We will demonstrate this methodology in Section 4.[2] In the remainder of this section we will discuss some important distinctions between top-level ontologies and domain ontologies.

3.1 A 'model-theoretic' view of domain ontologies

Before we can begin to develop our ecosystem domain ontology we need to discuss an important distinction between the usage of top-level terms like 'sub-universal-of' in top-level ontologies and in domain ontologies.

The terms of a top-level ontology refer to *classes of relations* which satisfy the relevant axioms of the top-level ontology. For example, in TLO the term 'sub-universal-of' (abbreviated by $\sqsubseteq$) refers to the class of all relations which satisfy the axioms AU1-4, $D_{\sqsubseteq}$, and $D_{\sqsubset}$. In domain ontologies top-level terms often refer *specific relations*. 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.

[2] [22] applies a similar methodology to produce a formal description of RNA-structures using the RCC-theory and textbook definitions.

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 are more than one relations of a given type. We will see below in our ecosystem example, that there are three distinct relations of type sub-universal-of and three distinct relations of type universal-part-of in the domain of ecosystem classification.

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

3.2 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 we will also say that R is a binary relation on $\mathcal{D}(R)$. We 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 we can define the relations uR and dR on the powerset of $\mathcal{D}(R)$, i.e., $\mathcal{D}(uR) = \mathcal{D}(dR) = \mathcal{P}(\mathcal{D}(R))$, as follows:

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

R is called an *individual-level* relation on $\mathcal{D}(R)$ while uR and dR are *class-level* relations on $\mathcal{P}(\mathcal{D}(R))$ [23–25].

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). Moreover, since class-level relations are defined in terms of individual-level relations, the properties of individual-level relations determine the properties of the class-level relations. For example, if R is reflexive then uR and dR are reflexive and if R is transitive then uR and dR are transitive. (See [24] for details.)

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

Table 2. A selection of properties which a binary relation R with domain $\mathcal{D}(R)$ either has or lacks.

For any binary (individual-level or class-level) relation R, we can define the *immediate-R*-relation, R_i on $\mathcal{D}(R)$ in terms of R. (R_i may be the empty relation

[3] As the formal language for the domain ontology we use the language of set theory.

if $\mathcal{D}(R)$ is dense or R is the identity relation. For most non-empty R_i, R_i is intransitive.)

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

We will use R_i-relations in Section 4 to draw graphs of specific binary relations between individual ecoregions and ecosystem classes. These graphs serve as graphic representations of our 'formal' domain ontology. Consider Figure 1 which depicts the graph of the relation R_i such that the set of nodes of the graph is the domain of R_i and $(x,y) \in R_i$ if and only if there is an edge from node x to node y in the graph. If R_i is the relation depicted in the graphs of Figures 1–4, then R is the reflexive and transitive closure of R_i.

3.3 Types of relations

We can classify binary relations according to their logical properties. A *Partial ordering* is a type of relations which are reflexive, antisymmetric, and transitive. According to the top-level ontology of Section 2.1 *individual-part-of* is a type of relations that are reflexive, antisymmetric, transitive, and in addition satisfy axiom AM4 (hence individual-part-of is a sub-type of partial ordering).

In the remainder of this paper we use SMALL CAPITAL LETTERS to signify top-level terms for types of entities such as INDIVIDUAL and UNIVERSAL, and for types of relations such as INSTANCE-OF, 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. We use `typewriter font` and superscripts to distinguish specific relations in the domain ontology from relations-types denoted by top-level terms in the top-level ontology. I.e., we 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.

4 Ecosystem classification and delineation

We now apply the top-level ontology TLO from Section 2 and the methodology introduced in the beginning of the Section 3 to the problem of building a domain ontology for ecosystem classification and delineation. We thereby show how rather imprecise and ambiguous 'definitions' by Bailey [19] can be made more precise by (a) using the top-level terms listed in Table 1 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 particular we show how to explicitly distinguish a number of relations which have been confused and been taken to be a single relation before. As pointed out above, we use graphic rather than symbolic representations for those relations. The symbolic representations can be obtained as described in Section 3.2.

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

According to our methodology we start by analyzing domain-specific definitions from the scientific literature:

(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 we rephrase (BL1) as follows: Ecoregions are INDIVIDUALS that are INDIVIDUAL-PARTS-OF the surface of the Earth (an INDIVIDUAL) that are of geographic scale or larger.[4] Geographic ecosystems (ecosystems for short) are UNIVERSALS which are SUB-UNIVERSALS-OF the UNIVERSAL *ecosystem*. Ecosystem universals are INSTANTIATED BY ecoregions with similar functional characteristics. Every ecoregion has smaller ecoregions as INDIVIDUAL-PARTS all of which are INSTANCES-OF the UNIVERSAL *ecosystem*. Thus, according to our interpretation of (BL1) 'are' (in BL1) is intended to mean INSTANTNCE-OF, 'contain' is intended to mean INDIVIDUAL-PART-OF, '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) tells us that the *differentia* used for distinguishing ecosystem UNIVERSALS are environmental factors that create the variations between the ecoregions that INSTANTIATE distinct ecosystems. 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 [27].[5]

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 Figure 1 where the nodes are UNIVERSALS that form the domain of the relation $\texttt{sub-universal-of}^1$ and the directed edges represent the relation $\texttt{sub-universal-of}_i^1$ (in the sense of equation 2). The root of the tree is the ecosystem UNIVERSAL *geographic ecosystem*.

[4] To specify formally what 'being of geographic scale or larger' means we would need a theory of qualitative size relations like the one presented in [26].

[5] The classification of the quality universals such as *humid temperate climate* or *prairie climate* that serve as differentials, too, are in need of an ontological analysis. For the purpose of this paper, however, we take them as given.

The relevant differentia, which tell us 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 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 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 prairie climate group, by the prairie climate type, and by bushland vegetation. In other words: *Prairie Bushland Ecosystem* is a `sub-universal-of`^{1} *Prairie Ecosystem* which in turn is a `sub-universal-of`^{1} *Humid temperate domain* and thus, according to definition ($D_{\sqsubseteq}$), every INSTANCE-OF *Prairie Bushland Ecosystem* is an INSTANCE-OF *Prairie Ecosystem* and is also an INSTANCE-OF *Humid Temperate Ecosystem*.

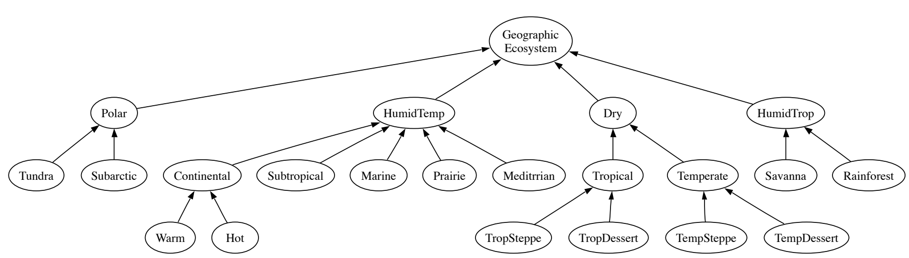


Fig. 1. Graphs of the relations `sub-universal-of`$_i^1$ and `upwards-universal-part-of`$_i^1$. Notice that in Bailey's classification all the leaf noded in the depicted graph have further sub-universals which are omitted here. See [19, p. 369] for a more complete tree.

[6] Bailey collapses Koeppen's subtropical and temperate climate groups into 'Humid temperate'.

4.2 Universal parthood relations between ecosystem universals

We can further refine (BL1) 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 Figure 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 Figure 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 Figure 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 our top-level ontology it is clear that these are two very different relations: `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]

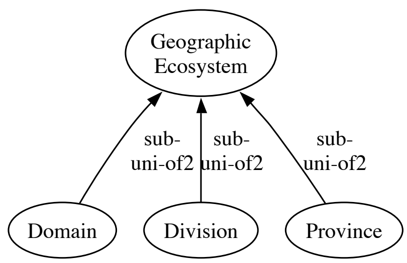
4.3 Classification of ecosystems according to kinds of climatic and vegetation characteristics

(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]

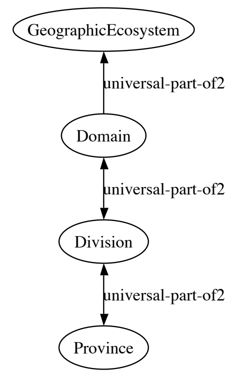
(BL3) tells us that, in addition to the classification of ecosystems according to *particular* climatic and vegetational affinities among the INSTANTIATING ecoregions (Figure 1), Bailey also classifies ecosystems according to the *kinds* of climatic and vegetation characteristics that characterize the INSTANTIATING ecoregions. Bailey [19] distinguishes the following (additional) geographic ecosystem

[7] This indicates some serious limitations of our extensional conception of relations as introduced in Section 3.2, which identifies the relations `sub-universal-of`^{1} and `upwards-universal-part-of`^{1}.

UNIVERSALS: *domains* are ecosystem UNIVERSALS, which INSTANTIATING ecoregions 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, Figure 1, p. 367].[8] The graph of the resulting relation $\texttt{sub-universal-of}_i^2$ is depicted in Figure 2(a), which shows ecosystem UNIVERSALS as nodes and the relation $\texttt{sub-universal-of}_i^2$ as directed edges.



(a) Graph of the relation $\texttt{sub-universal-of}_i^2$.



(b) Graphs of the relations (up- and downwards) $\texttt{universal-part-of}_i^2$.

Fig. 2. Relations among ecosystem universals which instances are charactrerized 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 maximal[9], 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 $\texttt{upwards-universal-part-of}^2$ be

[8] In this paper we omit *sections*.

[9] 'Maximal' here may mean different things: maximally singly-connected individual, maximal (but not necessarily singly-connected) part of a given continent, etc. Bailey seems to mean the latter. To clearly distinguish the possible interpretations we would have extend the top-level ontology by mereo-topological notions such as connectedness and self-connectedness. See, for example, [28] for details.

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 $\texttt{upwards-universal-part-of}_i^2$ is represented by the upwards arrows in Figure 2(b).

Between the universals *domain*, *division*, and *province* in addition the relation $\texttt{downwards-universal-part-of}^2$ 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 graph of the relation $\texttt{downwards-universal-part-of}_i^2$ is represented by the downwards arrows in Figure 2(b).

One can see that, in contrast to the graphs of $\texttt{sub-universal-of}_i^1$ and $\texttt{upwards-universal-part-of}_i^1$, the graphs of $\texttt{sub-universal-of}_i^2$ and $\texttt{upwards-universal-part-of}_i^2$ 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 [16] for a similar approach or 'intersecting' classification trees.)

Let $\texttt{sub-universal-of}^3$ 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 $\texttt{sub-universal-of}^3$ is, that *Geographic ecosystem* is the only UNIVERSAL that a has proper SUB-UNIVERSAL. Thus, the graph of $\texttt{sub-universal-of}_i^3$ is a flat but rather broad tree as indicated in Figure 3.

The graph of the relation $\texttt{sub-universal-of}_i^3$ is a *refinement* of the graph of the relation $\texttt{sub-universal-of}_i^2$ in Figure 2(a) in the sense that each of the nodes *domain*, *division*, and *province* in the graph of the relation $\texttt{sub-universal-of}_i^2$ is replaced by a set of jointly exhaustive and pairwise disjoint UNIVERSALS. For example, the node *Domain* in Figure 2(a) is replaced by the nodes *PolarDomain*, *Humid Temperate Domain*, *Dry domain*, and *Humid Tropical domain* in Figure 4. Similarly for the other nodes in Figure 2(a).

Let $\texttt{upwards-universal-part-of}^3$ be the UPWARDS-UNIVERSAL-PART-OF relation on the domain of $\texttt{sub-universal-of}^3$. The graph of $\texttt{upwards-universal-part-of}_i^3$ (Figure 4) is a refinement of the graph of $\texttt{upwards-universal-part-of}_i^2$

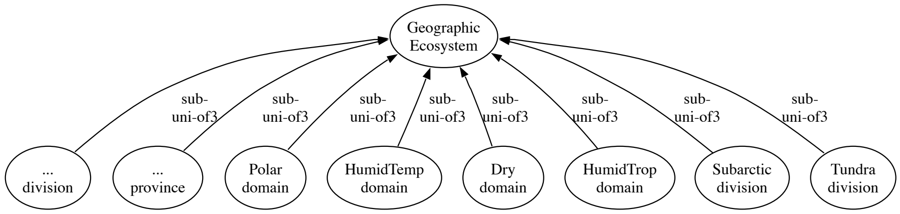


Fig. 3. Graph of the relation `sub-universal-of`^3_i.

(Figure 2(b)) in the sense that each of the nodes *domain*, *division*, and *province* of the graph of `upwards-universal-part-of`^2_i is replaced by a set of universals as described in the context of the graph of `sub-universal-of`^3_i above.

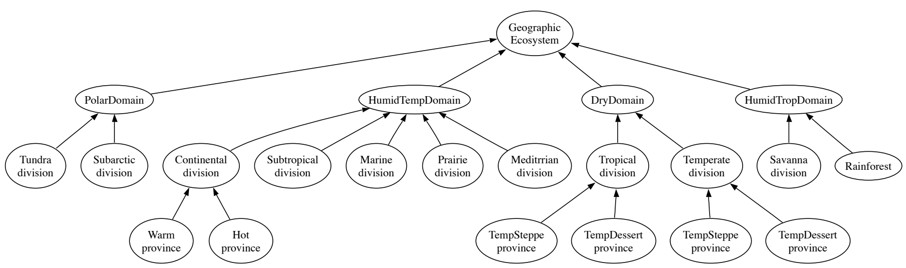


Fig. 4. Graph of the relation `universal-part-of`^3_i.

Notice, that there is a graph-isomorphism between the graphs of the relations `sub-universal-of`^1_i, `upwards-universal-part-of`^1_i, and `upwards-universal-part-of`^3_i, 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`^1_i and `sub-universal-of`^3_i are very different as easily recognizable in Figures 1 and 3.

4.5 Ecosystem delineation

So far we have 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 [29][10] that separate ecoregions INDIVIDUALS,

[10] Of course, these boundaries are subject to vagueness, as Bailey himself points out.

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 ecoregion which INSTANTIATES the UNIVERSAL *Dry domain* and the ecoregion which INSTANTIATES the UNIVERSAL *Humid temperate domain* in Figure 5. The former is an ecoregion which climate belongs to the dry climate group according to the Koeppen classification and the latter is an ecoregion which climate belongs to the humid temperate climate group. Since 'having instances which exclusively belong to climate group x and are of maximal size' is a differentia in the ecosystem classification, no ecoregion can belong to more than one climate group and every geographic ecoregion belongs to some climate group. In particular no two ecoregions, that INSTANTIATE the UNIVERSAL *domain* OVERLAP and there is no 'no-mans-land' between them.

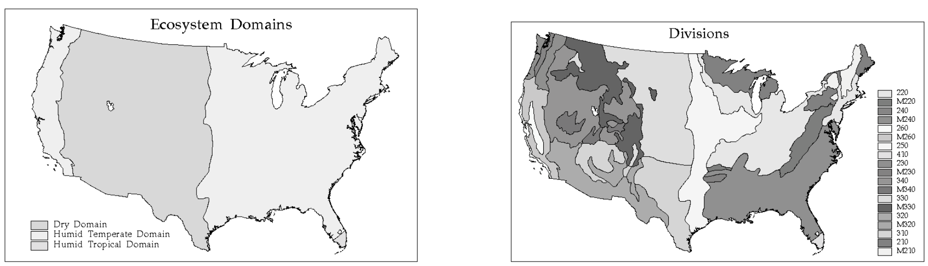
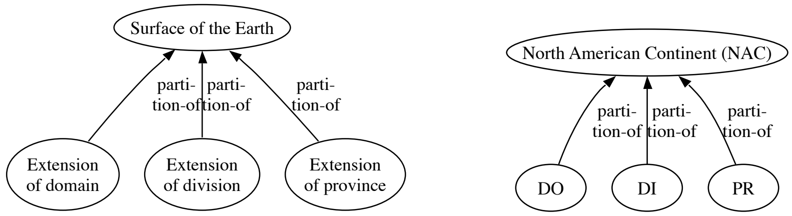


Fig. 5. Domain and division partitions of the USA [30, 31]

In general, the relation `sub-universal-of`[2] and its underlying domain of ecosystem universals are such that the EXTENSION of the UNIVERSAL *domain* (i.e., the COLLECTION of ecoregions that INSTANTIATE the UNIVERSAL *domain*) PARTITIONS 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 '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* PARTITION the INDIVIDUAL *Surface of Earth*. (Figures 5 and 6(a).)

4.6 Ecoregions in North America

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.



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

(b) Collections of geographic ecoregions that partition the North American Continent *NAC* .

Fig. 6. Partitions formed by collections of ecoregions.

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. For this purpose we use the notion of collection. 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 instantces-of the universal *division* and that are individual-part-of the North American Continent; and let *PR* be the collection of ecoregions that are instances-of the universal *province* and that are individual-part-of the North American Continent. 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` is depicted in Figure 6(b).

The collections *DO*, *DI*, and *PR* not only partition *NAC* they are also 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*.

5 Conclusions

We presented a methodology of how to use a top-level ontology to create a domain ontology from existing scientific texts by (1) identifying informal definitions of domain-specific terms, (2) substituting terms refering to top-level relations, by terms of the top-level ontology, and (3) refining the definitions of the domain-specific terms by taking into account additional vocabulary provided by the top-level ontology. We demonstrated this methodology by applying it to definitions extracted from Bailey's 'Delineation of Ecoregions' [19].

Notice that we do not claim 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 and can be interpreted in different ways, and thus leave 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 our aim to make as precise as possible how we understand Bailey by introducing 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 our interpretation.

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. Bittner, T., Donnelly, M., Smith, B.: A spatio-temporal top-level ontology for geographic information processing. Department of Philosophy, SUNY Buffalo (2007)
7. 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)
8. 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
9. Bodenreider, O.: The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Res (2004)
10. The Gene Ontology Consortium: Creating the gene ontology resource: Design and implementation. Genome Res **11** (2001) 1425–1433
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. Gnome 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 Lecture Notes in Computer Science, Springer Verlag (1999) 283–298
16. 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
17. Sorokine, A., Bittner, T., Renschler, C.: Ontological investigation of ecosystem hierarchies and formal theory for multiscale ecosystem classifications. In: Proceedings of GI-Science 2004. (2004)
18. 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
19. Bailey, R.G.: Delineation of ecosystem regions. Environmental Management **7** (1983) 365–373
20. 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 Lecture Notes in Bioinformatics., Springer Verlag (2004) 79–94
21. Sorokine, A., Bittner, T., Renscher, C.: Ontological investigation of ecosystem hierarchies and formal theory for multiscale ecosystem classifications. geoinformatica **10**(3) (2006) 313–335
22. Cohn, A.G.: Formalising bio-spatial knowledge. In Welty, C., Smith, B., eds.: Proc. 2nd International Conference on Formal Ontology in Information Systems (FOIS'01), ACM (2001) 198–209
23. Smith, B., Rosse, C.: The role of foundational relations in the alignment of biomedical ontologies. In Fieschi, M., Coiera, E., Li, Y.J., eds.: Proceedings of the 11th World Congress on Medical Informatics. (2004) 444–448
24. Donnelly, M., Bittner, T.: Spatial relations between classes of individuals. In Mark, D., Cohn, T., eds.: Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science. International Conference (COSIT 2005). Number 3693 in Lecture Notes in Computer Science, Springer Verlag (2005) 182 – 199
25. 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
26. Bittner, T., Donnelly, M.: A formal theory of qualitative size and distance relations. Technical report, Department of Philosophy, SUNY Buffalo (2007)
27. Koeppen, W.: Grundriss der Klimakunde. W. de Gruyter, Berlin (1931)
28. Varzi, A.: Parts, wholes, and part-whole relations: The prospects of mereotopology. Data and Knowledge Engineering **20**(3) (1996) 259–86
29. Smith, B.: Fiat objects. Topoi **20**(2) (2001) 131–48
30. Bailey, R.G.: Description of the ecoregions of the united states. Technical report, USDA Forest Service (1995)
31. Bailey, R.G.: http://www.fs.fed.us/land/ecosysmgmt/ecoreg1_home.html (1995)