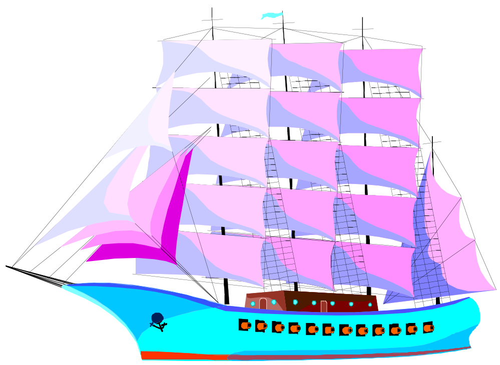
Ontologies

What is an Ontology?
What is a Good Ontology?
Rigidity, Identity, Unity, Dependence



Ontology in Philosophy

- Encarta: ***Metaphysics***, branch of philosophy concerned with the nature of ultimate reality. Metaphysics is customarily divided into:
  - ✓ ***Ontology***, questions how many fundamentally distinct sorts of entities compose the universe, and
  - ✓ ***Metaphysics proper***, describes the traits of reality.
- These traits together define reality and characterize any universe.
- Ontology, by contrast, investigates the ultimate divisions within our universe, and is more closely related to the physical world of human experience.

What is an Ontology?

- A characterization of the intended subject matter (class of applications) for a given conceptual model.
- We could do this in terms of **standard vocabularies**, which characterize what you can talk about, given a particular conceptual model.
- But then two conceptual models which talk about the Euclidean space in terms of different coordinate systems would be different…
- For example, we might have an ontology of time which treats time as a continuous line of (time) points (or, discrete, branching or other)

An Ontology is not a Vocabulary!

- For example,
  - ✓ $Vocabulary_1$={**Ent**,**Rel**}, where R ∈ **Rel** iff
    $R= D_1 x D_2 x \ldots x D_n$ where $D_i \in$ (**Ent** ∪ **Rel**)
  - ✓ $Vocabulary_2$={**Thing**}, where T ∈ **Thing** iff $T = D_i$
    or $T = T_1 \, x \, T_2 \, x \ldots x \, T_n$ where $T_i \in$ **Thing**

  model exactly the same aspects of an application.
- For a given ontology, one may define one or more **standard vocabularies** (**reference models**), in terms of which one can talk about the intended subject matter.
- For example, to talk about **Time**, there are two standard vocabularies, one involving points {**Point**, {**<**, **=**, **>**}}, and the other intervals {**Interval**, **TempRelationship**), with 13 possible temporal relationships.

...and it's not a Conceptual Schema

- There may be a university ontology which define concepts such as student, course, degree etc.
- A conceptual schema -- say, for the UofT student information system -- may use these concepts; but they are specialized in meaning: for example the student concept may be meant to have as instances 2003-2004 UofT students only.
- An ontology is ***not*** application- or system-specific, but a conceptual schema is. An ontology is meant to be reusable, a conceptual schema is less so.

Ontologies in AI and Beyond

- An ***ontology*** in AI means a collection of concept definitions: "...In the context of knowledge sharing, I use the term ontology to mean a specification of a conceptualization. That is, an ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist… it is certainly a different sense of the word than its use in philosophy…" [Gruber91].
- ***Ontologies*** are used in AI to facilitate "knowledge sharing" among agents. Ontologies have a big role to play in data integration as well.

Ontologies in Practice

- Netscape was developing a "world ontology" (taxonomy) to classify web sources. The taxonomy was similar to Yahoo's, however:
  - ✓ The taxonomy and its instances were supposed to be public;
  - ✓ It was specified in RDF;
  - ✓ Netscape was asking for volunteers to serve as editors for entries…

Where Do Ontologies Come From?

- Philosophy, recently Lexicography, Linguistics, some areas within Artificial Intelligence interested in the subject.
- There have been attempts to define general ontologies, e.g., [Bunge77], or use these in Information Systems Engineering [Wand90] to evaluate the adequacy of particular information models.
- One could define more coarse- or fine-grain ontologies (**upper-level** and **lower-level**).

Levels of Ontological Depth

- **Lexicon** -- a vocabulary with NL definitions.
- **Simple Taxonomy** -- captures taxonomic relationships.
- **Thesaurus** -- taxonomy plus related-terms; captures synonymy, homonymy, etc.
- **Relational Model** -- Unconstrained use of arbitrary relations
- **Fully Axiomatized Theory** -- universal, ontologically neutral language; can specify/characterize fully a conceptualization.

Engineering "Good" Ontologies

- What distinguishes a good ontology from bad ones? [Welty01] addresses this question for taxonomic ontologies (i.e., class diagrams or subsumption hierarchies).
- It is assumed that the taxonomy is defined by assertions of the form

  $\forall x[P(x) \Rightarrow Q(x)]$ (e.g., $\forall x[Student(x) \Rightarrow Person(x)]$)
- In such cases, we'll say that Q **subsumes** P.

Meta-Properties of Ontologies

- *We'll analyze taxonomic ontologies in terms of four fundamental meta-properties (i.e., properties that hold or not for properties such as Student and Person):*
  - ✓ *rigidity -- once a dog, always a dog!*
  - ✓ *identity -- is this Chris' dog?*
  - ✓ *unity -- is the collar part of Chris' dog?*
  - ✓ *dependence*

Notation and Terminology

- *We assume that properties (Person, Student,...) are predicates in a first order language L_0. Predicates generally have an extra time argument, e.g., P(x,t) means that x has property P at time t; for time invariant properties we will just write P(x).*
- *Meta-properties are predicates in another first order language, L_1, which has predicates of L_0 as constants; if P is an L_0 predicate and **R** an L_1 predicate, we will write P^{+R} if P has property **R**, and P^{-R}, $P^{\sim R}$ if it doesn't (...negation comes in two flavours...)*
- *"∀x" means "for all x in all possible worlds"; this means that the property Unicorn is not empty even though no unicorn exists (…in this world!).*
- *We write E(x,t) to indicate that x actually exists at time t.*

Rigidity

- A property P is ***rigid*** (denoted **+R**)if it is essential to all its instances: $\forall x[P(x) \Rightarrow \quad P(x)]$; this says something like "For all x, if P(x), then necessarily P(x)"; you can also interpret to mean "in all possible worlds."
- For example, *Person* is rigid (*Person*$^{+R}$)., meaning that if someone is a person, she can't cease to be one.
- A property is ***non-rigid*** (**-R**) if it is not essential to some of its instances: $\exists x[P(x) \Rightarrow \neg \quad P(x)]$
- A property is ***anti-rigid*** (**~R**) if it is not necessary for all its instances: $\forall x[P(x) \Rightarrow \neg \quad P(x)]$
- *Student* is non-rigid (*Student*$^{-R}$), which assumes some "students for life", while *President* is (…should be!) anti-rigid (*President*$^{\sim R}$).

Identity

- Identity for a class means being able to distinguish between different instances by having a characteristic property.
- An ***identity condition*** (IC) is a formula Σ (sameness condition) such that either

  $E(x,t) \wedge P(x,t) \wedge E(y,t') \wedge P(y,t') \wedge x=y \Rightarrow \Sigma(x,y,t,t')$

  or $E(x,t) \wedge P(x,t) \wedge E(y,t') \wedge P(y,t') \wedge \Sigma(x,y,t,t') \Rightarrow x=y$

  in the first case, Σ is a ***necessary*** IC, in the second it is a ***sufficient*** IC.
- For example, say that SI#(x) = SI#(y) is an IC for Person; (...in an ideal world) we can treat this as a sufficient IC.
- In a less ideal world, we can treat it as a necessary IC.

Sortals: Properties with Identity

- A property P **supplies** an IC iff (i) it is rigid, (ii) there is a necessary or sufficient IC for it, and (iii) The same IC is not carried by any of the properties subsuming P.
- This means that, if P inherits different (but compatible) ICs from multiple properties, it still supplies an IC.
- A property P **carries** an IC iff it is subsumed by a property supplying that IC.
- A property carrying an IC has meta-property **I** (**-I** otherwise.) A property supplying an IC has meta-property **O** ('owns" an IC).
- It follows from these definitions that *Person*$^{+O}$, while *Student*$^{+I}$.

Local Identity

- Global ICs are rigid properties.
- A **local IC** identifies instances of P only when they are instances of P.
- For example, "same-wing-pattern" is an IC (necessary and sufficient, in fact) for butterflies, but only when an entity is a butterfly, not when **that entity** is a caterpillar.
- Likewise, same-registration# for students is a sufficient condition, but only while someone is a student.
- Local IC (**+L**) are non-rigid properties.
- A global IC identifies an entity for its entire existence (only for +R properties.)

Unity

- Determines whether a property has a notion of "wholeness", a criterion for determining the boundary of an instance of that property. Related to the *PartOf* relationship.
- PartOf(x,y) means "x is part of y".
- We assume that PartOf satisfies the following axioms:
  - ✓ $PPartOf(x,y,t) =_{def} PartOf(x,y,t) \wedge \neg(x = y)$ (proper part)
  - ✓ $O(x,y,t) =_{def} \exists z(PartOf(z,x,t) \wedge PartOf(z,y,t))$ (overlap)
  - ✓ $PartOf(x,y,t) \Rightarrow E(x,t) \wedge E(y,t)$ (actual existence of parts)
  - ✓ $PartOf(x,y,t) \wedge PartOf(y,x,t) \Rightarrow x = y$ (antisymmetry)
  - ✓ $PartOf(x,y,t) \wedge PartOf(y,z,t) \Rightarrow PartOf(x,z,t)$ (transitivity)
  - ✓ $PPartOf(x,y,t) \Rightarrow \exists z(PPartOf(z,y,t) \wedge \neg O(z,x,t))$ (weak supplementation)

Wholes

- An object x is a ***whole under*** ω iff ω is an equivalence relation such that all the parts of x are linked by ω, and nothing else is linked by ω.
- Depending on the ontological nature of ω (generalized connection), we distinguish three main kinds of unity for concrete (i.e., spatio-temporal) entities:
  - ✓ ***Topological unity***: ω is topological/physical connection, e.g., parts of a coal piece, or apple.
  - ✓ ***Morphological unity***: ω is combination of topological unity and shape, such as a ball, or a morphological relation between wholes such as a constellation.
  - ✓ ***Functional unity***: ω is combination of other kinds of unity with some notion of purpose (e.g., a hammer), or a functional relation between wholes such as a bikini.

Having the Unity Meta-Property

- *A property P carries a **unity condition** (or UC) iff there exists a single equivalence relation ω such that each instance of P is a whole under ω. We write P^{+U}.*
- *A **non-unity** property (**-U**) doesn't have a single UC that applies for all its instances, but has some UC for each one of its instances.*
- *An **anti-unity** property (**~U**) doesn't have a UC for any of its instances.*
- *LegalEntity is an example of a non-unity property. Its instances (people or corporations) have different UCs.*
- *AmountOfMatter is an anti-unity property, since none of its instances can be wholes.*

(External) Dependence

- A property P depends on another Q if Q is not a part or a constituent of P, and for every instance of P there is necessarily an instance of Q.
- For example, *Parent* depends on *Child* (*Parent*$^{+D}$), since one has to have a child to be a parent, and moreover *Child* is not a part or constituent of *Parent*.
- More precisely, a property P is ***externally dependent*** on a property Q if, for all its instances x of P, necessarily some instance y of Q must exist, which is not a part nor a constituent of x:

$$\forall x \quad [P(x) \Rightarrow \exists y(Q(y) \land \neg PartOf(y,x) \land \neg Const(y,x))]$$

- Const(y,x) means that y is a constituent of x, e.g., bricks are a constituent of a house.

Subsumption Constraints

- It follows from these definitions that if P, Q are properties,
  - ✓ $P^{\sim R}$ can't subsume Q^{+R}
  - ✓ P^{+I} can't subsume Q^{-I}
  - ✓ P^{+U} can't subsume Q^{-U}
  - ✓ $P^{\sim U}$ can't subsume Q^{+U}
  - ✓ P^{+D} can't subsume Q^{-D}

Using the Framework

- Two well-known ontologies define:
  - ✓ Physical object isA amount of matter (WordNet)
  - ✓ Amount of matter isA Physical Object (Pangloss)
- Which one is correct?
- Analyze each
  - ✓ Physical-object: **+O+U+R-D**
  - ✓ Amount of matter: **+O~U+R-D**
- Result: According to a common sense understanding, both ontologies are wrong, each concept should be at the top-level.

Exposing Differences

- Let's say that two people characterize *Social Entity* as follows:
  - ✓ Person 1: Social Entity **+O+U+R-D**
  - ✓ Person 2: Social Entity **+O+U+R+D**
- Problem?
  - ✓ Person 1: A social entity is a group of people **who are together for some social reason**. Hence, **-D**.
  - ✓ Person 2: A social entity is an entity **recognized by society**, therefore **+D.**

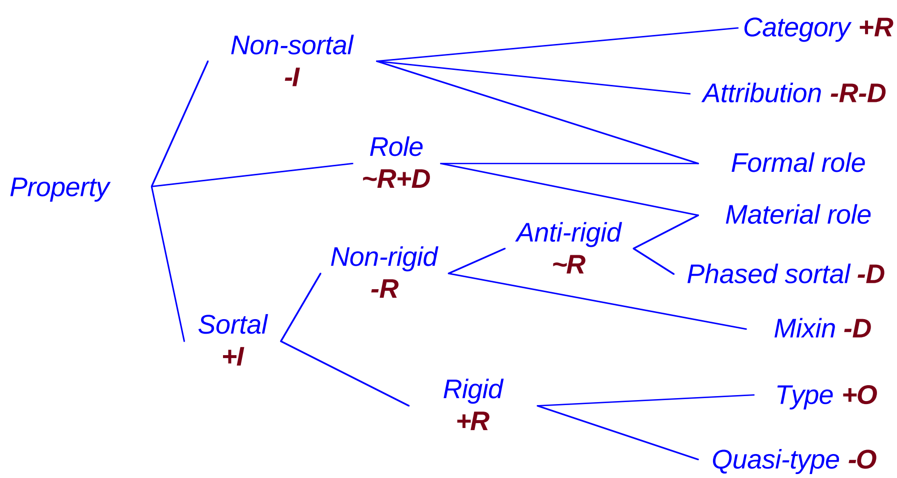
Sortals and Categories

- **Sortals** (Horse, Triangle, Amount of Matter, Person, Student...) -- carry identity, hardly *definable* in terms of a few primitives, high organizational utility.
- **Categories** (Universal, Particular, Event, Substance...) -- no identity, useful generalizations for sortals, characterized by a set of (only necessary) formal properties, good organizational utility
- Other non-sortals (red, big, decomposable, eatable, dependent, singular...) -- no identity, span across different sortals, limited organizational utility (but high semantic value)

Types and Roles

- *Types are rigid sortals that own an identity (**+I+R+O**). For example, Integer is a type.*
- *Roles are anti-rigid, depend on something else (**~R+D**).*
- *For example, to be a student there has to be an educational institution.*
- *Roles can be partitioned into **formal** and **material roles**.*
- *Formal roles don't have an identity (**-I**). For example, the property of being an agent is an example of a formal role and applies to many things from people to chemicals.*
- *Material roles do have an identity and are often subsumed by types.*

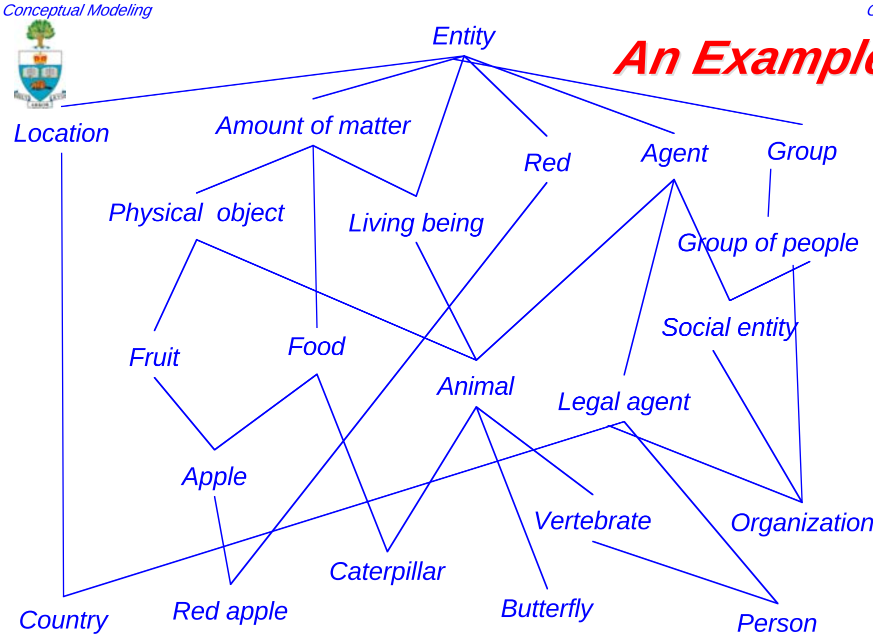
A Formal Ontology of Properties



Basic Property Kinds

O	I	R	D		
+	+	+	±	Type	Location
-	+	+	±	Quasi-type	Vertebrate
-	+	-	-	mixin	
-	+	~	+	Mat. role	
-	+	~	-	Phased sortal	Country
-	-	+	±	Category	Entity
-	-	~	+	Formal role	Agent
-	-	-	-	Attribution	Red

An Example



Property Analysis

Entity
- Everything is an entity
- **-I-U-D+R**
- Category

Location
- Generalized region of space.
- **+O**: by its parts (mereol. extensional).
- **~U**: no way to isolate location
- **-D+R**
- Type

Amount of Matter
- unstructured /scattered "stuff" as lumps of clay or some bricks
- **+O**: mereologically extensional
- **~U**: intrinsically non-unity
- **-D+R**
- Type

More Analysis

Red
- Really Red-thing, the set of all red-colored entities
- **-I-U-D-R**
- Formal Attribution

Group

An *unstructured* collection of wholes

+O: same-members

~U: unstructured, no unity.

-D+R

Type

Agent

An entity playing a part in some event

-I-U: no *universal* IC/UC

+D: on the event/action participating in

~R: no instance is necessarily an agent

Formal role

...More...

Butterfly

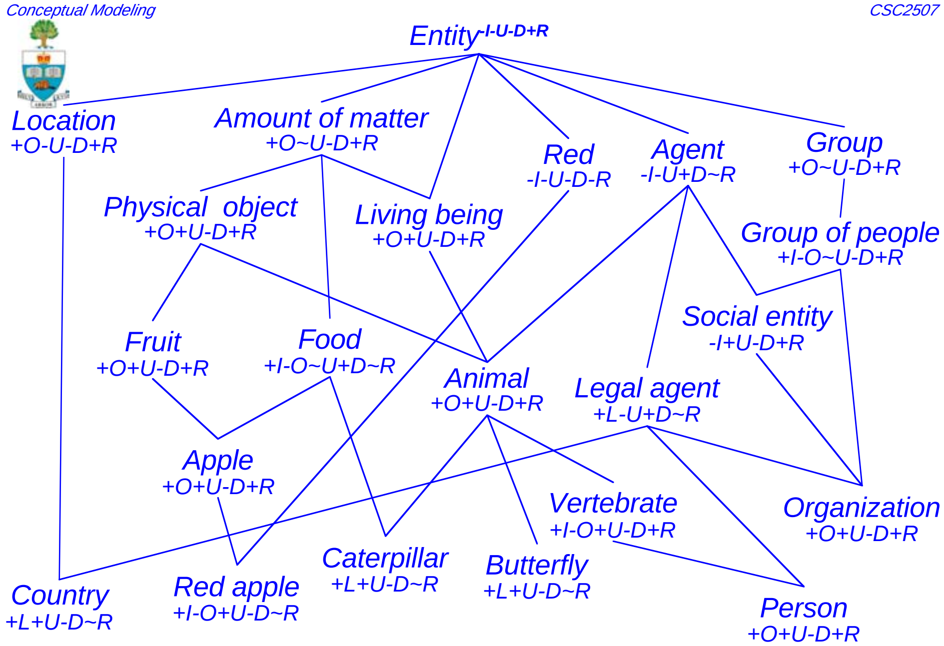
- **+L**: same-wing-pattern
- **-D+U**: biological
- **~R**: the *same entity* can be something else (a caterpillar)
- Phased sortal

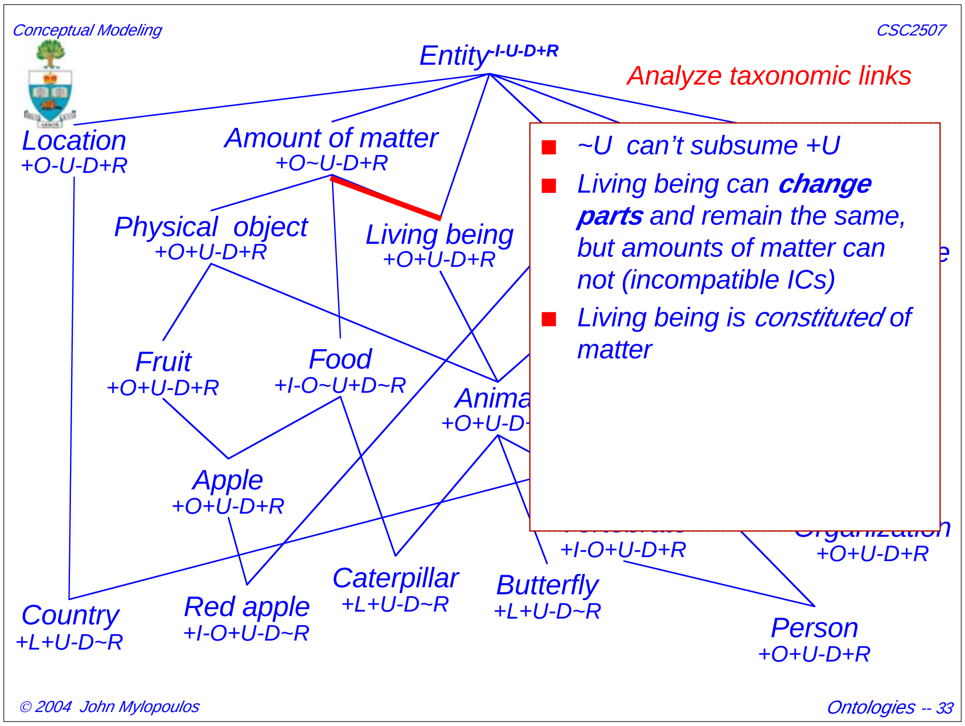
Caterpillar

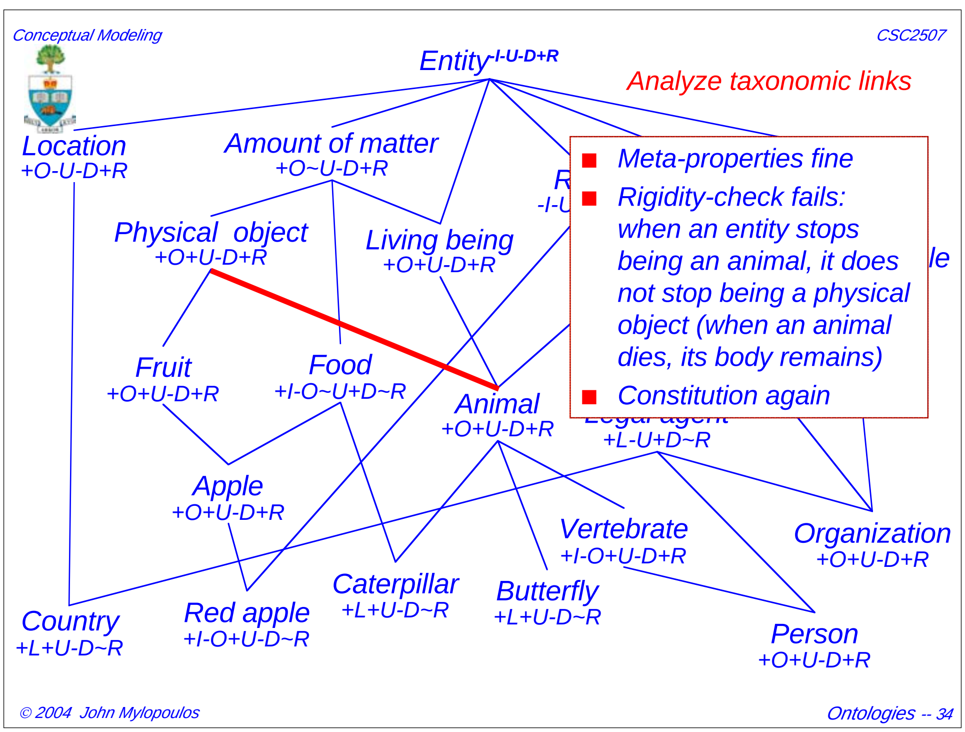
- **+L**: spots, legs, color
- **-D+U**: biological
- **~R**: caterpillars become butterflies, change their IC
- Phased sortal

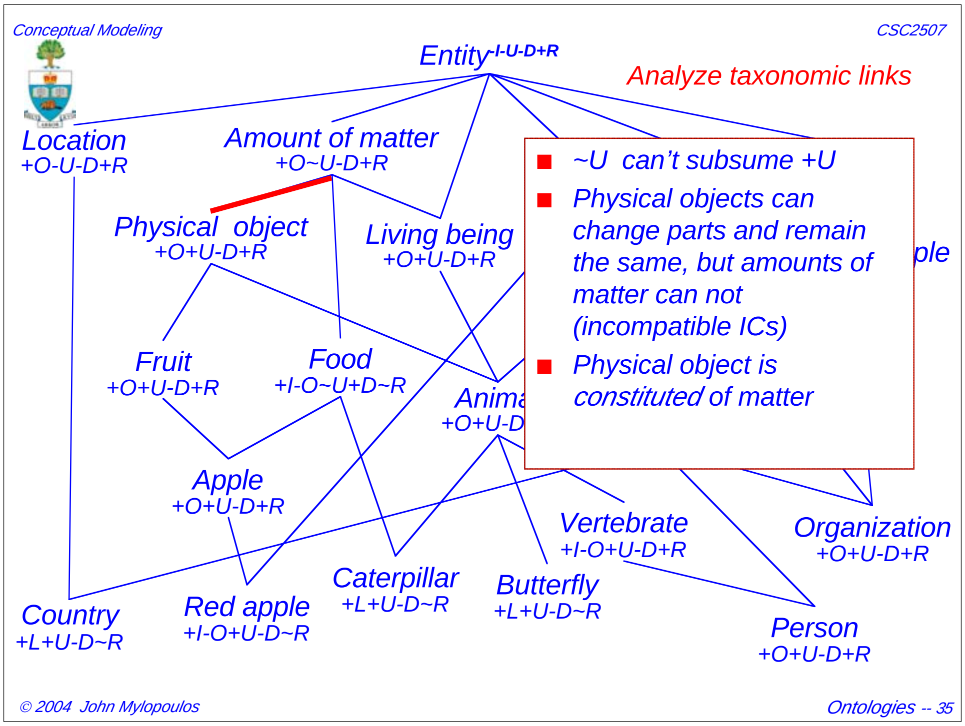
Country

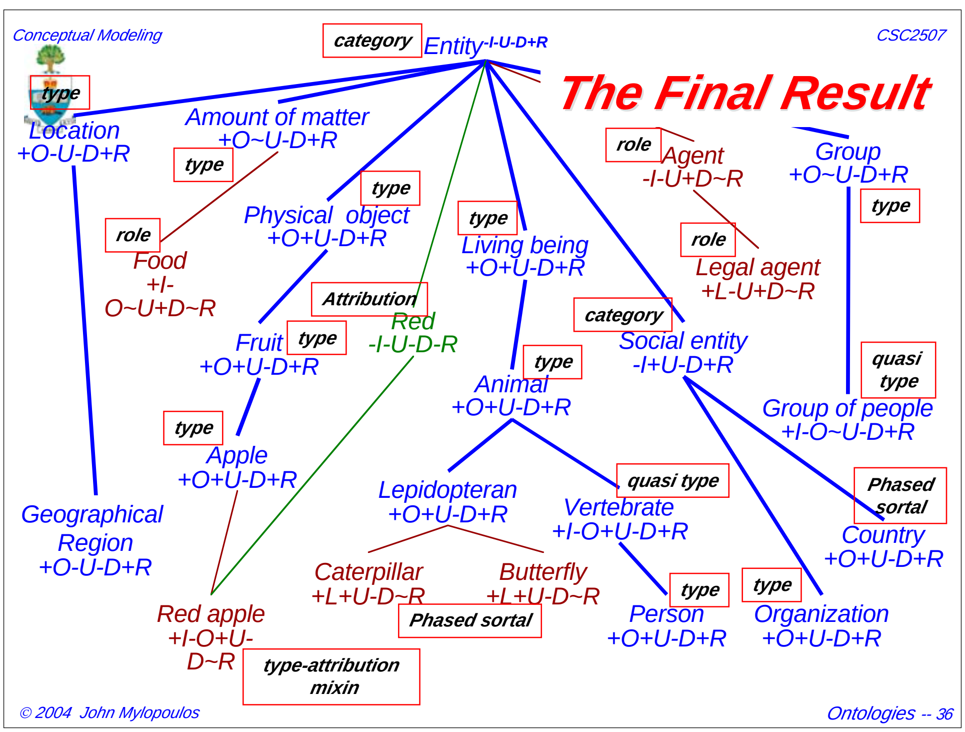
- A place recognized by convention as autonomous
- **+L**: government, sub-regions
- **+U**: countries are countable (heuristic)
- **-D**
- **~R**: some countries do not exist as countries any more (e.g. Prussia) but are still places
- Phased sortal



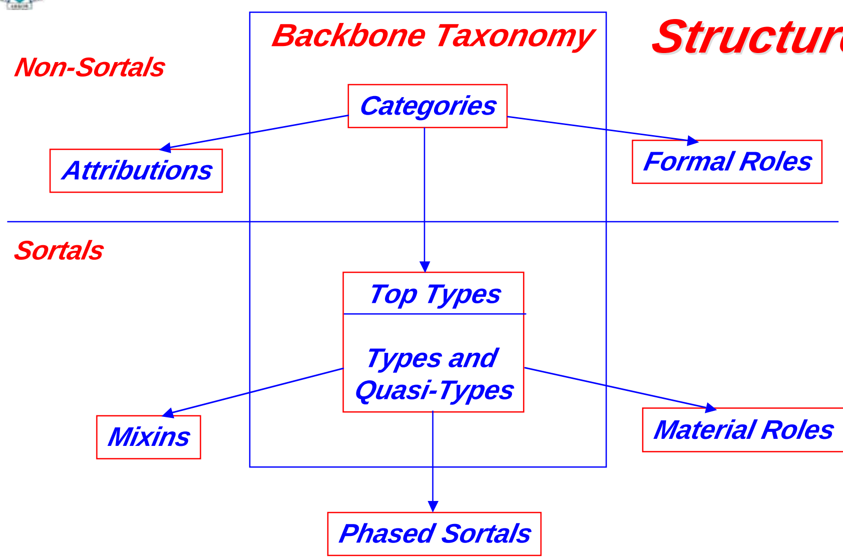

Conceptual Modeling
CSC2507
Entity-I-U-D+R
Analyze taxonomic links
Location
+O-U-D+R
Amount of matter
+O~U-D+R
Physical object
+O+U-D+R
Living being
+O+U-D+R
Fruit
+O+U-D+R
Food
+I-O~U+D~R
Apple
+O+U-D+R
Red apple
+I-O+U-D~R
Caterpillar
+L+U-D~R
Butterfly
+L+U-D~R
+I-O+U-D+R
+O+U-D+R
Country
+L+U-D~R
Person
+O+U-D+R
~U can't subsume +U
Living being can change parts and remain the same, but amounts of matter can not (incompatible ICs)
Living being is constituted of matter
© 2004 John Mylopoulos
Ontologies -- 33


Conceptual Modeling
CSC2507
Entity-I-U-D+R
Analyze taxonomic links
Location
+O-U-D+R
Amount of matter
+O~U-D+R
Physical object
+O+U-D+R
Living being
+O+U-D+R
Fruit
+O+U-D+R
Food
+I-O~U+D~R
Animal
+O+U-D+R
+L-U+D~R
Apple
+O+U-D+R
Vertebrate
+I-O+U-D+R
Organization
+O+U-D+R
Red apple
+I-O+U-D~R
Caterpillar
+L+U-D~R
Butterfly
+L+U-D~R
Country
+L+U-D~R
Person
+O+U-D+R
Meta-properties fine
Rigidity-check fails: when an entity stops being an animal, it does not stop being a physical object (when an animal dies, its body remains)
Constitution again
© 2004 John Mylopoulos
Ontologies -- 34


Conceptual Modeling
CSC2507
Entity -I-U-D+R
Analyze taxonomic links
Location
+O-U-D+R
Amount of matter
+O~U-D+R
Physical object
+O+U-D+R
Living being
+O+U-D+R
~U can't subsume +U
Physical objects can change parts and remain the same, but amounts of matter can not (incompatible ICs)
Physical object is constituted of matter
ple
Fruit
+O+U-D+R
Food
+I-O~U+D~R
Anima
+O+U-D
Apple
+O+U-D+R
Vertebrate
+I-O+U-D+R
Organization
+O+U-D+R
Caterpillar
+L+U-D~R
Butterfly
+L+U-D~R
Country
+L+U-D~R
Red apple
+I-O+U-D~R
Person
+O+U-D+R
© 2004 John Mylopoulos
Ontologies -- 35


Conceptual Modeling
CSC2507
category
Entity -I-U-D+R
The Final Result
type
Location
+O-U-D+R
Amount of matter
+O~U-D+R
type
role
Agent
-I-U+D~R
Group
+O~U-D+R
type
type
Physical object
+O+U-D+R
type
Living being
+O+U-D+R
role
Legal agent
+L-U+D~R
role
Food
+I-
O~U+D~R
Attribution
Red
-I-U-D-R
category
Social entity
-I+U-D+R
Fruit
type
+O+U-D+R
type
Animal
+O+U-D+R
quasi type
Group of people
+I-O~U-D+R
type
Apple
+O+U-D+R
Lepidopteran
+O+U-D+R
quasi type
Vertebrate
+I-O+U-D+R
Phased sortal
Country
+O+U-D+R
Geographical Region
+O-U-D+R
Caterpillar
+L+U-D~R
Butterfly
+L+U-D~R
Phased sortal
type
Person
+O+U-D+R
type
Organization
+O+U-D+R
Red apple
+I-O+U-
D~R
type-attribution mixin
© 2004 John Mylopoulos
Ontologies -- 36

Ideal Taxonomic Structure



References

- *[Bunge77] Bunge,M., Treatise on Basic Philosophy: Ontology I, The Furniture of the World, Reidel, 1977.*
- *[Gruber91] Gruber, T., "The Role of a Common Ontology in Achieving Sharable, Reusable Knowledge Bases," Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning, Cambridge, 601-602.*
- *[Guarino00] Guarino, N. and Welty, C., "Ontological Analysis of Taxonomic Relations," in A. Länder and V. Storey (eds.), Proceedings of ER-2000: The International Conference on Conceptual Modeling, Springer Verlag LNCS 1920.*
- *[Guarino01] Guarino, N. and Welty, C., "Conceptual Modeling and Ontological Analysis", tutorial notes, 2001.*
- *[Wand90] Wand, Y. and Weber, R., "An Ontological Model of an Information System", IEEE Transactions on Software Engineering 16, 1282-1292.*