Ontologies
An Introduction

(http://www.dfki.de/~horacek/knowl-onto.html)

Helmut Horacek

Saarland University / DFKI

Language Technology, old bldg. 1.25
Tel: 85775-2450
email:helmut.horacek@dfki.de

WHAT IS AN ONTOLOGY?

A branch of metaphysics concerned with the nature and relations of being

A particular theory about the nature of being or the kind of existents

Merriam Webster (http://m-w.com/home.htm)

Specification of a conceptualization

Tom Gruber

Important aspect - supports/regulates agreement about a set of terms

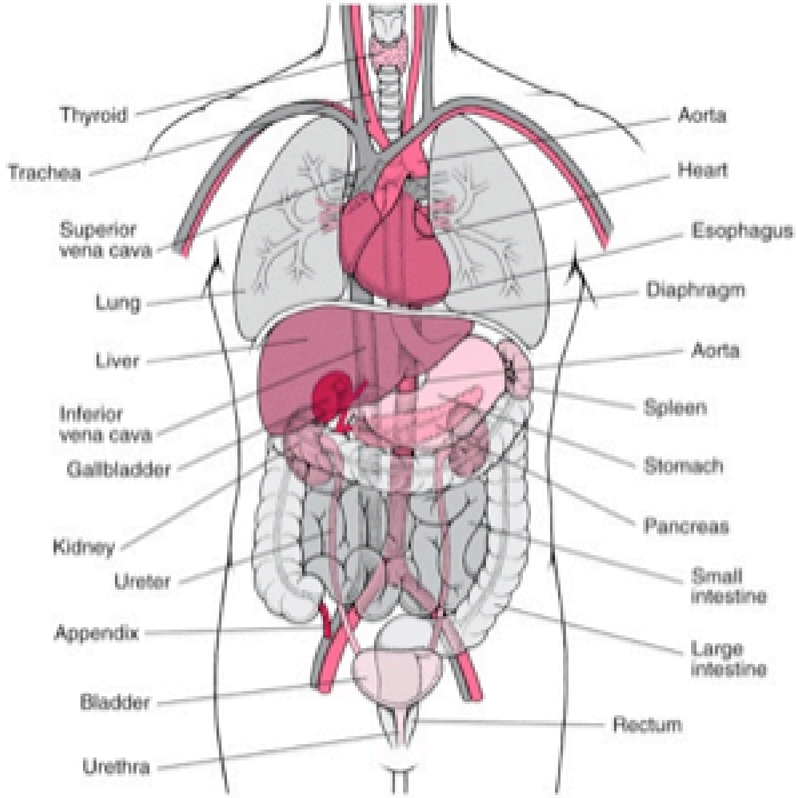
What is an Ontology?

A model of (some aspect of) the world

What is an Ontology?

A model of (some aspect of) the world

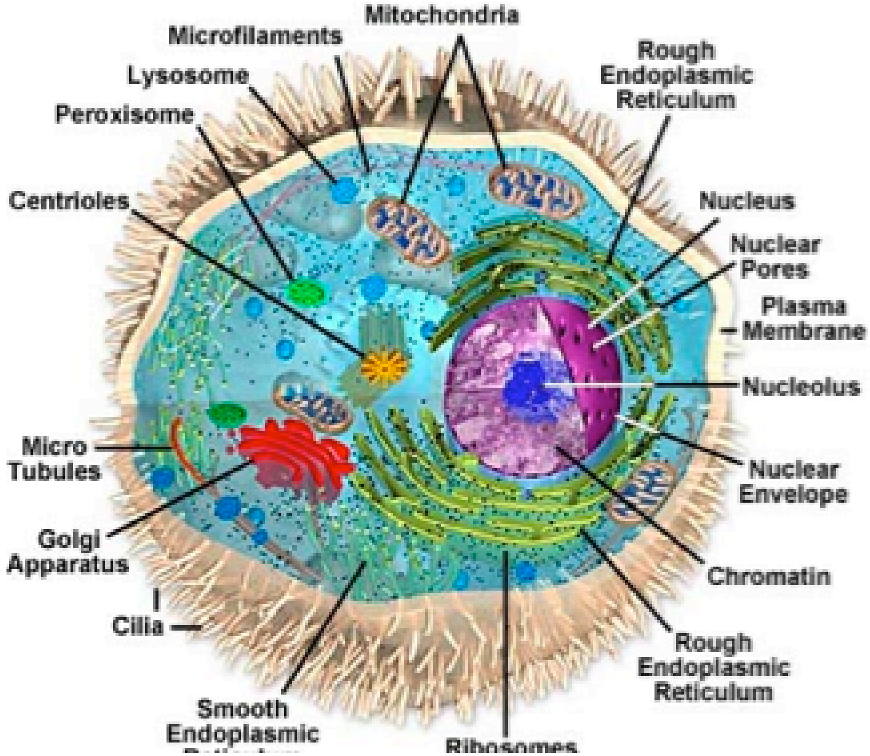
- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy



What is an Ontology?

A model of (some aspect of) the world

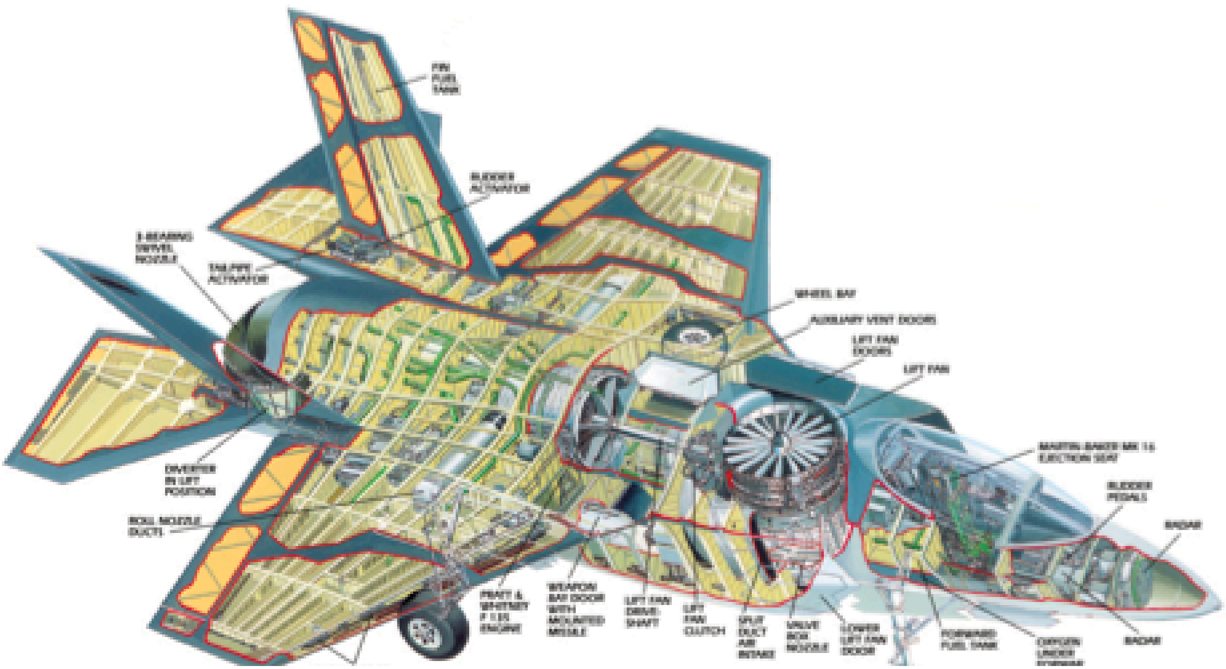
- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
  - Cellular biology



What is an Ontology?

A model of (some aspect of) the world

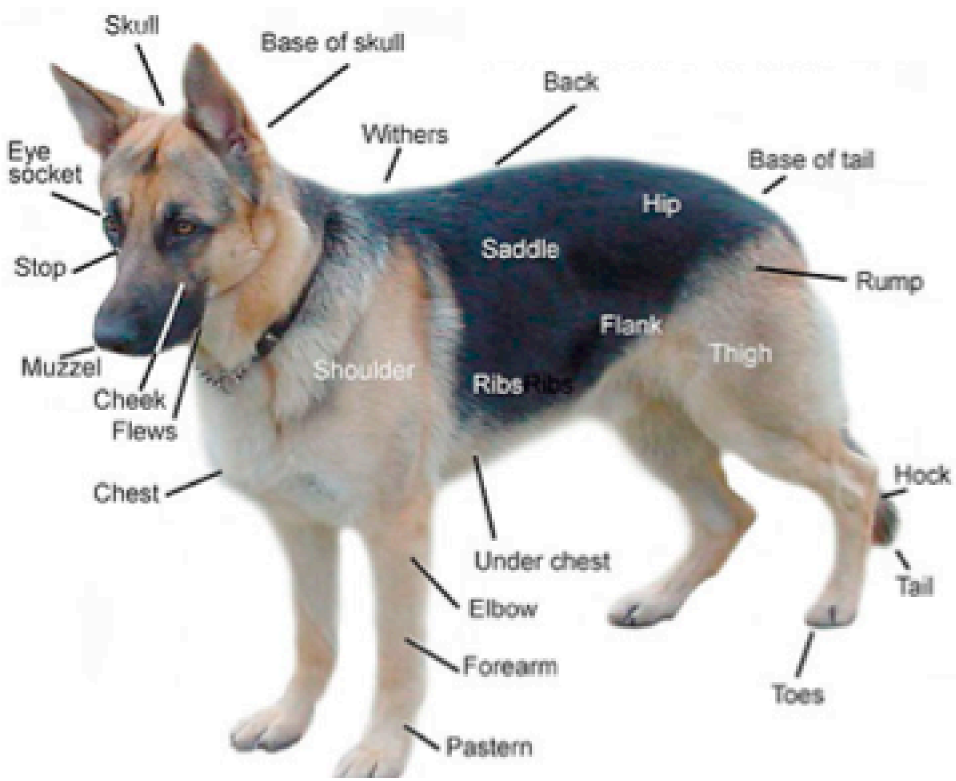
- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
  - Cellular biology
  - Aerospace



What is an Ontology?

A model of (some aspect of) the world

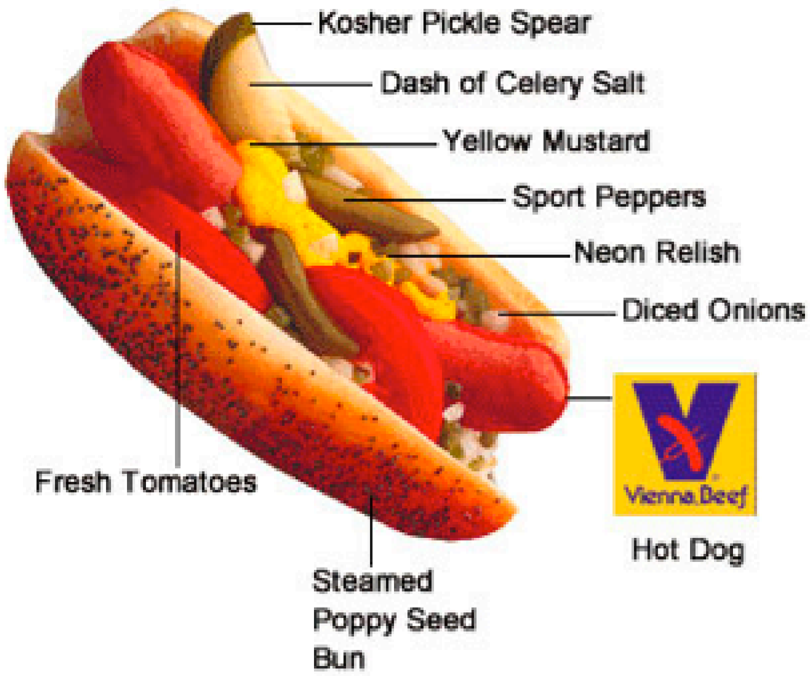
- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
  - Cellular biology
  - Aerospace
  - Dogs



What is an Ontology?

A model of (some aspect of) the world

- Introduces **vocabulary** relevant to domain, e.g.:
  - Anatomy
  - Cellular biology
  - Aerospace
  - Dogs
  - Hotdogs
  - ...

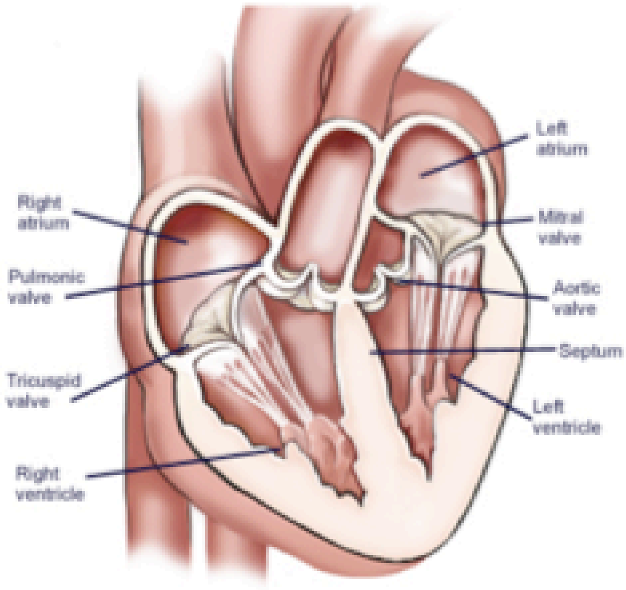


What is an Ontology?

A model of (some aspect of) the world

- Introduces **vocabulary** relevant to domain
- Specifies **meaning** (semantics) of terms

Heart is a muscular organ that
is part of the circulatory system



What is an Ontology?

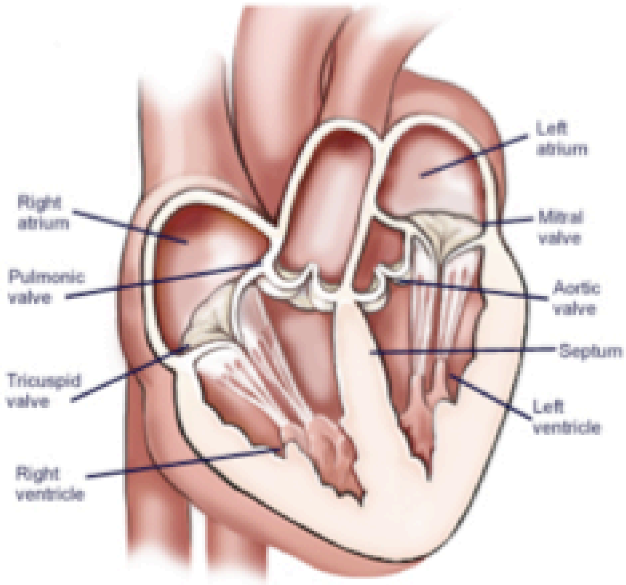
A model of (some aspect of) the world

- Introduces **vocabulary** relevant to domain
- Specifies **meaning** (semantics) of terms

  Heart is a muscular organ that
  is part of the circulatory system

- **Formalised** using suitable logic

$$\forall x.[\mathrm{Heart}(x) \rightarrow \mathrm{MuscularOrgan}(x) \wedge \exists y.[\mathrm{isPartOf}(x,y) \wedge \mathrm{CirculatorySystem}(y)]]$$



SOME MOTIVATIONS

Several efforts come from the *database area*:

New kinds of databases are studied,

to manage *semi-structured* (XML), and probabilistic data.

***Information integration* is one of the major challenges for the future or IT.**

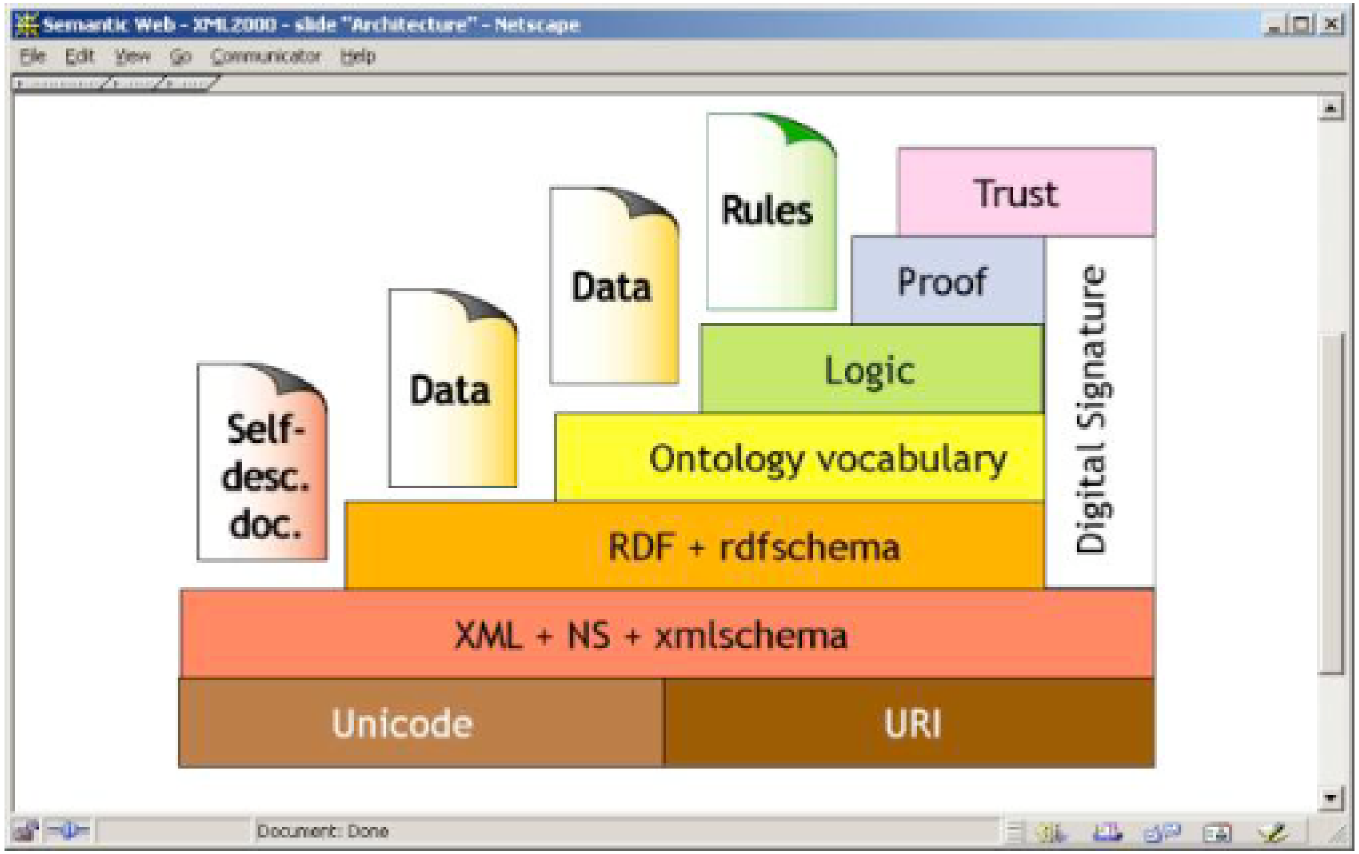
E.g., the market for information integration software has been estimated to grow from $2.5 billion in 2007 to $3.8 billion in 2012 (+8.7% per year) [IDC. Worldwide Data Integration and Access Software 2008-2012 Forecast. Doc No. 211636 (2008)].

Other motivations include

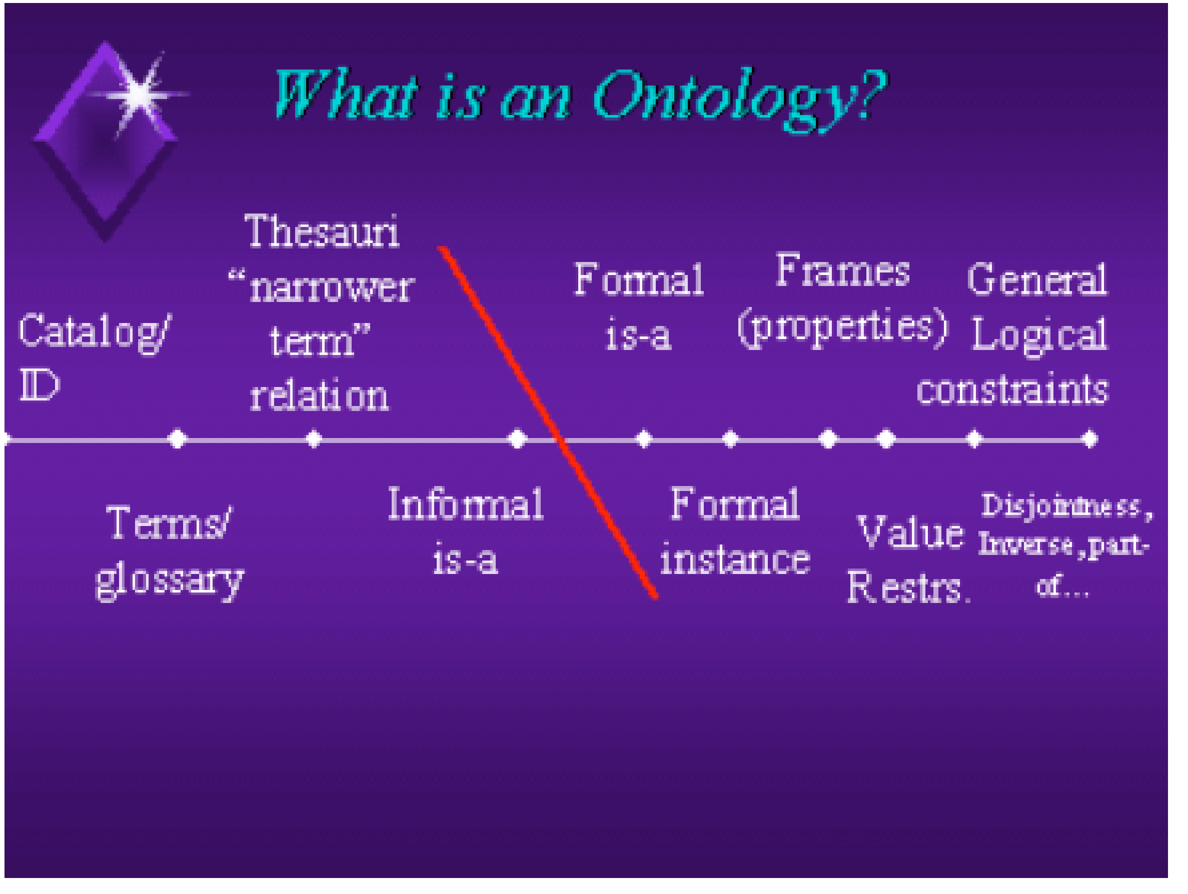
***Mediating* between ontologies – for naturally grown, distributed knowedge sources**

***Modeling* knowledge and *understanding* formally represented knowledge is both, much needed and intellectually difficult**

BERNERS-LEE'S ARCHITECTURE



A SPECTRUM OF ONTOLOGIES



SIMPLE ONTOLOGIES

Catalogs

Controled vocabulary, list of terms

Glossary

Terms and meanings, may be ambiguous

Thesaurus

Relations between terms (e.g., synonyms)

Term hierarchy

Informal, not always strict, categories may be mixed up

STRUCTURED ONTOLOGIES

Is-a hierarchies

Using inheritance

Frames

Properties associated with terms

Value restrictions

Specifications of property filler categories

Logical constraints

Arbitrarily complex relations

USES OF SIMPLE ONTOLOGIES

Same vocabulary

Site organization and navigation support

Expectation setting

“Umbrella” structure for high level terms

Browsing support

Search support

Sense disambiguation support

USES OF STRUCTURED ONTOLOGIES

Consistency checking

Completion

Interoperability support

Support validation and verification testing

Encode test suites

Configuration support

Support structured, comparative, customized search

Exploit generic/specific information

ONTOLOGY ENVIRONMENT

Collaboration and distributed workforce support

Platform interconnectivity

Scale

Versioning

Security

Life cycle issues

Ease of use

Diverse user support

Presentation style

Extensibility

MODELING WITH ONTOLOGIES

Distinctions along several perspectives

Formal vs. informal

(e.g., for inferencing or for topic groups)

Granularity

(e.g., hair color for "general purposes" or for a beauty salon)

Task

(e.g., natural language processing or information gathering)

Some difficult modeling problems

(e.g., how to represent "door"? - opening vs. physical object

ASSISTANCE IN USING ONTOLOGIES

Design and maintain high quality ontologies

- **Meaningful — all named classes can have instances**
- **Correct — captured intuitions of domain experts**
- **Minimally redundant — no unintended synonyms**
- **Richly axiomatised — (sufficiently) detailed descriptions**

Store (large numbers) of instances of ontology classes

- **Annotations from web pages**

Answer queries over ontology classes and instances

- **Find more general/specific classes**
- **Retrieve annotations/pages matching a given description**

Integrate and align multiple ontologies

BUILDING ONTOLOGIES

***Knowledge elicitation* plays a prominent role**

Knowledge often in the hand of domain experts - knowledge needs to be elicited

- **The information is often locked away in the heads of domain experts**
- **The experts may not be aware of the implicit conceptual models that they use**
- **We have to draw out and make explicit all the known & unknown knowns**

Uncovering the "obvious", "self-evident" may be important

it must be expressed explicitly for the machine (if needed for inferencing)

Huge differences across domains - explicit models may or may not exist

BUILDING ONTOLOGIES

Essential steps in modeling

1. **Establish the purpose**
2. **Informal/semiformal knowledge elicitation**
   (collect terms, organize them, paraphrase and clarify, diagram informally)
3. **Refine requirements and tests**
4. **Implementation**
   (Paraphrase and comment at each stage before implementing
   Scale up a bit, check performance, populate)
5. **Evaluate**
   (Against goals, include tests for evolution and change management)
6. **Monitor use**

ESSENTIAL DISTINCTIONS: "PART-OF" COMPOSITION

Properties of "part-of" (meronymic) relations (3 basic ones) **[Winston]**

- ***Configuration***

  whether or not the parts bear a particular functional or structural relationship to one another or to the object they constitute

- ***Homeomerous***

  whether or not the parts are the same kind of thing as the whole

- ***Invariance***

  whether or not the parts can be separated from the whole

Kinds of composition based on

particular combinations of these three basic properties

KIND OF COMPOSITION (1)

Component-integral object composition

The parts are required to bear a particular functional or structural relationship to one another--as well as to the object they constitute

Examples

Bristles are part of a toothbrush

Wheels are part of a grocery cart

Scenes are parts of films

Projective geometry is part of mathematics

Some properties

Objects may be tangible, abstract, organzational, or temporal

Components that lost connection to the whole are not parts anymore

KIND OF COMPOSITION (2)

Material object composition

This relationshp defines an invariant configuration of parts within a whole.

It defines what an object is made of.

Examples

A cappuccino is partly milk

A car is partly iron

Bread is partly flour

Some properties

Components are not separable

Components that lost connection make the whole not existing anymore

KIND OF COMPOSITION (3)

Portion object composition

This relationship defines a homeomeric configuration of parts within a whole. Thus, the parts are of the same kind as the whole.

Examples

A slice of bread is a portion of a loaf of bread.

This chunk is part of my Jell-O.

A meter is part of a kilometer

Some properties

Inheritance from whole to components (except to quantities)

KIND OF COMPOSITION (4)

Place area composition

This relationship defines a homeomeric and invariant configuration of parts within a whole. Thus, the pieces cannot be separated from the whole.

Examples

San Francisco is part of California

A peak is part of a mountain

The 50-yard line is part of a football field

Some properties

Places are similar to the whole (in some properties)

KIND OF COMPOSITION (5)

Member bunch composition

This relationship defines a collection of parts as a whole. Therer are no requirements on structural or functional properties, just membership.

Examples

A tree is part of a forest

An employee is part of a union

That ship is part of a fleet

Some properties

Relationship based on some sort of spatial proximity or social connection

Not to be confused with classification

KIND OF COMPOSITION (6)

Member partnership composition

This relationship defines an invariant collection of parts as a whole.

Hence, it is an invariant form of member bunch composition.

Examples

Ginger and Fred are a waltz couple.

Steven Fink is a managing partner in Fink and Josephson, attorneys at law.

Stan Laurel is part of Laurel and Hardy

Some properties

Members cannot be removed without destroying the partnership.

Replacing a member results in a different partnership.

A COMPARISON PROPERTIES OF COMPOSITIONS

	Configurational	*Momeomeric*	*Invariant*
Component integral object	yes	no	no
Material object	yes	no	yes
Portion object	yes	yes	no
Place area	yes	yes	yes
Member bunch	no	no	no
Member partnership	no	no	yes

META PROPERTIES (Guarino)

Identity

- **How are instances of a class distinguished from each other**
- **How are all instances of a class identified (Especially over time)**

Unity

- **How are all the parts of an instance connected**

Essence

- **Can a property change over time**
- **A property is essential to an entity if it must hold for it, not just things that happen to be true all the time.**

Dependence

- **Can an entity exist without some others**

META PROPERTIES (2)

Rigidity

- **A *rigid* property is a property that is essential to all of its instances.**
- **Every entity that can exhibit the property must do so.**

  For example, every entity that is a person must be a person

  There are no entities that can be a person but aren’t.

- **An *anti-rigid* property is one that is never essential**

  For example every instance of student isn’t necessarily a student -

  - students may cease to be students at some point without ceasing to exist

- **A *semi-rigid* property is one that is essential to some instances but not to others**

Distinctions essential for building top-level ontologies

THE ROLE OF META-PROPERTIES

Identity - distinguishing according to identity conditions (IC)

Property with IC - sortals (nouns), without IC - non-sortals (adjecives)

Often difficult to decide in practice

Unity

Related to parthood, related axioms (Simons):

- **proper part**
- **overlap**
- **actual existence of parts**
- **antisymmetry**
- **transitivity**
- **weak supplementation**

THE ROLE OF META-PROPERTIES (2)

Unity - some distinctions

topological **unity (a piece of coal)**

morphological **unity (a ball)**

functional **unity (a hammer)**

Individuation and countability

Property	***Identity***	***Unity***
Apple (a countable)	**yes**	**yes**
Apple piece (biological unity)	**yes**	**yes**
Apple food (a mass-sortal)	**yes**	**no**
Intrinsic whole (e.g., tokens, electrons)	**no**	**yes**
Red	**no**	**no**

CONSEQUENCES ON TAXONOMY BUILDING

Inheritance constraints

No inheritance from different/incompatible IC

No inheritance from UC to non-UC property

Some famous examples of "sloppy" modeling

"A physical object is an amount of matter" (Pangloss)

"An amount of matter is a physical object" (Wordnet)

Neither view is acceptable, given IC/UC analysis

There is a tendency to incorporate as much as possible in hierarchies

(does not always work properly)

USE OF RELATIONS IN MODELING

Exhibiting ontological accuracy is difficult

Natural language use is sloppy, may be ambiguous

("is" in "x is y" highly ambiguous)

e.g., polysemous terms ("the book is heavy", "I read the book yesterday")

Model classes of each meaning separately, use multiple subsumption links

Accuracy important for inferences

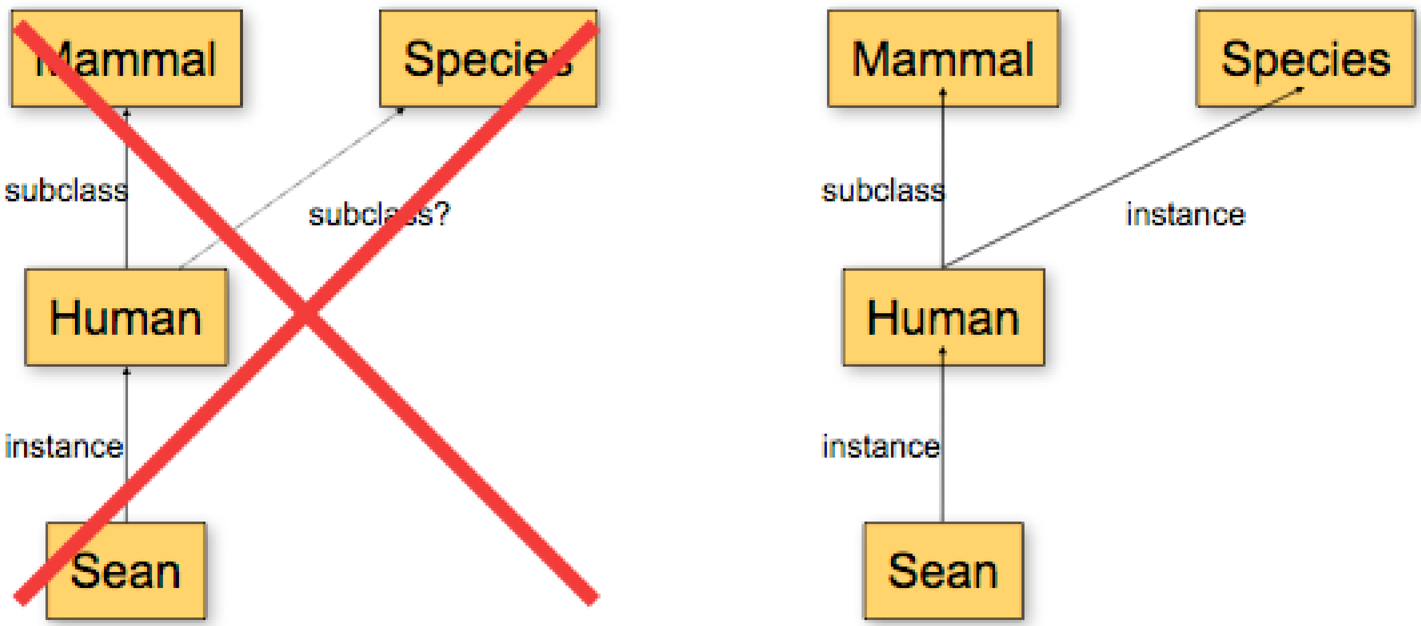
Transitivity in part-of only works for subcategorizations

Subsumption often misused

- **Subsumption allows/justifies inheritance of properties**
- **Subsumption used to obtain inheritance,**
  rather to proper represent a specialization

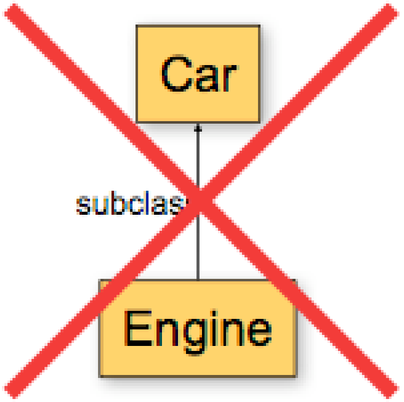
MODELING ERRORS (1)

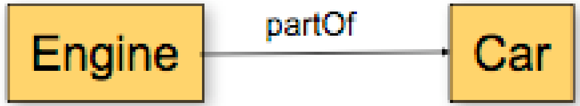
Miscategorizing instantiation as subclass



MODELING ERRORS (2)

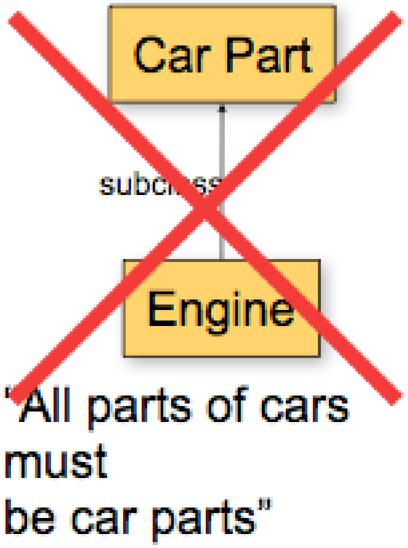
Miscategorizing part-whole as subclass

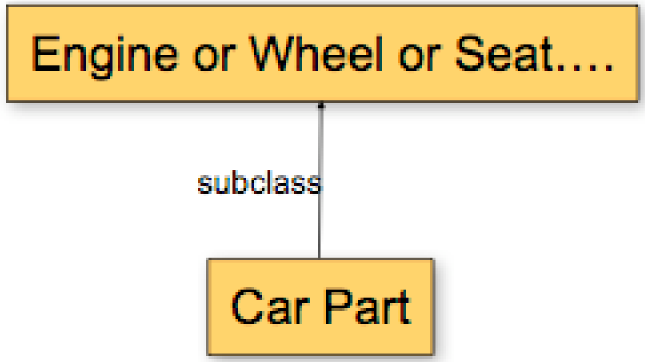




MODELING ERRORS (3)

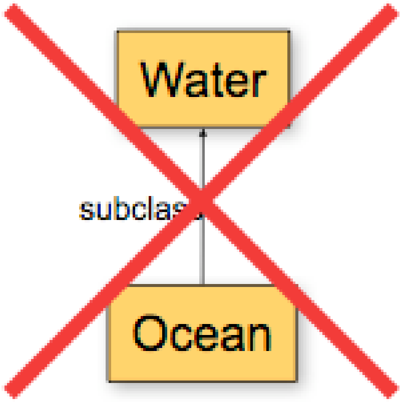
Miscategorizing disjunction as subclass





MODELING ERRORS (4)

Miscategorizing constitution as subclass



Oceans are *made up* of amounts of water

EXCERPTS FROM 2 CONTRASTIVE APPLICATIONS

Formal end of the spectrum

Logical reasoning (theorem proving)

Many widely informal uses (at current state of development)

Requirements engineering

Some role of natural language representations in both

ONTOLOGIES IN THEOREM PROVING

Purposes of Ontologies

Inferencing in a problem-solving context

Interfacing a system with others

also with natural language (accessing it or presenting its results)

Challenging problems

Increased functionality for the interfacing purpose

Interoperability between ontologies, also serving *different* purposes

Goal

Increasing quality in using combinations of formal systems

Illustrating (and acessing) the functionality of formal systems

THE GENERAL APPROACH

Overall aim

Building a knowledge base that can be interfaced with reasonable effort

Domain is mathematics – theorem proving, presentation, tutoring, …

Knowledge bases (including presentation)

Proof development system *ΩMEGA*

Proof presentation system *P.rex*

State of affairs

Knowledge bases for both purposes separately

Minimal form of connection maintained

Better understanding of representation demands to explore extensions

ROLE OF PRESENTATION

Functionalities of presentation

Expressing contents in a variety of forms

(descriptions, abstractions, partially implicit forms)

Requires general techniques and some specific to inference-rich discourse or domain

Contribution of the knowledge base

Ensuring expressibility of domain concepts and their compositionality

Supporting variations (specializations, aggregation, rhetorical preferences)

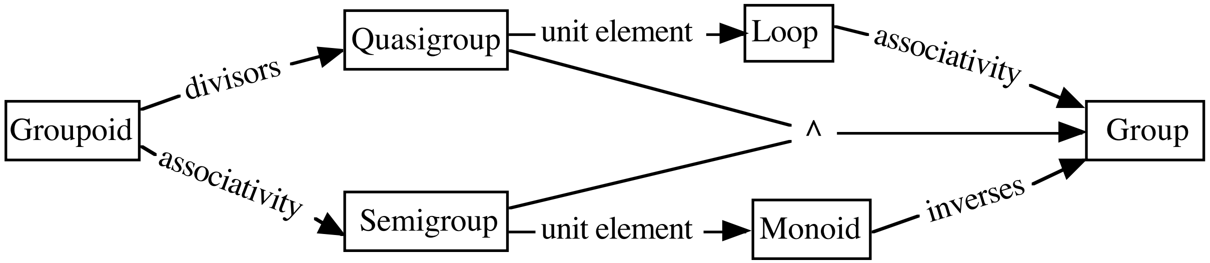
Organization forms - semantic categories

Vocabulary for restrictions on compositions

Split into two parts: *Upper model* and *textual semantic* categories

EXAMPLE – DEFINING A MATHEMATICAL GROUP

Objects with inherited properties



Alternative definitions

1) Non-empty set with an operation that is closed and associative, has unit element and inverses

2) Associative loop

REPRESENTATION FOR INFERENCING

Requirements

Efficient access for proving

Avoiding redundancy in storing objects

Representation system

Inheritance network to percolate information effectively

(symbols, definition, assertions, proofs, theories)

Support for correctness

Relations of varying granularity

(definition-entailment, used-for, depends-on, theory-inheritance)

Exploited for accomplishing proof obligations in proving equivalences

ROLE OF INFERENCING

Purpose of proving

Searching for a proof in context given by hypotheses

Accessing definitions, applying rules of calculus, transforming results

Contribution of the knowledge base

Provision of data

(definition, axioms, theorems, lemmas, …)

Specific issue – handling equivalences

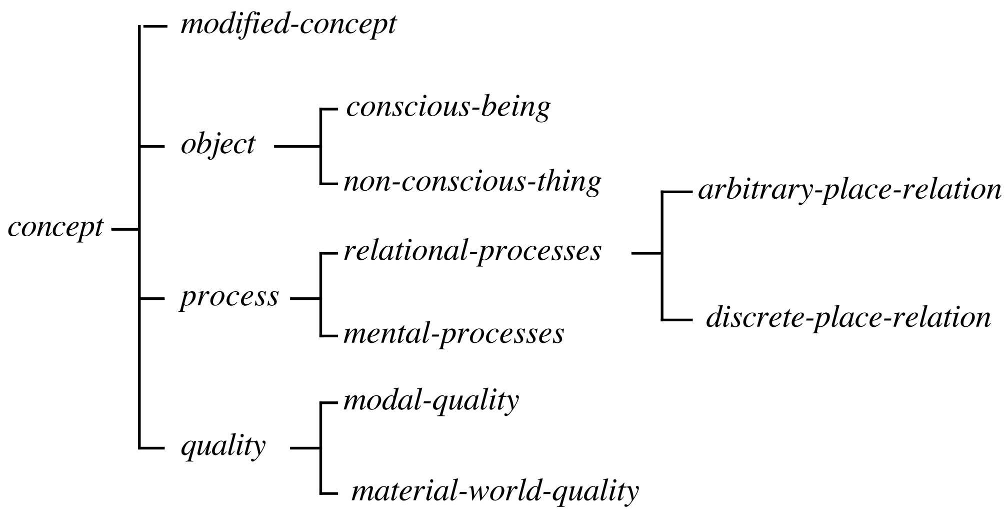
Options for defining mathematical objects

***Minimal* properties versus *commonly used* ones**

Amounts to logical redundancies versus ease of use

DISTINCTIONS FOR NL PURPOSES
THE UPPER MODEL

Fragment of concepts



Hierarchical organization according to linguistic *expressibility*

Driven by specializations and type restrictions

ANOTHER VIEW
TEXTUAL SEMANTIC CATEGORIES

Fragment of categories

- category
  - text
  - sentence
  - clause
  - vp
  - np
  - modifier
    - clause-modifier
      - clause-modifier-rankingII
      - clause-modifier-rankingI
      - clause-modifier-embedded
    - vp-modifier
    - np-modifier
    - intensifier

Hierarchical organization according to linguistic *realization*

Provides alternatives for upper model concepts to support paraphrasing

INTEROPERABILITY

Relating the mathematical database to the Upper Model

Re-representation of mathematical concepts in the Upper Model

Consequence of diverging organization principles

- **Adding hierarchical organisation**
- **Precise logical definitions not accessible**

Maintenance effort

Duplication unavoidable for more than minimal linguistic variations

For pure name reference – use of a “catch-all” concept

Independence versus functionality

REQUIREMENTS ENGINEERING

Collecting, maintaining, verifying, and using specifications for systems to be built

Bridging between natural language descriptions and system specifications

Providing a development environment for organizing specifications

- **In different degrees of (in)formality**
- **Representing portions of specifications and relating them to one another**

Ontologies support

Term glossary

Tracing specifications to architecture and system elements

Reuse of specifications from similar cases

THE SCIENTIFIC FIELD

Main regular conferences

FOIS - Formal Ontology in Information Systems

(International Conference, bi-annual, since 1998)

KR - Principles of Knowledge Representation and Reasoning

(International Conference, bi-annual, 12 conferences so far)

New - Interdisciplinary ontology conference

Ontology and knowledge representation prominent in conferences on

Artificial intelligence, reasoning

Knowledge engineering, modeling

Computational linguistics, natural language

Semantic web

LEARNING GOALS IN THE SEMINAR

Major goals

Acquaintance with basic concepts and problems in building and using ontologies

Language elements, expressiveness, complexity, purpose

Understanding consequences of the purpose of ontologies on their design (principles)

Getting a sense of degrees of difficulties and effort in handling ontologies

(assessing benefits and limitations of methods)

Major goals not included

Intensive study of formal properties of ontology languages and inference techniques

Application of learning techniques to build/match ontologies