Data and Knowledge Representation Lecture 3

Last Time We Talked About

- Boolean Algebra
- Predicate Logic (First order logic)

Today We Will Talk About

- Ontology
- Data Model

Tell me what's in this room

- Tables, chairs, windows, computers, papers, pens, people, etc..
- We can write $\exists x.y.Table(x) \wedge Room(y) \wedge In(x, y)$
- But what is a table? What is a room?
- Logic has no vocabulary of its own

Ontology Fills the Gap

- Ontology is a study of existence, of all kinds of existence, of all kinds of entities
- It supplies the predicates of predicate logic and labels that fill the boxes and circles of conceptual graph

Webster's Definition of Ontology

- "**1 :** a branch of metaphysics concerned with the nature and relations of being **2 :** a particular theory about the nature of being or the kinds of existents" -- http://www.webster.com/cgi-bin/dictionary

My Simplified Understanding

- Ontology seeks to describe entities through classification of relations among entities
- Domain ontology limits the its scope to a specific domain such as medicine
- In informatics, we further limit domain ontology to what is needed by a application or certain kinds of applications such clinical guideline, retrieval of pathology information

Why Ontology in Biomedical Domain

- Encode data
  - E.g. Patient A is diabetic and HIV positive
- Represent knowledge
  - E.g. Blood Glucose test is a diagnostic test for diabetes.

Sources of Ontology

- Observation: provides knowledge of the physical world
- Reasoning: make sense of observation by generating a framework of abstractions called “metaphysics”.

Ontology Development in Biomedical Domain

- Areas that directly involve ontology
  - Data model
  - Vocabulary/terminology
  - Knowledge based system

Philosopher's Approach to Ontology

- Top-down
  - Concerned with the entire universe
  - Build top level ontology first
- Long history
  - Lao Zi (Book of Tao)
  - Plato
  - Aristotle
  - Kant (1787)

Computer/Information Science's Approach

- Bottom Up
  - Start with limited world or specific applications
    - Exception: Cyc system
  - Designed with computing in mind
- Short History
  - First use of the term "ontology" in computer science community: McCarthy, J. 1980 "Circumscription – A Form of Non-Monotonic Reasoning", *Artificial Intelligenc*e, 5: 13, 27–39.

Problem Faced by Computer/Information Scientists

- Tower of Babel
  - Ontology used/developed by different groups for applications
    - Terminological and conceptual incompatibilities
    - Problem arise in system development and maintenance as well as data/knowledge exchange
- Insufficient expressive power

Example

- Problem Oriented Medical Record
  - Weed LL. Medical records that guide and teach. 1968. MD Comput. 1993 Mar-Apr;10(2):100-14.
  - Where "SOAP" comes from…
  - The gist: organizing medical data/information by patient problem
- Many EMRs has a place for "problem list"

Example

- Which one of the following is a "problem"
  - Cough
  - Anxiety
  - Pregnancy
  - Sleep disorder
  - Rash
- Physicians can not agree
  - Cited by a number of POEMRs as one of the reasons of failure

Another Example

- What does "acute" mean?
  - sharpness or severity e.g. *acute* pain
  - having a sudden onset, sharp rise, and short course, e.g. *acute* pancreatitis
- In a data model for finding, we had severity as an attribute. Thus need to decide where acute fit in.

To Solve the Problem

- Develop formalism for sharing (e.g. KIF, CGIF)
- Develop standard ontology
- Develop new formalism to increase expressive power

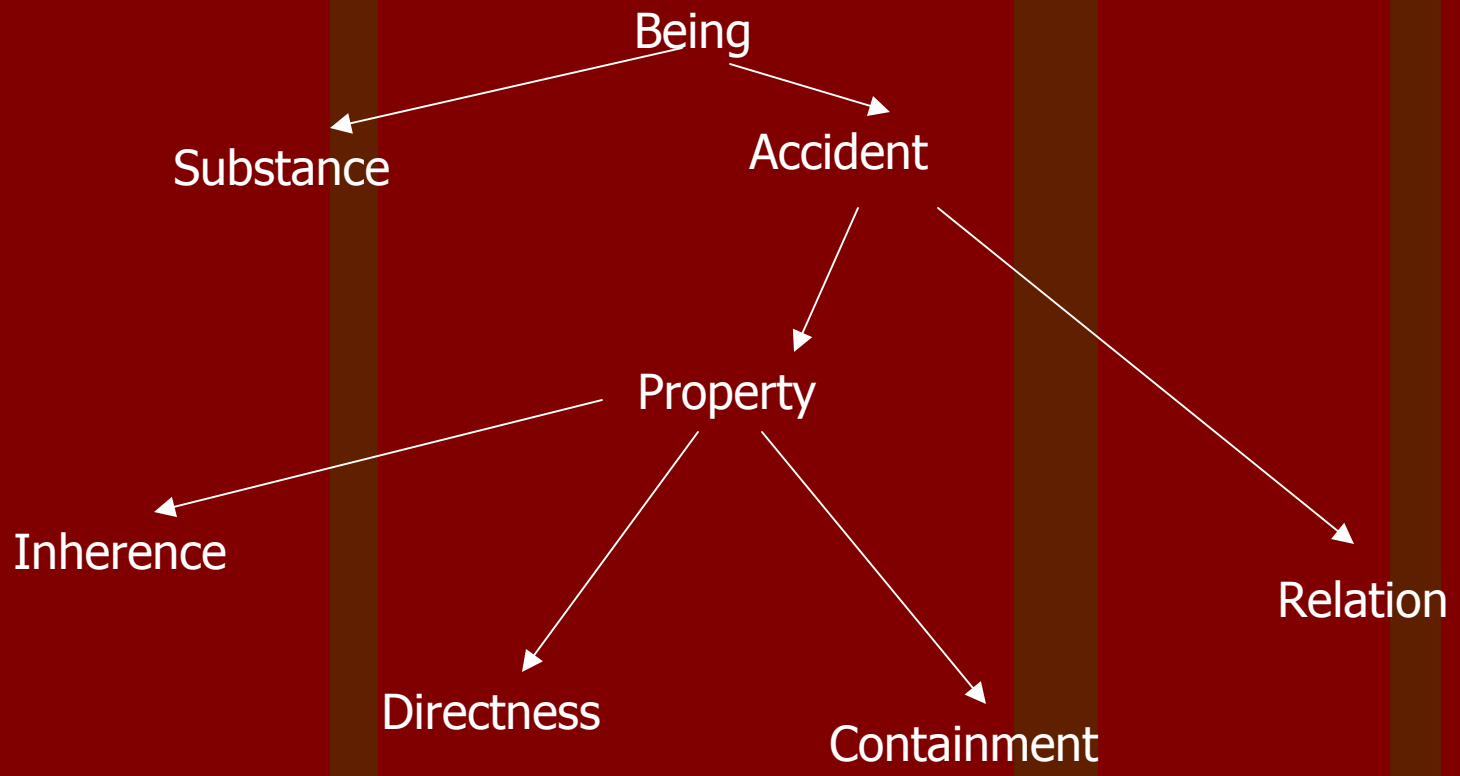
Ontological Categories

- Making a choice on ontological categories is first step in system design – John Sowa
- Ontological Categories is
  - "Class" in OO system
  - "Domain" in database theory
  - "type" in AI theory
  - "type" or "sort" in logic

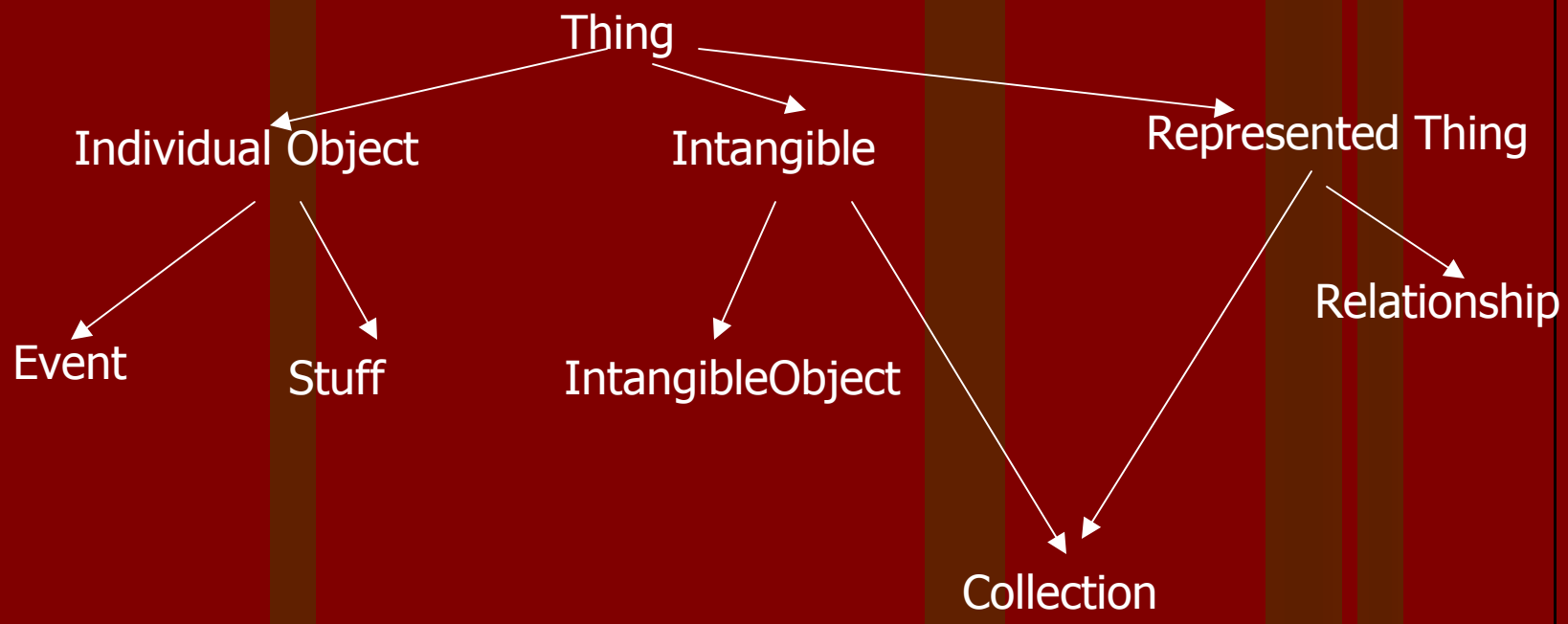
Ontological Categories

- Making a choice on ontological categories is first step in system design – John Sowa
- Ontological Categories is
  - "Class" in OO system
  - "Domain" in database theory
  - "type" in AI theory
  - "type" or "sort" in logic

Brentano's tree of Aristotle's Categories



CYC Ontology



Contrast -> Distinction

- All perceptions start with contrast
  - Bright – dark
  - Tall – short
  - Healthy – ill
  - Happy – sad
- Distinction (discrete/continuous) conceptual interpretations of perceptual contrasts

Contrast -> Distinction

- All perceptions start with contrast
  - Bright – dark
  - Tall – short
  - Healthy – ill
  - Happy – sad
- Distinction (discrete/continuous) conceptual interpretations of perceptual contrasts

Distinction -> Categories

- Distinctions maybe combined to generate categories. E.g.
  - Classify patients.
  - Distinctions: (insured, uninsured), (inpatient, outpatient), (infant, child, adult), (emergency, urgent, general)……..
  - Categories: insured pediatric emergency patient, uninsured adult inpatient……

Sowa's Ontology (Peirce and Whitehead)

- AXIOMS:
  - Physical: physical entities have location in space and a point in time. E.g. hand, hair, computer.
  - Abstract: abstract entities do not have location in space or a point in time. E.g. theorem, knowledge, story.

Sowa's Ontology

- AXIOMS:
  - Independent: independent entities can exist without being dependent on the existence of another entity. E.g. person, diary, song.
  - Relative: relative entities require the existence of some other entity. E.g. joints between bones, middle child, remission after a disease episode.
  - Mediating: mediating entities require the existence of (at least) two other entities and establish new relationship among them. E.g. theory of relativity, diagnostic strategy, cardiovascular system.

Sowa's Ontology

- AXIOMS:
  - Continuant: has only spatial parts and no temporal parts; identity cannot depend on location in space and time. E.g. gender, alert and reminder system, medication formula.
  - Occurrant: has both spatial parts (participants) and no temporal parts (stages); can only identify by location in space and time. E.g. disease episode, clinical event, medication order.

Matrix of Central Categories

	Physical		Abstract	
	Continuant	Occurrent	Continuant	Occurrent
Indepen-dent	Object	Process	Schema	Script
Relative	Juncture	Participation	Description	History
Mediating	Structure	Situation	Reason	Purpose

Data Modeling

- Is about how to represent a piece of data in a standard format.
- Often involve the specification of
  - Properties the classobject/construct (e.g. data type, relationship to other class/object/construct)
  - Attributes/elements of the data class/object/construct
  - Properties (e.g. data type) of attributes
  - Methods/functions pertaining to the data

Data Model Example – HL7 RIM

- It is good practice to give a free-text description of a attribute, as well as document the rationale and open issues
  - ***"3.1.6 status_cd : CS***

    The state of the action (e.g., newly ordered, in process, completed.) The state is communicated in coded form. The codes are strictly defined by the state-transition model of a service class. No alternative coding system can be used for the status_cd attribute (CNE, coded no exceptions.) ……" -- *Schadow, Russler, Mead, Case & McDonald, "The Unified Service Action Model," Regenstreif Institute for Health Care, 2000*

Further Reading

- http://www.togethersoft.com/services/practical_guides/umlonlinecourse/index.html
- http://www.hl7.org/, then search for “rim”

Major Considerations

- Expressiveness (Domain Complete)
- Efficiency (concise)
- Computability
- Clarity (non ambiguous)
- Generalizability
- Consistency
- Low Redundancy

Exercise

Assume you are developing an alert system to monitor errors in laboratory information systems. Identify some distinctions for categorizing the errors and describe which distinctions are in contrast with which other distinctions. How would this influence our data modeling

Reading

- Sowa: Chapter 2