Ontologies

Vinay K. Chaudhri
Mark Musen

CS227
Spring 2011
Classes and Relations Needed for SIRI
Classes and Relations Needed for Inquire Biology

Protein

A functional biological molecule consisting of one or more polypeptides folded and coiled into a specific three-dimensional structure.

Parts of Protein:
- Carbon Skeleton
- Hydrogen
- Carbon
- Functional Group
- Polypeptide

Kinds of Protein:
Actin, Allosteric Protein, Antibody, Antimicrobial Protein, Cadherin, Calmodulin, Chaperone protein, Chaperonin, more...

Protein Diagram:
- Protein
  - has parts
  - has parts
    - carbon skeleton
    - hydrogen
    - carbon
    - functional group
    - organic molecule
      - polymer
      - amphipathic molecule
      - polypeptide
# Classes and Relations Needed for Wolfram Alpha

## Input Interpretation:

<table>
<thead>
<tr>
<th>Input</th>
<th>Amount</th>
<th>Container</th>
<th>Total Calories</th>
</tr>
</thead>
<tbody>
<tr>
<td>corn flakes</td>
<td>amount</td>
<td>1 small bowl</td>
<td>total calories</td>
</tr>
<tr>
<td>orange juice</td>
<td>amount</td>
<td>1 small glass</td>
<td>total calories</td>
</tr>
</tbody>
</table>

**Result:**
274.7 Cal (dietary Calories)

## Total Nutrition Facts:

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>Value</th>
<th>% Daily Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serving Sizes</td>
<td>291 g</td>
<td></td>
</tr>
<tr>
<td>corn flakes</td>
<td>1 small bowl (43 g)</td>
<td></td>
</tr>
<tr>
<td>orange juice</td>
<td>1 small glass (248 g)</td>
<td></td>
</tr>
<tr>
<td>total calories</td>
<td>275</td>
<td></td>
</tr>
<tr>
<td>fat calories</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>% daily value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>total fat</td>
<td>1 g</td>
<td>2%</td>
</tr>
<tr>
<td>saturated fat</td>
<td>410 mg</td>
<td>2%</td>
</tr>
<tr>
<td>trans fat</td>
<td>0 g</td>
<td></td>
</tr>
<tr>
<td>cholesterol</td>
<td>43 mg</td>
<td>0%</td>
</tr>
<tr>
<td>sodium</td>
<td>305 mg</td>
<td>13%</td>
</tr>
<tr>
<td>total carbohydrates</td>
<td>64 g</td>
<td>8%</td>
</tr>
<tr>
<td>dietary fiber</td>
<td>2 g</td>
<td></td>
</tr>
<tr>
<td>sugar</td>
<td>26 g</td>
<td></td>
</tr>
<tr>
<td>protein</td>
<td>5 g</td>
<td>10%</td>
</tr>
</tbody>
</table>

## Unit Conversions:

- 1.15 MJ (megajoules)
- 1.15 x 10³ J (joules)
- 1.15 x 10⁻³ ergs (unit officially deprecated)
- 275 kcal₂₉₃ (thermodynamic kilocalories) (unit officially deprecated)
- 275 kcal₁₅₉ (kilocalories (International Steam Table 1956)) (unit officially deprecated)
Outline

• Defining an ontology and its uses
  – Lexicon vs ontology
• Ontology Design
  – Some key upper level distinctions
  – Correct choice of relationships (subclass-of, part-of)
• Ontology Engineering
  – Manual
  – Semi-Automatic
• Ontology Evaluation
Definition of Ontology

- Ontology as a philosophical discipline
  - Study of what there is
  - Study of the nature and structure of reality
- A philosophical ontology is a structured system of entities assumed to exist, organized in categories and relations

(A category enumerates all possible kinds of things that can be the subject of a predicate)
Definition of Ontology

- An ontology defines a set of representational primitives with which to model a domain of knowledge or discourse
  - The representational primitives are classes or relationships
  - Their definitions include information about their meaning and constraints on their logically consistent application

- The above definition is too permissive as it allows almost anything to be an ontology

Adapted from: http://tomgruber.org/writing/ontology-definition-2007.htm
Levels of Ontological Precision

Glossary

Taxonomy

Thesaurus

Catalog

Axiomatic theory

DB/OO scheme

ontological precision
From Logical Level to Ontological Level

- **Logical Level (Flat, no constrained meaning)**
  \[ \exists x \ (\text{Apple} (x) \land \text{Red} (x)) \]

- **Epistemological Level (structure, no constraint)**
  Many sorted logic
  \[ \exists x : \text{Apple} \ \text{Red} (x) \]
  \[ \exists x : \text{Red} \ \text{Apple} (x) \quad \text{(Axiom A1)} \]

  Structured Description
  - a is a Apple with Color = Red
  - a is a Red with Shape = apple \quad \text{(Axiom A2)}

- **Ontological level (structure, constrained meaning)**
  - Axioms A1 and A2 are not allowed
  - Apple carries an identity criterion, Red does not

---

Slide adapted from Nicola Guarino
From Logical Level to Ontological Level

- A painter may interpret the words ``Apple” and ``Red” in a completely different way
  - Three different reds on my palette: Orange, Apple, Cherry
- So an expression such as $\exists x: \text{Red Apple (x)}$ may mean that there is an Apple Red
- Two different ontological assumptions behind the red predicate:
  - Adjectival interpretation: being a red thing does not carry an identity criterion
  - Nominal interpretation: being a red color does carry an identity criterion

Formal ontological distinctions help making intended meaning explicit
Ontology vs Lexicon

- Lexicon works at the language level which is different from the ontological level
- To better understand that, let us take a detour and first understand what is a lexicon
Ontology vs Lexicon

• Lexicon is a list of words in a language
  – A vocabulary along with some knowledge about how each word is used

Example WordNet entry

Organized as synsets or synonym sets

http://wordnet.princeton.edu
Lexical Relationships

• **Synonymy**
  – Two words are synonymous if one may substitute for the other without changing the meaning
    • lodger, boarder, roomer

• **Hyponymy / Hyerpnymy**
  – A word whose meaning is included in that of another word
    • Scarlet, vermilion, and crimson are hyponyms of red

• **Meronymy / holonymy**
  – A semantic relation that holds between a whole and a part
    • Relationship between bicycle and wheel

• **Antonymy**
  – Words that are opposite to each other
    • Hot/cold
Why Can't a Lexicon be an Ontology?

- Isn't hyperonymy relation very similar to subclass-of?
- Isn't meronymy relation very similar to has-part?
Overlapping Word Senses

- In an ontology the sub-categories of a category are usually taken to be mutually exclusive
  - This breaks down for the hyponymy/hypernymy relations
  - Words are usually near synonyms

Example:
  error, mistake, blunder, slip, lapse, faux pas, bull, howler, boner

  Error and mistake overlap in meaning
  Slop and lapse overlap
  A faux pas could also be a lapse, blunder, or howler

  One cannot really create a hierarchy out of these words
Gaps in Lexicon

• A lexicon will omit any reference to ontological categories that are not lexicalized in the language
  – Usually the categories that require multiple words to describe them
    • English has not word for embarrassing bureaucratic error (bavure in French)
• Some categories are not lexicalized in any language
  – Sniglets: the words that should appear in a dictionary but should
• Higher level concepts
  – Tangible Entity, Partially Tangible, etc.
Linguistic Characterizations are not Ontological

- Semantic categorizations that are needed for correct word choice are not necessary from an ontological point of view
  - Whether a vehicle can be used as a container (bus vs canoe vs bicycle)
- Even though language distinguishes between countable and mass nouns but it is not consistent
  - Spaghetti is a mass noun but noodle is a count noun
Lexically Based Ontologies

- Technical domains
  - In technical domains the language more closely corresponds to the ontology of the domain
    - International Classification of Diseases
      - 724 Unspecified disorders of the back
      - 724.0 Spinal stenosis, other than cervical
      - 724.00 Spinal stenosis, unspecified region
      - 724.01 Spinal stenosis, thoracic region
      - 724.02 Spinal stenosis, lumbar region
      - 724.09 Spinal stenosis, other
      - 724.1 Pain in thoracic spine
      - 724.2 Lumbago
      - 724.3 Sciatica
      - 724.4 Thoracic or lumbosacral neuritis
      - 724.5 Backache, unspecified
      - 724.6 Disorders of sacrum
      - 724.7 Disorders of coccyx
      - 724.70 Unspecified disorder of coccyx
      - 724.71 Hypermobility of coccyx
      - 724.71 Cocecygodynia
      - 724.8 Other symptoms referable to back
      - 724.9 Other unspecified back disorders
## Different Levels of Representation

<table>
<thead>
<tr>
<th>Level</th>
<th>Primitives</th>
<th>Interpretation</th>
<th>Main Feature</th>
<th>Example Relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logical</td>
<td>Predicates, functions</td>
<td>Arbitrary</td>
<td>Formalization</td>
<td>None</td>
</tr>
<tr>
<td>Epistemological</td>
<td>Structuring relations</td>
<td>Arbitrary</td>
<td>Structure</td>
<td>Instance-of, subclass-of</td>
</tr>
<tr>
<td>Ontological</td>
<td>Ontological relations</td>
<td>Constrained</td>
<td>Meaning</td>
<td>Has-part, quality, role</td>
</tr>
<tr>
<td>Linguistic</td>
<td>Linguistic terms</td>
<td>Subjective</td>
<td>Language dependence</td>
<td>Hyponymy, antonymy</td>
</tr>
</tbody>
</table>

Adapted from Nicola Guarino
Goals of Developing an Ontology

- To share a **common understanding** of the entities in a given domain
  - among people
  - among software agents
  - between people and software
- To enable **reuse** of data and information
  - to avoid re-inventing the wheel
  - to introduce standards to allow interoperability and automatic reasoning
- To create **communities of researchers**
Common Uses of Ontology

- Support navigation of information
  - Example: Yahoo’s open directory (http://dir.yahoo.com)
- Serve as a controlled vocabulary
  - Example: Gene Ontology (http://www.geneontology.org)
- Provide a set of terms for semantic interoperability
  - Example: iCalendar standard (http://en.wikipedia.org/wiki/iCalendar)
- Provide schema for an information system
  - Example: Class definitions in a Java program
- Provide a clear description of the domain
- Enable problem solving behavior
  - Example: Question answering, problem solving
Ontology for the Purpose of this Course

• A set of classes and relations and their definitions in
  – a frame language
  – a structured descriptions language
Outline

- Defining an ontology and its uses
  - Lexicon vs ontology
- **Ontology Design**
  - Some key upper level distinctions
  - Correct choice of relationships (subclass-of, part-of)
- Ontology Engineering
  - Manual
  - Semi-Automatic
- Ontology Evaluation
Ontological Distinctions

John:

- height: 6 Feet
- has-part: hands
- has-employee: Jane
- kissed: Mary
- Job: researcher

Each kind of relationship has specific properties and can be studied separately

Adapted from Nicola Guarino
Formal Tools of Ontological Analysis

- Theory of essence and identity
- Theory of parts (mereology)
- Theory of unity and plurality
- Theory of dependence
- Theory of composition and constitute
- Theory of properties and qualities

A detailed treatment of each of these requires a full course in itself. We will consider a few principles that are of most practical use.

Adapted from Nicola Guarino
Event vs Entity

- **Entity**: Things that continue over a period of time maintaining their identity
  - Cell, Ribosome, Nucleus, …
- **Event**: Things that happen, unfold or develop in time
  - DNA Replication, Mitosis, Cell Division, …

- Other commonly used phrases used for this distinction
  - Occurrent, Perdurant
  - Continuant, Endurant
Upper Ontology

• An upper ontology captures a set of basic distinctions that are useful across multiple domains
  – The distinctions in an upper ontology may or may not be always useful to an application depending on its requirements
    • SIRI leverages little from the kind of upper ontology we will consider here
    • Inquire Biology does exploit the upper ontology distinctions
  – We will take a look at two specific upper ontologies and some of the basic distinctions they introduce
    • Basic Foundational Ontology
    • DOLCE
      – There are many others (Cyc, SUMO, CLIB, …)
From Ontology for 21st Century by Andrew Spear
DOLCE

Object (endurant)
- Physical
  - Amount of matter
  - Physical object
  - Feature
- Non-Physical
  - Mental object
  - Social object
  ...

Event (perdurant)
- Static
  - State
  - Process
- Dynamic
  - Achievement
  - Accomplishment

Quality
- Physical
  - Spatial location
  ...
- Temporal
  - Temporal location
  ...
- Abstract
  - Quality region
  - Time region
  - Space region
  - Color region
  ...

Adapted from Nicola Guarino
Properly Using subclass-of Relation

• If a class A is subclass of class B
  – Every instance of A is also an instance of B (ie, subset relationship)
    • Every human is also a mammal
  – Values of template slots of B are inherited by instances of A
    • Every human is an air-breathing vertebrate animal

• There are many examples where the use of subclass-of relation can be incorrect in subtle ways
Use of subclass-of

- Consider an event called: My Day
  - It has several sub events: Get UP, Go To Gym, Work, Go Home, Sleep

These are *not* subclass-of relationships
Use of Subclass-of Relation

time-duration

one hour, two hours, ....

? time-interval

1:00-2:00 next Tuesday
3:00-5:00 Wednesday
A Helpful Tool: Identity Criteria

- Identity criteria are the criteria that we use to answer questions like “Is that my dog?”
  - Identity criteria are conditions that we use to determine equality and are entailed by equality
  - Identity criteria are necessary properties

<table>
<thead>
<tr>
<th>time-duration</th>
<th>Identity criteria: same length</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>Identity criteria: same start and end time</td>
</tr>
</tbody>
</table>

Hence the two cannot be subclasses of each other
Which of the Two is Correct?

Adapted from Nicola Guarino
Possible Solution

Adapted from Nicola Guarino
Using Part-of Relationship

• There are many different flavors of part-of relationships
  – Component (e.g., handle of a car door)
  – Stuff (e.g., flour in bread)
  – Portion (e.g., a slice from a loaf of bread)
  – Area (e.g., city in a country)
  – Member (e.g., ship in a fleet of ships)
  – Partner (e.g., Laurel in Laurel & Hardy)
  – Piece (e.g., handle when removed from the door)
Outline

• Defining an ontology and its uses
  – Lexicon vs ontology
• Ontology Design
  – Some key upper level distinctions
  – Correct choice of relationships (subclass-of, part-of)
• Ontology Engineering
  – Manual
  – Semi-Automatic
• Ontology Evaluation
What is Ontology Engineering?

- Defining entities in the domain (classes)
- Arranging the entities in a taxonomy (creating class-subclass hierarchy)
- Defining slots of classes and constraints on their values
- Defining slots values

You already started to do this process on a small scale as part of HW1

As part of HW2 you will have an opportunity to do this on a larger scale
Ontology Development Process

In reality - an iterative process:
Different Philosophies for Scoping the Ontologies

• Be as encyclopedic as possible (more you can model the better)
  – The Cyc Knowledge Base, National Cancer Institute Thesaurus
• Let a thousand flowers bloom: create small scale ontologies tailored for a relatively few tasks
Competency Questions

Start by asking what questions should the ontology be able to answer?
- Which characteristics should I consider when choosing a wine?
- Is Bordeaux a red or a white wine?
- Does Cabernet Sauvignon go well with seafood?
- What is the best choice of wine for grilled meat?
- Which characteristics of a wine affects its appropriateness for a dish?
- What were good vintages for Napa Zinfandel?
Knowledge Acquisition Techniques

Adapted from Guus Schreiber
Ontology Learning

\[ \forall x \ (\text{country}(x) \rightarrow \exists y \ \text{capital_of}(y, x) \land \forall z \ (\text{capital_of}(z, x) \rightarrow y = z)) \]

disjoint(river, mountain)

capital_of \leq_r \text{located_in}

flow_through(dom: river, range: GE)

capital \leq_c \text{city}, \text{city} \leq_c \text{Inhabited GE}

c := \text{country} := (i(c), \|c\|, \text{Ref}_c(c))

\{\text{country, nation, Land}\}

river, country, nation, city, capital,...
## Algorithms for Ontology Learning

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Generic use</th>
<th>Use in ontology learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Association rule discovery</td>
<td>Discovery of “interesting” transactions in itemsets (e.g., customer data)</td>
<td>Discovery of interesting associations between words</td>
</tr>
<tr>
<td>(Hierarchical) Clustering</td>
<td>Discovery of groups in data (unsupervised)</td>
<td>Clustering of words</td>
</tr>
<tr>
<td>Classification (e.g., SVMs,</td>
<td>Prediction (supervised)</td>
<td>Classification of new concepts into an existing hierarchy</td>
</tr>
<tr>
<td>Naive Bayes, kNN, etc.)</td>
<td>Induction of rules from data (supervised)</td>
<td>Discovery of new concepts from extensional data</td>
</tr>
<tr>
<td>Inductive logic programming</td>
<td>Concept discovery (extension and intension)</td>
<td>Learning concepts and concept hierarchies</td>
</tr>
<tr>
<td>([48])</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conceptual clustering</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Outline

• Defining an ontology and its uses
  – Lexicon vs ontology
• Ontology Design
  – Some key upper level distinctions
  – Correct choice of relationships (subclass-of, part-of)
• Ontology Engineering
  – Manual
  – Semi-Automatic
• Ontology Evaluation
Ontology Evaluation

• Accuracy
  – Do the axioms comply to the expertise of one or more users?
  – Does the ontology correctly capture aspects of the real world?
• Adaptability
  – Can it be used for a range of anticipated tasks?
• Clarity
  – Does the ontology communicate the intended meaning of terms?
  – Are definitions objective and independent of context?
• Completeness
  – Is the domain of interest appropriately covered?
  – Are competency questions defined? Can it answer them?
• Conciseness
  – Does the ontology include axioms irrelevant to the domain?
• Consistency
  – Are the formal and informal representations consistent?
Ontology Evaluation

• SIRI
  – Does ontology support the kinds of things I want to do using my assistant?
  – Is the ontology easy to use?
  – Does it enable efficient software engineering?
  – Can it deal with integration of data across web services?

• Inquire Biology
  – Is ontology easily understood by the students?
  – Does it capture the textbook correctly?
  – Does it meet the teaching standards?
Evaluating Taxonomic Knowledge

- Inconsistency
  - Circularity Errors
    - Subclass partition with common Instances
    - Subclass partition with common classes
  - Partition Errors
    - Exhaustive subclass partition with common instances
    - Exhaustive subclass partition with common classes
    - Exhaustive subclass partition with external instances
  - Semantic Errors
    - Incomplete Concept Classification
    - Omission of disjoint knowledge
      - Subclass Partition Omission
      - Exhaustive subclass partition omission

- Incompleteness
  - Incomplete Concept Classification
  - Omission of disjoint knowledge

- Redundancy
  - Grammatical
    - Redundancies of subclass of relations
    - Redundancies of instance of relations
  - Identical formal definition of some classes
  - Identical formal definition of some instances
Exhaustive subclass partition with common classes
Summary

- Everyone uses and has an ontology regardless of whether they know it.
- Ontology provides a representation that is somewhere in between an uninterpreted logical representation and the natural language.
- There are some upper level distinctions and design tools available to help guide the process.
- The ontology construction is an engineering process no different than any other software artifact.
- Ontologies should be evaluated just like any other software system.
Readings

• Required readings (both on the course website)
  1. What are ontologies and why do we need them?
  2. Ontology Development 101: A Guide to Creating your First Ontology

• Optional Readings
  – Ontology and the Lexicon by Graham Hirst
  – Why Evaluate Ontology Technologies? Because They Work! By S. Staab
Forums for Recent Research on Ontologies

- International Conference on Formal Ontology in Information Systems (See [www.formalontology.org](http://www.formalontology.org))
- Knowledge Capture Conference (see [www.k-cap.org](http://www.k-cap.org))