Realism vs. Pragmatism

Ontologies as sustainable knowledge representation artefacts

Workshop on knowledge management and the future of our society

Trondheim, Norway,
September 8th, 2014
meaning of domain terms

"methanal" is a synonym of "formaldehyde"
"cell division" is broader than "mitosis"
"eau" is French for "water"

Universal properties of domain entities

"all cell membranes contain lipids"
"all fetuses were embryos"
"the surgical removal of a gallbladder is named "cholecystectomy"
"all brains develop inside animals"

Contingent characteristics of entities

"Ebola infections are rare"
"adult humans have typically 32 teeth"
"Lmn-2 interacts with Elf-2"
"fungi are not plants"
"most plant cells have cell walls"
Redesigning the map
Redesigning the map

RICH KNOWLEDGE

ONTOLOGY

TERMINOLOGY

Alan Rector (2008): "very few interesting items of knowledge that are truly ontological..."
Bill Woods (1975): "conceptual coat rack"
"methanal" is a synonym of "formaldehyde"

"cell division" is broader than "mitosis"

eau" is French for "water"

"all cell membranes contain lipids"

"all fetuses were embryos"

"all brains develop inside animals"

"fungi are not plants"

"Ebola infections are rare"

"adult humans have typically 32 teeth"

"most plant cells have cell walls"

"Lmn-2 interacts with Elf-2"

"the surgical removal of a gallbladder is named "cholecystectomy"

"all brains develop inside animals"

"fungi are not plants"
"cell division" is broader than "mitosis"

"methanal" is a synonym of "formaldehyde"

eau" is French for "water"

"all cell membranes contain lipids"

"all fetuses were embryos"

"the surgical removal of a gallbladder is named "cholecystectomy"

"all brains develop inside animals"

"fungi are not plants"

"Lmn-2 interacts with Elf-2"

"Ebola infections are rare"

"adult humans have typically 32 teeth"

"most plant cells have cell walls"
Focusing on Formal Ontology

Universal properties of domain entities:

- "all cell membranes contain lipids"
- "all fetuses were embryos"
- "the surgical removal of a gallbladder is named "cholecystectomy"
- "all brains develop inside animals"
- "fungi are not plants"
- "Lmn-2 interacts with Elf-2"

FORMAL ONTOLOGY

Boundary issues:

"Ontological Purism"
"Ontological Idiosyncrasy"
"Ontological Relativism"
"Ontological Pragmatism"
Ontological "Purism" (Smith / Ceusters)

Source: Campaign: "You can always tell a place that uses Lexware"
Ontological "Purism" (Smith / Ceusters)

- Ontologies represent universals (types) in reality
- The world is split into universals and individuals and there are objective criteria for this
- Everything is either a continuant or an occurrent
- Ontologies are independent of concrete applications
- Axioms in ontologies state what is universally true for all instances of a type
- Small set of relations
- Relations between continuant individuals are time-indexed
- First-order logics appropriate representation language

OGMS, based on BFO, distinguished:
- "Disorder" subclassOf Material object
- "Disease" subclassOf Disposition
- "Disease course" subclassOf Process

Medical terms are ambiguous: How to represent "gastric ulcer"?
- is a piece of anatomically altered stomach wall → material Object
- is a process (ulceration)
Ontological Purism: Problems (II)

- BFO 2 uses FOL to introduce ternary relations between continuants
  - located-in \((a, b, t_1)\) AND located-in \((b, c, t_1)\) $\rightarrow$ located-in \((a, c, t_1)\)
  - located-in \((a, b, t_1)\) AND located-in \((b, c, t_2)\) $\rightarrow$ ?
- FOL is undecidable
- In Description logics only two-valued relations (object properties)
  - located-in \((a, b)\) AND located-in \((b, c)\) $\rightarrow$ located-in \((a, c)\)
    - If transitive, leads to wrong entailments.
  - Otherwise, incomplete

Smith B et al. Basic Formal Ontology (BFO).
Ontological Relativism (Noy / McGuinness)
Ontological Relativism (Noy / McGuinness)

- Ontologies represent "shared conceptualizations"
- Ontologies + instances = knowledge bases
- Terminologies / vocabularies are kinds of (informal) ontologies
- Whether something is modeled as a class or an instance depends on granularity and context
- Ontologies are built to represent the knowledge needed for specific applications
- Ontology reuse is highlighted but no clear provisions for interoperability taken
- Upper-level ontology not explicitly recommended

Ontological Relativism: problems

- Ontologies as shared conceptualizations:
  - Things are represented how they are perceived / known, not as they are (philosophically: ontological realism)
  - Potentially contradictory representations of the same thing

- Example
  - *Glucose* instanceOf *Hexose*
  - What about *L-Glucose*?

- Terminologies, thesauri (e.g. UMLS, MeSH) are also understood as ontologies?
  - How to formally describe them?
  - If not, how to differentiate them?
Ontological Idiosyncrasy / Syncretism
Unprincipled, naïve (undisciplined?) approach to ontologies

Assumptions:

- informal vocabularies or database schemes wrapped into a formal language (e.g. OWL) become ontologies
- Everything which represents knowledge in the Semantic Web is an ontology

The way an ontology is shaped depends on its specific purpose

- "A little semantics goes a long way"
- "Anything goes" with regard to upper-level classes and relations (their need is often questioned)
Ontological Idiosyncrasy / Syncretism: problems

- Embedding modal, negative, or probabilistic notions. Example: NCI Thesaurus:
  \[ \text{Ureter}_\text{Small}_\text{Cell}_\text{Carcinoma} \text{ subclassOf Disease}_\text{May}_\text{Have}_\text{Finding} \text{ some Pain} \]

- Improper co-ordinations
  \[ \text{Calcium}-\text{Activated}_\text{Chloride}_\text{Channel}-2 \text{ subclassOf Gene}_\text{Product}_\text{Expressed}_\text{In}_\text{Tissue} \text{ some Lung and Gene}_\text{Product}_\text{Expressed}_\text{In}_\text{Tissue} \text{ some Mammary}_\text{Gland and Gene}_\text{Product}_\text{Expressed}_\text{In}_\text{Tissue} \text{ some Trachea} \]

- Weak or non-existing upper level and undefined primitives:
  Relies on implicit human language understanding.
  Barrier to shared conceptualizations.
  Examples:
  - Unclear whether "animal" includes "human"
  - Unclear whether events and processes are the same
  - Unclear whether "part-of" ranges over all times
  - etc...

Schulz S et al. The Pitfalls of Thesaurus Ontologization - the Case of the NCI Thesaurus.
Ontological pragmatism: the GoodOD approach
Ontological pragmatism: the GoodOD approach

- Ontologies as formal systems (using OWL DL)
- Ontological engineering supported by
  - clearly defined upper-level categories
  - closed set of basic relations
  - constraining axioms
  - understandable labels
- Criteria of dividing between classes and individuals
- Aristotelian definitions (genus – differentia)
- Naming conventions, design patterns and guidelines
- Upper ontology BioTopLite2

http://purl.org/biotop/btl2.owl

Guideline on Developing Good Ontologies in the Biomedical Domain with Description Logics

URL: http://www.purl.org/goodod/guideline

Version 1.0
December 2012

Send feedback to:
martin.boeker@uniklinik-freiburg.de
ludger.jansen@uni-rostock.de

Schulz S\textsuperscript{1,3}, Seddig-Raufie D\textsuperscript{1}, Grewe N\textsuperscript{2}, Röhl J\textsuperscript{2},
Schober D\textsuperscript{1}, Boeker M\textsuperscript{1}, Jansen L\textsuperscript{2}

\textsuperscript{1}: Institute of Medical Biometry and Medical Informatics,
University Medical Center Freiburg
\textsuperscript{2}: Institute of Philosophy, University of Rostock
\textsuperscript{3}: Department of Medical Informatics, University of Graz

11th December 2012

http://purl.org/goodod/guideline
Class-individual distinction not discretionary

Every human is a primate, every primate is a vertebrate

Every human is a vertebrate

Whether a class is the extension of a universal (type) or not is secondary

Human subClassOf Primate

∀a, t: 'instance of' (A, a, t) → 'instance of' (B, a, t)
Intuitive hierarchies ≠ good taxonomies

\[ \forall x, t: \text{'instance of'} \ (X, x, t) \leftrightarrow \text{'instance of'} \ (Y, x, t) \leftrightarrow \forall t \neg \exists x: \text{'instance of'} \ (X, x, t) \land \neg \text{'instance of'} \ (Y, x, t) \]

**OWL-DL:**

\[ X \text{ subClassOf } Y \]

\[ X \text{ and not } (Y): \text{unsatisfiable} \]

- Test: \text{there is no neoplasms that is not an oncology}  
- \text{there is no prostate that is not a neoplasm}  
- \text{there is no oncology that is not a clinical medicine}

Labelling!


Aristotelian Definitions do not permit exceptions

FOL: \[ \forall x, t: \text{'instance of'} ('Viral hepatitis', x, t) \leftrightarrow \text{'instance of'} ('Viral infection', x, t) \land \exists z: \text{'instance of'} (Liver, z, t) \land \text{'is included in'} (x, z, t) \]

OWL-DL: 'Viral Hepatitis' equivalentTo ViralInfection and 'is included in' some Liver

Test:
- There is no viral hepatitis that is not located in a liver
- There is no viral hepatitis that is not a viral infection
Always investigate the ontological commitment

- Which are exactly the instances?
- Does the label tell us what is meant?
- Is there an implicit context?

Test:
- There is no neoplasm in both lungs that is not a neoplasm in the left lung OR There is no patient with neoplasm in both lungs that is not a patient with the neoplasm in the left lung
- There is no varicose vein in the lower limb that is not a chronic peripheral venous insufficiency OR There is no patient with varicose lower limb veins that is not a patient with a chronic peripheral venous insufficiency

Guarino N, Carrara M, Giaretta P. Formalizing ontological commitment. AAAI, 1994 - aaai.org
Upper level ontologies partition the domain into disjoint and exhaustive categories.

- Upper level ontologies enforce a strict categorization
- Constraints on upper-level categories
- Upper level ontology for the biomedical domain **BioTopLite**
BioTopLite provides a small set of toplevel classes, relations, and axioms

- Precise formulations about generic and defining properties of basic categories of a domain
- Logical Framework (Description logics)
- OWL – DL (Web Ontology Language) complete and decidable language - compromise between expressiveness and performance

Automated reasoning enables checking consistency, equivalence and subsumption

Ontologies play an increasing role in new generation of biomedical terminology systems
BioTopLite2: Dealing with ambiguity

- "Every gastric ulcer is in the stomach wall"
- "Every stomach wall is part of a stomach"
- "Every gastric ulcer is in the stomach"

- 'is part of' subPropertyOf 'is included in'
  (both transitive)

- Condition equivalentTo 'Material object' or
  Disposition or Process

- 'Gastric ulcer' subClassOf Condition
  'Gastric ulcer' 'is included in' some 'Stomach wall'
  'Stomach wall' 'is part of' some Stomach
  'Gastric ulcer' 'is included in' some Stomach

<table>
<thead>
<tr>
<th><strong>BioTopLite2</strong></th>
<th><strong>BFO2</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ontological Realism &quot;light&quot;: no commitment to universalism</strong></td>
<td><strong>Ontological Realism: ontologies describe universals</strong></td>
</tr>
<tr>
<td><strong>Description Logic (OWL-DL)</strong></td>
<td><strong>First-Order-Logic (FOL), only OWL-DL class-only and experimental versions</strong></td>
</tr>
<tr>
<td><strong>Binary object properties</strong></td>
<td><strong>binary and ternary (time-indexed) relations</strong></td>
</tr>
<tr>
<td><strong>Classes: 53</strong></td>
<td><strong>Classes: 36</strong></td>
</tr>
<tr>
<td><strong>Relations: 37</strong></td>
<td><strong>Relations (in experimental TR version): 78</strong></td>
</tr>
<tr>
<td><strong>Axioms: 527</strong></td>
<td><strong>Axioms (in experimental TR version): 1572</strong></td>
</tr>
<tr>
<td><strong>Coverage: domain-independent upper-level classes and relations + few biomedical classes</strong></td>
<td><strong>Coverage: domain-independent classes and relations</strong></td>
</tr>
</tbody>
</table>

BioTopLite2: harmonization intended with BFO2 once its DL version stable
The Boundaries of Formal Ontology

**Meaning of Domain Terms**
- "methanal" is a synonym of "formaldehyde"
- "cell division" is broader than "mitosis"
- "eau" is French for "water"

**Universal Properties of Domain Entities**
- "all cell membranes contain lipids"
- "all fetuses were embryos"
- "all brains develop inside animals"
- "fungi are not plants"

**Contingent Characteristics of Entities**
- "Ebola infections are rare"
- "adult humans have typically 32 teeth"
- "most plant cells have cell walls"

**Terminology**
expressible as annotation properties

**Formal Ontology**

**Knowledge**
beyond expressiveness of formal ontologies
Universal properties of domain entities

"all cell membranes contain lipids"

"all fetuses were embryos"

"the surgical removal of a gallbladder is named "cholecystectomy"

"all brains develop inside animals"

"fungi are not plants"

"Lmn-2 interacts with Elf-2"

Definition of term as newly defined class

Interpretation as a dispositional statement

The Boundaries of Formal Ontology

FORMAL ONTOLOGY
Towards terminologies:

- Definition of meaning of a term as new class; expression of ambiguities by disjunction:
  
  'Gastric ulcer' equivalentTo 'Gastric ulcer structure' or 'Gastric ulcer process'

- Cholecystectomy equivalentTo 'Surgical removal' and 'has participant' some Gallbladder

Towards "rich" knowledge bases

- Qualitative dispositional predicates:
  
  'Lmn-2' subClassOf 'is bearer of'
  some Disposition and 'has realization' only (Interaction and 'has participant' some Elf-2)
Conclusions

- Domain ontologies are the most sustainable part of the representation of domain knowledge and they should be limited to
- Formal ontologies express what is universally true for all members of a class (all instances of a type)
- Large parts of interesting domain knowledge are not ontological
- Reusable ontologies should be
  - philosophically grounded and expressible in a computable language
  - user-friendly in terms of labelling
- This should be supported
  - by educational material
  - by expressive upper-level ontologies
  - appropriate editor and visualization tools
- Compromises are needed
  - understandability and intuitiveness of toplevel classes and relations
  - representation of ambiguous terms as disjoint classes
  - decidable and tractable logic (e.g. DL only allowing for binary relations)
Purism vs. Pragmatism
Ontologies as sustainable KR artefacts

Slides downloadable from
http://user.medunigraz.at/stefan.schulz/presentations.htm
Further readings
Ontology on the Web

- Description Logics: http://dl.kr.org/
- Protégé: http://protege.stanford.edu/
- Bioontology: http://www.bioontology.ch/
- Buffalo Ontology Site: http://ontology.buffalo.edu/smith/
- OBO Foundry: http://obofoundry.org/
- Bioportal: http://bioportal.bioontology.org/
- SNOMED CT: http://www.ihtsdo.org/snomed-ct/
  http://terminology.vetmed.vt.edu/sct/menu.cfm
- CO-ODE (Pizza ontology): http://www.co-ode.org/
- GoodOD Guideline: http://www.iph.uni-rostock.de/GoodOD-Guideline.1299.0.html
- BioTop: http://purl.org/biotop