The Relevance of Ontologies in Biology and Medicine

Stefan Schulz
Freiburg University Hospital
Medical Language and Ontology Group (MediLOG)
Department of Medical Informatics
448 in the last year
Identification of the “Thirteen most highly prioritised areas”:

1. Medical Genetics Databases and Initiatives
2. Gene Expression Information in Medical Diagnostics & Prognostics
3. Modelling & Simulation of Biological Structures & Processes/Diseases
4. Data Integration from Biosensors & Med. Devices with clinical information systems
5. Integration of patient molecular data in Electronic Health Records
6. Systems for Clinical Decision Making
7. Semantic Interoperability and Ontologies in Biomedicine
8. Technologies for Biomedical Information Integration
9. Data Interoperability & Standards
10. Connecting Biobanks to large scale databases to enable data mining
11. Patient Risk Profiling and Lifestyle Management
12. Applied Pharmaceutical Research
13. Clinical and Ethical Issues related to biomedical data processing
Content

- A cruise through the O-Space
- The "O-word": Terminological Clarification
- Purposes of Ontologies
- Mapping the O-Space
  - What is represented
  - How is it represented
- Practice of Good Ontology
A cruise through the archipelago of ontologies
MeSH
Medical Subject Headings
Bacteria [B03]
  Atypical Bacterial Forms [B03.110] +
  Bacteria, Aerobic [B03.120]
  Bacteria, Anaerobic [B03.130]
  Bacteroides [B03.140] +
  Bifilins [B03.150]
  Blood-Borne Pathogens [B03.165]
  Cholobac [B03.220] +
  Cholorela [B03.275] +
  Cyanobacteria [B03.280] +
  Endospore-Forming Bacteria [B03.300] +
  Fimbribacteria [B03.370] +
  Gram-Negative Bacteria [B03.440] +
  Gram-Positive Bacteria [B03.510]
    Actinobacteria [B03.510.024] +
    Gram-Positive Cocci [B03.510.400] +
    Gram-Positive Endospore-Forming Bacteria [B03.510.415] +
    Gram-Positive Rods [B03.510.480] +
  Proteobacteria [B03.660] +
  Spirochaetales [B03.851] +
  Spores [B03.887] +
  Sulfur-Reducing Bacteria [B03.900] +
Bacteria [B03]
  Gram-Positive Bacteria [B03.510]
    Staphylococcus [B03.510.400.790]
      Staphylococcus aureus [B03.510.400.790.750]
      Staphylococcus epidermidis [B03.510.400.790.750.543]
      Staphylococcus haemolyticus [B03.510.400.790.750.400]
      Staphylococcus hominis [B03.510.400.790.750.425]

Return to Entry Page  Link to NLM Cataloging Classification
**MeSH Descriptor Data**

**MeSH Heading**: Staphylococcus aureus

**Tree Number**: B03 510.400.790.750.100

**Annotation**: Infections = STAPHYLOCOCCAL INFECTIONS & do not bother to code with S. aureus unless particularly discussed (index IM); DF STAPH AUREUS

**Scope Note**: Potentially pathogenic bacteria found in nasal membranes, skin, hair follicles, and perineum of warm-blooded animals. They may cause a wide range of infections and intoxications.

**Allowable Qualifiers**: CH CL CY DE EN GD GE IM IT ME PH PY RE UL VI

**Entry Version**: STAP AUREUS

**Previous Indexing**: Staphylococcus (1966-1974)

**Online Note**: Use STAPHYLOCOCCUS AUREUS to search MICROCOCCUS PYOGENES 1975-91; use STAPHYLOCOCCUS 1966-74

**History Note**: 76, was MICROCOCCUS PYOGENES see under STAPHYLOCOCCUS 1963-75, MICROCOCCUS PYOGENES was see STAPHYLOCOCCUS AUREUS 1976-91

**Unique ID**: D013211

---

**MeSH Tree Structures**

- **Bacteria [B03]**
  - Gram-Positive Bacteria [B03 510]
    - Gram-Positive Cocci [B03 510.400]
      - Staphylococcaceae [B03 510.400.790]
        - Staphylococcus [B03 510.400.790.750]
          - Staphylococcus aureus [B03 510.400.790.750.100]
            - Staphylococcus epidermidis [B03 510.400.790.750.343]
            - Staphylococcus haemolyticus [B03 510.400.790.750.400]
            - Staphylococcus hominis [B03 510.400.790.750.425]
GO
Gene Ontology
Search GO

Exact Match
Terms
Gene Symbol/Name
Anfrage senden

Advanced Query
Query by Sequence

Gene Product Filters
Species
All
A. thaliana
B. anthracis str. Am

Datasource?
All
CGD
dbNBase

Evidence Code?
All Curator Approved
IGI
IEP

Ontology Filter
All
Biological Process
Cellular Component
Molecular Function

Set Filters

XML
FlatFile
Permalink

Help
GOst
The Gene Ontology
GO Request
AmiGO Request

Last updated: 2006-07-16
Copyright The Gene Ontology Consortium
GO:0043204: perikaryon (1)
GO:0009986: cell surface (688)
GO:0030312: external encapsulating structure (834)
GO:0044462: external encapsulating structure part (380)
GO:0042763: immature spore (23)
GO:0005622: intracellular (70290)
GO:0044424: intracellular part (69594)
  GO:0031255: lateral part of motile cell (0)
  GO:0031252: leading edge (208)
GO:0016020: membrane (21224)
  GO:0030673: axolemma (4)
  GO:0009941: chloroplast envelope (90)
  GO:0048475: coated membrane (238)
  GO:0012505: endomembrane system (1706)
  GO:0044425: membrane part (15359)
GO:0031090: organelle membrane (3785)
  GO:0005789: endoplasmic reticulum membrane (606)
  GO:0010008: endosome membrane (62)
  GO:0031312: extrinsic to organelle membrane (19)
  GO:0020017: flagellar membrane (1)
    GO:0046860: glycosome membrane (4)
GO:0000139: Golgi membrane (310)
  GO:0030660: Golgi-associated vesicle membrane (78)
    GO:0012507: The lipid bilayer surrounding any of the compartments of the Golgi apparatus (29)
      GO:0012508: Golgi to ER transport vesicle membrane (0)
      GO:0012509: inter-Golgi transport vesicle membrane (0)
      GO:0012510: trans-Golgi network transport vesicle membrane (45)
GO:0031228: intrinsic to Golgi membrane (77)
  GO:0030173: integral to Golgi membrane (66)
    GO:0046859: hydrogenosomal membrane (0)
    GO:0031300: intrinsic to organelle membrane (311)
    GO:0031903: microbody membrane (102)
    GO:0031966: mitochondrial membrane (1447)
    GO:0031965: nuclear membrane (353)
    GO:0019866: organelle inner membrane (1296)
    GO:0031968: organelle outer membrane (260)
    GO:0042170: plastid membrane (63)
    GO:0031095: platelet dense tubular network membrane (0)
    GO:0042651: thylakoid membrane (406)
    GO:0005774: vacuolar membrane (294)
ICD
International Classification of Diseases
Internationale Statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme
10. Revision
Version 2006

German Modification

Vierstellige Ausführliche Systematik

Kapitelübersicht

<table>
<thead>
<tr>
<th>Kapitel</th>
<th>Gliederung</th>
<th>Titel</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>A00-B99</td>
<td>Bestimmte infektiöse und parasitäre Krankheiten</td>
</tr>
<tr>
<td>II</td>
<td>C00-D48</td>
<td>Neubildungen</td>
</tr>
<tr>
<td>III</td>
<td>D50-D59</td>
<td>Krankheiten des Blutes und der blutbildenden Organe sowie bestimmte Störungen mit Beteiligung des Immunsystems</td>
</tr>
<tr>
<td>IV</td>
<td>E00-E99</td>
<td>Endokrine, Ernährungs- und Stoffwechselkrankheiten</td>
</tr>
<tr>
<td>V</td>
<td>F00-F93</td>
<td>Psychische und Verhaltensstörungen</td>
</tr>
<tr>
<td>VI</td>
<td>G00-G99</td>
<td>Krankheiten des Nervensystems</td>
</tr>
<tr>
<td>VII</td>
<td>H00-H59</td>
<td>Krankheiten des Auges und der Augenanhangsgewichte</td>
</tr>
<tr>
<td>VIII</td>
<td>H60-H85</td>
<td>Krankheiten des Ohres und des Warzenfortsatzes</td>
</tr>
<tr>
<td>IX</td>
<td>J00-J99</td>
<td>Krankheiten des Kreislaufsystems</td>
</tr>
<tr>
<td>X</td>
<td>K00-K93</td>
<td>Krankheiten des Atmungssystems</td>
</tr>
<tr>
<td>XI</td>
<td>L00-L99</td>
<td>Krankheiten der Haut und der Unterhaut</td>
</tr>
<tr>
<td>XII</td>
<td>M00-M99</td>
<td>Krankheiten des Muskel-Skelett-Systems und des Bindegewebes</td>
</tr>
<tr>
<td>XII</td>
<td>N00-N99</td>
<td>Krankheiten des Urogenitalsystems</td>
</tr>
<tr>
<td>XIV</td>
<td>O00-O99</td>
<td>Schwangerschaft, Geburt und Wochenbett</td>
</tr>
<tr>
<td>XV</td>
<td>P00-P95</td>
<td>Bestimmte Zustände, die ihren Ursprung in der Perinatalperiode haben</td>
</tr>
<tr>
<td>XVII</td>
<td>Q00-Q99</td>
<td>Angeborene Fehlbildungen, Deformitäten und Chromosomenanomalien</td>
</tr>
<tr>
<td>XVIII</td>
<td>R00-R93</td>
<td>Symptome und abnorme klinische und Laborbefunde, die anderenorts nicht klassifiziert sind</td>
</tr>
<tr>
<td>XIX</td>
<td>S00-T98</td>
<td>Verletzungen, Vergiftungen und bestimmte andere Folgen außerer Ursachen</td>
</tr>
<tr>
<td>XX</td>
<td>V01-V98</td>
<td>Äußere Ursachen von Morbidität und Mortalität</td>
</tr>
<tr>
<td>XXI</td>
<td>Z00-Z99</td>
<td>Faktoren, die den Gesundheitszustand beeinflussen und zur Inanspruchnahme des Gesundheitswesens führen</td>
</tr>
<tr>
<td>XXII</td>
<td>U00-U99</td>
<td>Schlüsselnummern für besondere Zwecke</td>
</tr>
</tbody>
</table>

Kapitel 1:

Bestimmte infektiöse und parasitäre Krankheiten
(A00-B99)
Kapitel II:

Neubildungen
(C00-D48)

C00-C07 Bosartige Neubildungen
C00-C07 Bosartige Neubildungen an genau bezeichneten Lokalisationen, als primär festgestellt oder vermutet, ausgenommen lymphatisches, blutbildendes und verwandtes Gewebe
C00-C14 Lippe, Mundhöhle und Pharynx
C15-C26 Verdauungsorgane
C30-C39 Atmungsorgane und sonstige intrathorakale Organe
C40-C41 Knochen und Gelenkknorpel
C43-C44 Haut
C45-C49 Mesotheliales Gewebe und Weichteilgewebe
C50 Brustdrüse [Mamma]
C51-C58 Weibliche Genitalorgane
C60-C63 Mannliche Genitalorgane
C64-C68 Hormone
C69-C72 Auge, Gehirn und sonstige Teile des Zentralnervensystems
C73-C77 Schilddrüse und sonstige endokrine Drüsen
C78-C80 Bosartige Neubildungen ungenau bezeichnet, sekundärer und nicht näher bezeichnet Lokalisationen
C81-C86 Bosartige Neubildungen des lymphatischen, blutbildenden und verwandten Gewebes, als primär festgestellt oder vermutet
C97 Bosartige Neubildungen als Primartumoren an mehreren Lokalisationen
D00-D09 In-situ-Neubildungen
D10-D36 Gutartige Neubildungen
D37-D48 Neubildungen unsicheren oder unbekannten Verhaltens [siehe Hinweis am Anfang der Krankheitsgruppe D37-D48]

Kapitel III:

Krankheiten des Blutes und der blutbildenden Organe sowie bestimmte Störungen mit Beteiligung des Immunsystems
(D50-D90)

D50-D52 Alimentäre Anämien
D55-D59 Hämatytische Anämien
D58-D64 Aplastische und sonstige Anämien
D65-D69 Koagulopathien, Purpura und sonstige hämorrhagische Diathesen
D70-D77 Sonstige Krankheiten des Blutes und der blutbildenden Organe
D80-D90 Bestimmte Störungen mit Beteiligung des Immunsystems

Kapitel IV:
Inkl.: Mittelohr
Exkl.: Mesothelium (C45.0)

C30.0 Nasenöhle
Conchae nasales
Naseninnenraum
Nasenknorpel
Nasenseptum
Vestibulum nasi
Exkl.: Bulbus olfactorius (C72.2)
Haut der Nase (C43.0, C44.3)
Hinterrand des Nasenseptums und der Choanen (C11.3)
Nase o.n.A. (C75.0)
Nasenbein (C41.02)

C30.1 Mittelohr
Cellulae mastoideae
Innenohr
Tuba auditiva (Eustachio)
Exkl.: Gehörgang (äußerer) (C43.2, C44.2)
Haut des (äußeren) Ohres (C43.2, C44.2)
Knöcherner Gehörgang (Meatus) (C41.01)
Ohrknorpel (C43.0)

C31.0 Bösartige Neubildung der Nasennebenhöhlen
Sinus maxillaris [Kieferhöhle]
Antrum maxillare [Highmore-Höhle]
Sinus ethmoidalis [Schleimeinzel]
Sinus frontal is [Stirnhöhle]
Sinus sphenoidal is [Keilbeinhöhle]
Nasennebenhöhlen, mehrere Teilbereiche überschneidend
[Siehe Hinweis 5 am Anfang dieses Kapitels]
Nasennebenhöhle, nicht näher bezeichnet

C32.0 Bösartige Neubildung des Larynx
Glottis
Lig. vocale [echtes Stimmband] o.n.A.
Ventriculus laryngis
Supraglottis
Arepiglottische Falte, laryngeale Seite
Epiglottis (supraglottischer Anteil) o.n.A.
Hinterer (laryngealer) Flächen der Epiglottis
Plica vestibularis
Word NET

MeSH: Medical Subject Headings
**Noun**

- **S.** (n) **tongue**, *lugus, glosses, clapper* (a mobile mass of muscular tissue covered with mucous membrane and located in the oral cavity)
- **S.** (n) **natural language, tongue** (a human written or spoken language used by a community, opposed to e.g. a computer language)
- **S.** (n) **tongue, knife** (any long thin projection that is transient) "tongues of flame licked at the walls"; "rifles exploded quick knives of fire into the dark"
- **S.** (n) **tongue** (a manner of speaking) "he spoke with a thick tongue"; "she has a glib tongue"
- **S.** (n) **epit, tongue** (a narrow strip of land that juts out into the sea)
- **S.** (n) **tongue** (the tongue of certain animals used as meat)
- **S.** (n) **tongue** (the flap of maternal under the laces of a shoe or boot)
- **S.** (n) **clapper, tongue** (metal striker that hangs inside a bell and makes a sound by hitting the side)

**Verb**

- **S.** (v) **tongue** (articulate by tonguing, as when playing wind instruments)
- **S.** (v) **tongue** (lick or explore with the tongue)
Noun

- S (n) *tongue*, *lingua*, *glossa*, *clapper* (a mobile mass of muscular tissue covered with mucous membrane and located in the oral cavity)
  - *part meronym*
    - S (n) *taste bud*, *taste bud*, *gustatory organ* (an oval sensory and organ on the surface of the tongue)
  - *direct hypernym* / *inherited hypernym* / *der term*
    - S (n) *organ* (a fully differentiated structural and functional unit in an animal that is specialized for some particular function)
    - S (n) *articulator* (a movable speech organ)
  - *part holonym*
    - S (n) *mouth*, *oral cavity*, *oral fissure*, *rima oris* (the opening through which food is taken in and vocalizations emerge) *he stuffed his mouth with candy*
    - S (n) *throat*, *pharynx* (the passage to the stomach and lungs, in the front part of the neck below the chin and above the collarbone)
- *derivationally related form*
  - *W (adj) lingual* (Related to *lingua*) (pertaining to or resembling or lying near the tongue) *lingual inflammation*; *the lingual surface of the teeth*
  - *W (adj) lingual* (Related to *lingua*) (consisting of or related to language) *linguistic behavior*; *a linguistic atlas*; *linguual diversity*
SNOMED
Clinical Terms
acute myocardial infarction of anterior wall
Content

- A cruise through the O-Space
- The "O-word": Terminological Clarification
- Purposes of Ontologies
- Mapping the O-Space
  - What is represented
  - How is it represented
- Practice of Good Ontology
Artifacts for ordering domain entities, relating word meanings or providing semantic reference

- Vocabularies
- Terminologies
- Thesauri
- Concept Systems
- Classifications
- Ontologies
ONTOLOGY: Unresolved Terminological Confusion…

- Artifacts for ordering domain entities, relating word meanings or providing semantic reference
  - Vocabularies
  - Terminologies
  - Thesauri
  - Concept Systems
  - Classifications
  - Ontologies
ONTOLOGY: Unresolved Terminological Confusion…

- Artifacts for ordering domain entities, relating word meanings or providing semantic reference

- Vocabularies
- Terminologies
- Thesauri
- Concept Systems
- Classifications

- Ontologies
Different scientific traditions:
Biology, Medicine, Philosophy, Logic, Linguistics, Library and Information Science, Computer Science, Cognitive Science

Different philosophical schools of thinking: Platonism, Aristotelian Realism, Conceptualism, Relativism, Idealism, Postmodernism, Constructivism, Nominalism, Tropism,…
Ontologies / Terminological Systems come in different flavors

Nodes and Links

(Inf)formal Definitions

**domain or region of DNA [GENIA]:** A substructure of DNA molecule which is supposed to have a particular function, such as a gene, e.g., c-jun gene, promoter region, Sp1 site, CA repeat. This class also includes a base sequence that has a particular function.

**ExtractionOfForeignBodyFromStomachByIncision ≡**
**RemovalOfForeignBodyFromDigestiveSystem**
**RemovalOfForeignBodyFromStomach**
**IncisionOfStomach**
∃ has-part.(∃ Method.RemovalAction)
∃ DirectMorphology.ForeignBody)
∃ has-part.(∃ Method.IncisionAction)
∃ ProcedureSite.stomachStructure)
What do the nodes in Ontologies / Terminological Systems stand for?

names
categories
types
universals
sets
synsets
sorts
properties
descriptors
classes
terms
concepts
entities
Content

- A cruise through the O-Space
- The "O-word": Terminological Clarification
- Purposes of Ontologies
- Mapping the O-Space
  - What is represented
  - How is it represented
- Practice of Good Ontology
Purposes of Ontologies: General

- Semantic Interoperability
- Terminology control
- Knowledge extraction
- Knowledge management
- Natural Language Processing
- Document retrieval
- Formal reasoning about knowledge structures
Purposes of Ontologies: Medicine

- Support of clinical coding (diagnoses, procedures):
  - Accounting
  - Health Statistics

- Support of Biomedical Science:
  - Interoperability between heterogeneous databases
  - Indexing of biomedical literature
Purposes of Ontologies: Biology

- Data and information retrieval and analysis
- Semantic Annotation of Genes, Proteins in terms of localization, pathways, functions...
- Intelligent text mining of literature abstracts: "Bibliomics"
Content

- A cruise through the O-Space
- The "O-word": Terminological Clarification
- Purposes of Ontologies
- Mapping the O-Space
  - What is represented
  - How is it represented
- Practice of Good Ontology
Mapping the space of Ontology

- instead of providing a definition...
Mapping the space of Ontology

Representation of arbitrary propositions

Representation of term meanings

Representation restricted to real world entities

What is represented?

How is it represented?

Axiomatic Theory

Conceptual Schema

Thesaurus

Taxonomy

Glossary

Catalog


Gunnar Klein & Barry Smith: ontology.buffalo.edu/concepts/ConceptsandOntologies
Content

- A cruise through the O-Space
- The "O-word": Terminological Clarification
- Purposes of Ontologies
- Mapping the O-Space
  - What is represented
  - How is it represented
- Practice of Good Ontology
Mapping the space of Ontology

Representation of arbitrary propositions

Representation of term meanings

Representation restricted to real world entities

What is represented?

Gunnar Klein & Barry Smith: ontology.buffalo.edu/concepts/ConceptsandOntologies
Mind
Independent
Reality

Intelligent Minds

Natural Language

Classes

Individual Entities (Instances, Particulars)
Mapping the space of Ontology: Realist perspective

Representation of arbitrary propositions

What is represented?

Representation of term meanings

How is it represented?

- Universals (types, kinds) are invariants in reality, e.g. cell, molecule, eye, inflammation, ...
- All universals refer to non-empty (at some moment) classes of (individual) entities in the world

Gunnar Klein & Barry Smith: ontology.buffalo.edu/concepts/ConceptsandOntologies
What is represented?

- Concepts do not necessarily imply the extension to classes in reality
- ("retinal transplant", "yin deficiency", "missing digit", "prevented pregnancy")
- Concepts as mind constructs may be oriented to prototypes, their extension exhibits large inter-individual variation
- Concepts can be related by imprecise conceptual relations such as "is broader as"

How is it represented?

- Representation restricted to real world entities
- Representation of term meanings
- Representation of arbitrary propositions

Gunnar Klein & Barry Smith: ontology.buffalo.edu/concepts/ConceptsandOntologies
Schools of Thinking

<table>
<thead>
<tr>
<th>Categories</th>
<th>Realist</th>
<th>Conceptualist</th>
<th>Nominalist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Universals</td>
<td><img src="realist_universals.png" alt="Image" /></td>
<td><img src="conceptualist_universals.png" alt="Image" /></td>
<td><img src="nominalist_universals.png" alt="Image" /></td>
</tr>
<tr>
<td>Concepts</td>
<td><img src="realist_concepts.png" alt="Image" /></td>
<td><img src="conceptualist_concepts.png" alt="Image" /></td>
<td><img src="nominalist_concepts.png" alt="Image" /></td>
</tr>
<tr>
<td>Names</td>
<td><img src="realist_names.png" alt="Image" /></td>
<td><img src="conceptualist_names.png" alt="Image" /></td>
<td><img src="nominalist_names.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Reference to Reality

Individual Entities (Instances, Particulars)
Mapping the space of Ontology
Nominalist perspective

- Names are created in an ad hoc fashion from linguistic predicates.
  
  Examples:
  - “People in SR 1048 at 7pm today”
  - “Nontraffic accident involving other off-road motor vehicle” (ICD9-CM: E821)
  - Tuberculosis of lung, bacteriological and histological examination not done (ICD-10: A16.1)
  - “Follow-up inpatient consultation for an established patient which requires at least two of these three key components: a detailed interval history; a detailed examination; medical decision making of high complexity. Counseling and/or coordination of care with other providers or agencies are provided consistent with the nature of the problem(s) and the patient’s and/or family’s needs. Usually, the patient is unstable or has developed a significant complication or a significant new problem. Physicians typically spend 30 minutes at the bedside and on the patient's hospital floor or unit.” (Current Procedural Terminology Code: HCPT06)

Gunnar Klein & Barry Smith: ontology.buffalo.edu/concepts/ConceptsandOntologies
A cruise through the O-Space

The "O-word": Terminological Clarification

Purposes of Ontologies

Mapping the O-Space

- What is represented
- How is it represented

Practice of Good Ontology
Mapping the space of Ontology

What is represented?

- Representation of arbitrary propositions
- Representation of term meanings
- Representation restricted to real world entities

How is it represented?

- Axiomatic Theory
- Conceptual Schema
- Thesaurus
- Taxonomy
- Glossary
- Catalog

Gunnar Klein & Barry Smith: ontology.buffalo.edu/concepts/ConceptsandOntologies

Mapping the space of Ontology Catalogs

- Representation of arbitrary propositions
- Representation of term meanings
- Representation restricted to real world entities

Hierarchical Ordering of Semantic Nodes

Axiomatic Theory  Conceptual Schema  Thesaurus  Taxonomy  Glossary  Catalog

Gunnar Klein & Barry Smith: ontology.buffalo.edu/concepts/ConceptsandOntologies

Mapping the space of Ontology
Catalogs

What is represented?

- **Catalog:** a set of terms without constraints (formal or informal) to characterize their meaning.

<table>
<thead>
<tr>
<th>Representation of arbitrary propositions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representation of term meanings</td>
</tr>
<tr>
<td>Representation restricted to real world entities</td>
</tr>
</tbody>
</table>

Gunnar Klein & Barry Smith: ontology.buffalo.edu/concepts/ConceptsandOntologies

Mapping the space of Ontology

Glossary

- Representation of arbitrary propositions
- Representation of term meanings
- Representation restricted to real world entities

Glossary: catalogue with glosses in natural language.

What is represented?

- **Taxonomy**: terms (and glosses) are organized into a subsumption (subclass) hierarchy with Property inheritance

---

**Mapping the space of Ontology**

- **Representation of arbitrary propositions**
- **Representation of term meanings**
- **Representation restricted to real world entities**

Gunnar Klein & Barry Smith: ontology.buffalo.edu/concepts/ConceptsandOntologies

What is represented?

- **Thesaurus**: taxonomy coupled with additional semantic relations (part-of, similar to, etc.).

Representation of arbitrary propositions

Representation of term meanings

Representation restricted to real world entities

---

Gunnar Klein & Barry Smith: ontology.buffalo.edu/concepts/ConceptsandOntologies

Mapping the space of Ontology Conceptual Schemas

What is represented?

- Representation of arbitrary propositions
- Representation of term meanings
- Representation restricted to real world entities

How is it represented?

- **Conceptual Schema**: Set of terms, attributes and relations (or frames and slots) with explicit descriptions (definitions)

Gunnar Klein & Barry Smith: ontology.buffalo.edu/concepts/ConceptsandOntologies

What is represented?

Axiomatic Theory: Formal system with a clear semantics that captures the meaning of the adopted vocabulary via logical formulas.

How is it represented?

Representation of arbitrary propositions

Representation of term meanings

Representation restricted to real world entities

Axiomatic Theory

Conceptual Schema

Thesaurus

Taxonomy

Glossary

Catalog

### Mapping the space of Ontology: Notions of Ontology

<table>
<thead>
<tr>
<th>Representation of arbitrary propositions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representation of term meanings</td>
</tr>
<tr>
<td>Representation restricted to real world entities</td>
</tr>
</tbody>
</table>

**Axiomatic Theory**  |  **Conceptual Schema**  |  **Thesaurus**  |  **Taxonomy**  |  **Glossary**  |  **Catalog**
---|---|---|---|---|---

Gunnar Klein & Barry Smith: [ontology.buffalo.edu/concepts/ConceptsandOntologies](http://ontology.buffalo.edu/concepts/ConceptsandOntologies)

Perspective of Philosophical Ontology

Mapping the space of Ontology:

Notions of Ontology

Representation
- restricted to real-world entities
- representation of term meanings
- representation of arbitrary propositions

Axiomatic Theory
Glossary
Taxonomy
Thesaurus
Conceptual Schema
Catalog


Gunnar Klein & Barry Smith: ontology.buffalo.edu/concepts/
ConceptsandOntologies
Mapping the space of Ontology:
Notions of Ontology

- Representation of arbitrary propositions
- Representation of term meanings
- Representation restricted to real world entities

Gunnar Klein & Barry Smith: ontology.buffalo.edu/concepts/ConceptsandOntologies

Mapping the space of Ontology: Biomedical Vocabularies

- Representation of arbitrary propositions
- Representation of term meanings
- Representation restricted to real world entities

SNOMED CT
WordNet
GO
ICD
MeSH

Content

- A cruise through the O-Space
- The "O-word": Terminological Clarification
- Purposes of Ontologies
- Mapping the O-Space
  - What is represented
  - How is it represented
- Practice of Good Ontology
Don’t mix up **universals** (Concepts, Classes) with **individuals** (Instances)

- **subclass-of** (Motor Neuron, Neuron) (FMA, OpenGALEN)
- **is-a** (Motor Neuron, Neuron)
- **instance-of** (Motor Neuron, Neuron) (FlyBase)

But:

- **instance-of** *(my Hand, Hand)*
- **instance-of** *(this amount of insulin, Insulin)*
- **instance-of** *(Germany, Country)*
- **not:** instance of (Heart, Organ)
- **not:** instance of (Insulin, Protein)
Don’t use superclasses to express roles

- is-a (Fish, Animal)
- is-a (Fish, Food) ??
- is-a (Acetylsalicylic Acid, Salicylate)
- is-a (Acetylsalicylic Acid, Analgetic Drug) ??

Be aware of the “rigidity” of classes
Partition the ontology by principled upper level categories

Example: DOLCE’s Upper Ontology

**Endurant (Continu.ant)**

*Physical*
- Amount of matter
- Physical object
- Feature

*Non-Physical*
- Mental object
- Social object

**Perdurant (Occurrent)**

*Static*
- State
- Process

*Dynamic*
- Achievement
- Accomplishment

**Quality**

*Physical Qualities*
- Spatial location
- ...

*Temporal Qualities*
- Temporal location
- ...

*Abstract Qualities*
- ...

*Abstract*

- Quality region
- Time region
- Space region
- Color region

Source: S. Borgo ISTC-CNR
Be aware of ambiguities

“Institution” (NCIT) may refer to
1. (abstract) institutional rules
2. (concrete) things instituted
3. act of instituting sth.

“Tumor”
1. evolution of a tumor as a disease process
2. having a tumor as a pathological state
3. tumor as a physical object

“Gene”
1. a (physical) sequence of nucleotides on a DNA chain
2. a collection of (1)
3. A piece of information conveyed by (1)
Use semantically precise Basic Relations

First version of the OBO Relation Ontology

Foundational relations

\textit{is\_a}
\par
\textit{part\_of}

Spatial relations (connecting one entity to another in terms of relations between the spatial regions they occupy)

\textit{located\_in}
\par
\textit{contained\_in}
\par
\textit{adjacent\_to}

Temporal relations (connecting entities existing at different times)

\textit{transformation\_of}
\par
\textit{derives\_from}
\par
\textit{preceded\_by}

Participation relations (connecting processes to their bearers)

\textit{has\_participant}
\par
\textit{has\_agent}
Nontaxonomic Relations between Classes are ambiguous!

- **has-part**(*Cell, Axon*) (Gene Ontology)
  - Do cells without axons exist?
  - Do axons without cells exist?

- **has-part**(*Neuron, Axon*) (FMA)
  - Does every neuron have an axon?
Nontaxonomic Relations between Classes are ambiguous!

A, B are classes, inst-of = class membership
rel: relation between instances  Rel: relation between classes

\[ \text{Rel} (A, B) = \text{def} \]

\[ \exists x: \text{inst-of} (x, A) \land \text{inst-of} (y, B) \land \text{rel} (x, y) \quad \text{OR} \]
\[ \forall x: \text{inst-of}(x, A) \rightarrow \exists y: \text{inst-of} (y, B) \land \text{rel} (x, y) \quad \text{OR} \]
\[ \forall x: \text{inst-of}(x, A) \rightarrow \exists y: \text{inst-of} (y, B) \land \text{rel} (x, y) \quad \text{AND} \]
\[ \forall y: \text{inst-of}(y, B) \rightarrow \exists x: \text{inst-of} (x, A) \land \text{rel} (x, y) \]
Nontaxonomic Relations between Classes are ambiguous!

\[ \exists x: \text{inst-of}(x, A) \land \text{inst-of}(y, B) \land \text{rel}(x, y) \lor \forall x: \text{inst-of}(x, A) \rightarrow \exists y: \text{inst-of}(y, B) \land \text{rel}(x, y) \lor \forall y: \text{inst-of}(y, B) \rightarrow \exists x: \text{inst-of}(x, A) \land \text{rel}(x, y) \land \text{inst-of} = \text{class membership} \]

\[ \text{rel}: \text{relation between instances} \quad \text{Rel}: \text{relation between classes} \]
Example: Part-of and Has-Part between Classes
MediLOG: Ontology activities

- Three EU projects
  - SemanticMining: Semantic Interoperability and Data Mining in Biomedicine
  - @neurist: Integrated decision support system to assess the risk of aneurysm rupture in patients and to optimize their treatments.
  - BootSTREP: Integration of biological fact databases and terminological repositories to implement a text analysis system which continuously increases their coverage by analyzing biological documents.
Semantic Mining - A Network of Excellence in the Field of Biomedical Informatics

by Hans Arntoft

The objective of the Network of Excellence entitled Semantic Interoperability and Data Mining in Biomedicine [SemanticMinIn] funded by the European Sixth Framework Programme, is to establish Europe as the international scientific leader in medical and biomedical informatics.

The long-term goal of the network will be the development of generic methods and tools supporting the critical tasks of the field: data mining, knowledge discovery, knowledge representation, abstraction and indexing of information, semantic-based information retrieval in a complex and high-dimensional information space, and knowledge-based adaptive systems for provision of decision support for dissemination of evidence-based medicine.

The general objective of a Network of Excellence (NoE) is to bridge gaps in the European research infrastructure and to facilitate cross-fertilisation between scientific disciplines. Traditionally academic departments in the domain have their roots either in computer science, system engineering (including a variety of engineering disciplines) or in medical or clinical context. The proposed network is composed of partners from these scientific areas, all bringing their experience and in-depth knowledge together into a common framework. An important aspect of this is the merging of medical or clinical informatics and bioinformatics including the new fields of genomics and proteomics.

Another bridging activity addressed by this NoE is knowledge transfer and co-operation between academic and organisations and SMEs in the health and welfare sector, including standardisation bodies and the different public and private institutions involved in health care delivery and management. The national institutes and organisations responsible for policy making and quality management with a regulatory and normative function will have an important role to play in the network. We believe that cooperation between these organisations and those involved in research departments needs to be strengthened, both in the early phase of research programme identification and in the later phases of implementation and large-scale evaluation of results and impact. The bridging activities between different levels of the healthcare system are exemplified in the figure.
MediLOG: Ontology activities

- Three EU projects
  - SemanticMining: Semantic Interoperability and Data Mining in Biomedicine
  - @neurist: Integrated decision support system to assess the risk of aneurysm rupture in patients and to optimize their treatments.
  - BootSTREP: Integration of biological fact databases and terminological repositories to implement a text analysis system which continuously increases their coverage by analyzing biological documents.
Welcome to @neurIST, a European initiative to integrate biomedical informatics in the management of cerebral aneurysms.
MediLOG: Ontology activities

- Three EU projects
  - SemanticMining: Semantic Interoperability and Data Mining in Biomedicine
  - @neurist: Integrated decision support system to assess the risk of aneurysm rupture in patients and to optimize their treatments.
  - BootSTREP: Integration of biological fact databases and terminological repositories to implement a text analysis system which continuously increases their coverage by analyzing biological documents.
BOOTStrap (Bootstrapping Of Ontologies and Terminologies STrategic REsearch Project) is funded in the EC’s 6th Framework Programme. The project will pull together already existing biological fact databases as well as various terminological repositories and implement a text analysis system which continuously increases their coverage by analysing biological documents.

Impact

Biological knowledge, up until now, is scattered in heterogeneous database formats and locked in unstructured natural language documents. The intended integration of biological knowledge in a homogeneous conceptual framework will ease access to this fragmented knowledge and substantially increase its usability for R & D purposes, e.g., in the European bio-tech and pharmaceutical industry.

BOOTStrap’s main innovations

Knowledge integration and reuse in the biology domain are the main goals of the BOOTStrap project. The resources and text mining tools developed within the project are expected to boost the performance in various bio application tasks. In particular, BOOTStrap aims at:

- exploiting already existing terminological resources (thesauri, classification systems, etc.) and combining them within a common, standardized conceptual representation framework. Based on this domain-specific background knowledge advanced natural language technologies are employed for the analysis of biological documents in order to fill conceptual gaps in these resources by automatically acquiring new terms, concepts and relations,
- creating, incrementally maintaining and continuously updating a repository of biological facts based on employing a comprehensive bio-lexicon and a standards-based formal bio-ontology for text analysis. Facts are extracted from biological documents in a fully automatic way, they are subsequently filtered and validated for novelty, redundancy, contradiction, etc.,
- developing resources and resource-building NLP tools for text-based knowledge harvesting in order to support information extraction and text mining in the biology domain,
- allowing multilingual public access to continuously updated and validated biological fact repositories.

Administrative details

BOOTStrap (FP6 – 038099) is a Specific Targeted Research Project (STREP) of the European Union’s 6th Framework Programme, Thematic Priority 2 (Information Society Technologies) within the fourth call of the programme. It addresses the strategic objective "Semantic-based Knowledge and Content Systems". The project started on April 1, 2006 and will end on March 31, 2009. The overall budget is 3.6 million euro. Six partners from four European countries (Germany, U.K., Italy, France) and one Asian partner from Singapore are involved in the project.
Thank you for your attention