

Effektive Interventionen zur Qualitätssteigerung von Ontologien

Martin Boeker

Department für Medizinische Biometrie und Medizinische
Informatik

Klinisches Krebsregister und IT, CCC Freiburg
Universitätsklinikum Freiburg

martin.boeker@uniklinik-freiburg.de

2015-04-13 conhIT-Satellitenveranstaltung 2015 von GMDS und BVMI



Gute Qualität von Ontologien in vier Schritten sichern

1. Die Zielsetzungen, Anforderungen und Nutzungsszenarien analysieren und die vorhandenen Ressourcen dabei berücksichtigen
2. Vorhandene Ontologien, Terminologien und andere Begriffssysteme wiederverwenden
3. Upper-Level Ontologien und Ontology Design Patterns einsetzen
4. Deskriptions-Logik und Ontologie-Editoren als Werkzeuge zu beherrschen lernen

Ziel- und Anforderungsanalyse

- ▶ Mit welcher Zielsetzung wird die Ontologie entwickelt?
- ▶ Was soll die Ontologie repräsentieren?
 - ▶ Inhalt
 - ▶ Granularität
 - ▶ Abgrenzung
- ▶ In welchem Kontext soll die Ontologie genutzt werden?
 - ▶ Software
 - ▶ Systeme
 - ▶ Performance

Mit Competency Questions Anforderungen formalisieren

- ▶ Competency Question sind ein „semi-formaler“ Weg Anforderungen (an Ontologien) zu formulieren
- ▶ Sie können — bei Verwendung von Top-Level Ontologien — auch formal repräsentiert werden
- ▶ Damit können sie auch zur (formalen) Evaluation von Ontologien eingesetzt werden

Ontologieentwicklung ist sehr personal- und zeitintensiv

- ▶ Ontologieentwicklung ist interdisziplinär: *Domänenexperten*, Ontologie- und Softwareentwickler sollten von Beginn an eingebunden werden
- ▶ Ontologieentwicklung sollte vor den Projekten starten, in denen die Zielontologie als Kommunikations- und Interoperabilitätsgrundlage genutzt werden soll

Auf vorhandene Ontologien/ standardisierte Terminologien aufbauen

- ▶ Nicht immer wieder neu definieren, was bereits an anderer Stelle vorhanden ist, sondern darauf basieren:
 - ▶ BioPortal¹
 - ▶ Open Biomedical Ontologies (OBO)²
 - ▶ UMLS Terminology Services³
 - ▶ DIMDI⁴

¹<http://bioportal.bioontology.org/>

²<http://www.obofoundry.org/>

³<https://uts.nlm.nih.gov/home.html>

⁴<https://www.dimdi.de/static/de/index.html>

Ist die Qualität und Nachhaltigkeit von vorhandenen Ontologien „institutionell“ gesichert?

- ▶ SNOMED CT: IHTSDO⁵
- ▶ LOINC: Regenstrief Institute⁶
- ▶ NCI Therausus: National Cancer Institute⁷
- ▶ FMA: University of Washington⁸

⁵<http://www.ihtsdo.org/snomed-ct>

⁶<https://loinc.org/>

⁷<http://ncit.nci.nih.gov/>

⁸<http://sig.biostr.washington.edu/projects/fm/>

Welche Qualität und formalen Eigenschaften haben vorhandene Ontologien und Terminologien

- ▶ Format
 - ▶ DL, OBO, Frames, informell
- ▶ formale Expressivität
- ▶ Vernetzung mit anderen Ontologien (crosslinking)
- ▶ Nutzung von Upper-Level Ontologien
- ▶ Annotation
- ▶ Dokumentation/ Publikation

Vorhandenes ontologisches Wissen mit Upper-Level Ontologien nutzen

- ▶ Mit Upper-Level Ontologien auf vorhandene und *bewährte* ontologische Strukturen zurückgreifen
- ▶ Upper-Level Ontologien als Rahmen und Startpunkt für eigene Entwicklungen einsetzen
- ▶ Die Definitionen der Upper-Level Klassen zur Konsistenzsicherung der eigenen Definitionen zur Entwicklungszeit nutzen
- ▶ Interoperabilität mit Ontologien erreichen, die die selben (ähnliche) Upper-Level Ontologien nutzen

Eine Reihe von Top- und Upper-Level Ontologien stehen zur Verfügung

- ▶ BFO: Basic Formal Ontology⁹
- ▶ DOLCE: Descriptive Ontology for Linguistic and Cognitive Engineering¹⁰
- ▶ BioTop¹¹
- ▶ GFO, SUMO, etc.

⁹<http://ifomis.uni-saarland.de/bfo/>

¹⁰<http://www.loa.istc.cnr.it/old/DOLCE.html>

¹¹<http://purl.org/biotop>

Basic Formal Ontology (BFO)

- ▶ BFO ist ein *de facto* Standard für OBO Ontologien
- ▶ BFO ist über die Taxonomie hinaus nicht in DL formalisiert
- ▶ BFO 2 soll in DL axiomatisiert werden (in Entwicklung)
- ▶ BFO orientiert sich an *realist ontology*

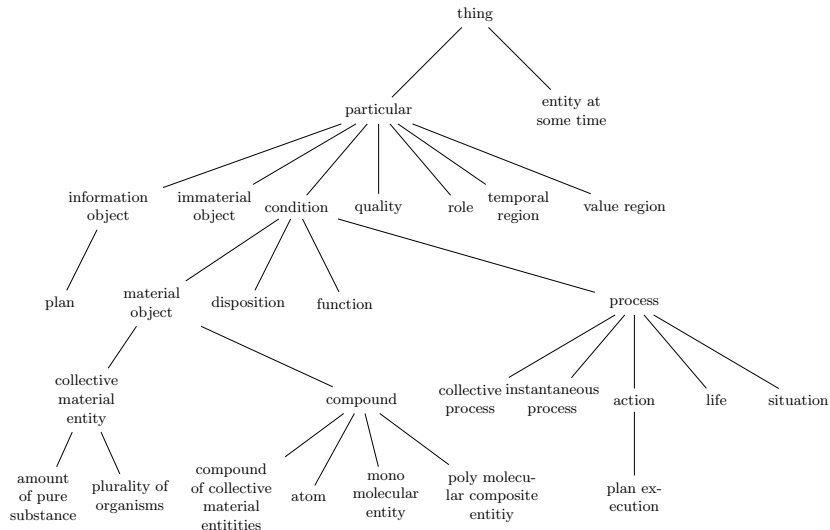
Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE)

- ▶ DOLCE ist in DL axiomatisiert
- ▶ DOLCE ist modular aufgebaut
- ▶ DOLCE nimmt einen *gemäßigt konstruktivistischen ontologischen Standpunkt* ein

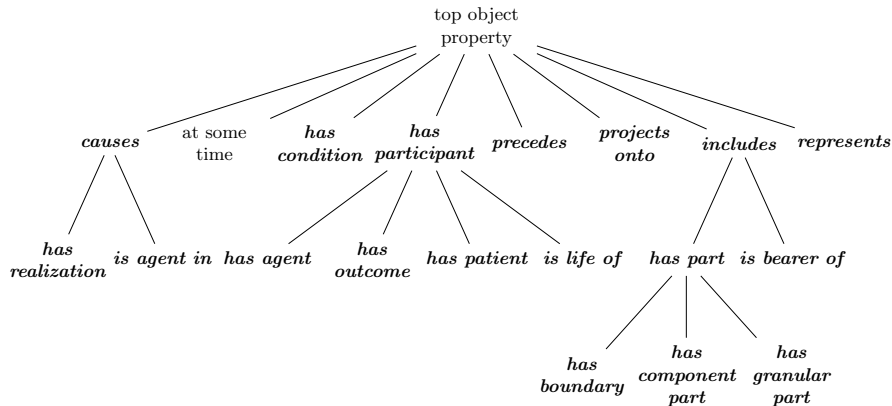
Toplevel Ontologie für biomedizinische Ontologien: BioTop

- ▶ BioTop ist „toplevel“ agnostisch
 - ▶ nutzt Eigenschaften von DOLCE und BFO
- ▶ BioTop repräsentiert für die Lifesciences wichtige Klassen und Relationen
- ▶ BioTopLite ist eine DL Repräsentation von BioTop

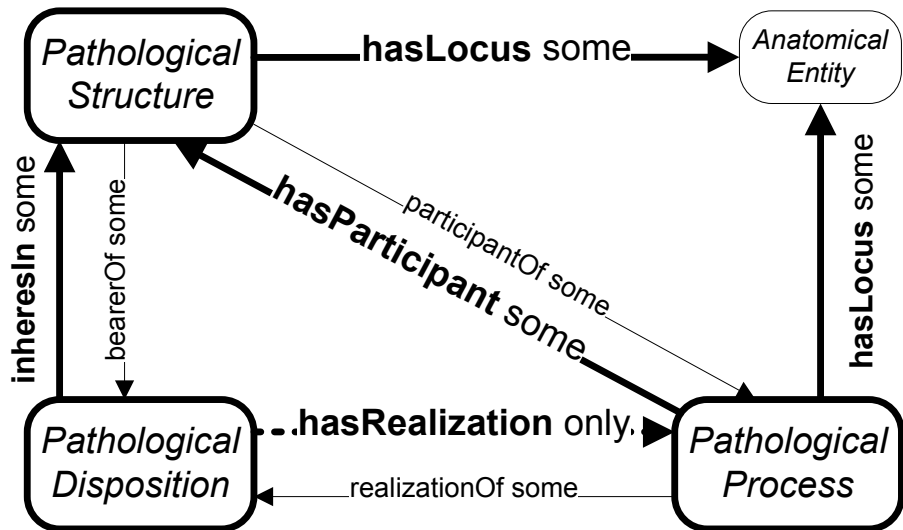
Taxonomie von BioTopLite 2



Relationen von BioTopLite 2



Beispiel für die Verwendung von BioTop: Disease Model



Wiederkehrende ontologische Entwurfsmuster können mit Ontologie Design Patterns abgebildet werden

- ▶ Closure
- ▶ Exception
- ▶ Normalisation
- ▶ Spatial Disjointness
- ▶ Viele Entwurfsmuster sind schon durch Upper-Level Ontologien abgebildet

Beispiel für ein einfaches Entwurfsmuster: Closure Pattern

BacterialPopulation equivalentTo
Population and
(**hasGranularPart** some *BacterialCell*) and
(**hasGranularPart** only *BacterialCell*)

Die Semantik der zur Verfügung stehenden Formalismen ist nicht intuitiv

- ▶ Der „natürliche Sprachgebrauch“ ist ambivalent
 - ▶ Beispiel: *is_a*
- ▶ DL wird als formale „Starrheit“ mit begrenzter Expressivität empfunden
 - ▶ Datenbankerfahrung (DL \neq ER)
 - ▶ Softwareentwicklung (DL \neq UML)
 - ▶ Wissensrepräsentation (DL \neq Frames)

Insbesondere die Verwendung der *Quantoren* ist schwierig

- ▶ Die Semantik der Quantoren ist auch für „erfahrene“ Ontologieentwickler problematisch
 - ▶ Die freie Variable, wie aus der Prädikatenlogik 1. Ordnung gewohnt, wird abgekürzt
 - ▶ SOME (\exists) impliziert Existenz (existentielle Abhängigkeit)
 - ▶ ONLY (\forall) führt bei transitiven Relationen zu unerwünschten Ergebnissen

„Visuelle“ Ontologie-Editoren (Protegé) erleichtern die Arbeit

The screenshot displays the Protegé ontology editor interface. The left pane shows a class hierarchy for 'TumorOfBreastWithDistantMetastasis', which is highlighted in blue. The right pane shows the logical axioms for this class, including its description and various restrictions.

Class Hierarchy (Left Pane):

- InformationObject
 - MaterialObject
 - CollectiveMaterialEntity
 - Compound
 - Atom
 - CompoundOfCollectiveMaterialEntities
 - MonoMolecularEntity
 - PolyMolecularCompositeEntity
 - StructuredBiologicalEntity
 - AnatomicalStructure
 - BodyPart
 - BodyPortion
 - MalignantAnatomicalStructure
 - Metastasis
 - PrimaryTumor
 - TumorAggregate
 - BreastTumorAggregate
 - TumorOfBreastWith10OrMoreMetastasisInAxilla
 - TumorOfBreastWith1to3MetastaticAxillaryLymph
 - TumorOfBreastWith4to9MetastaticAxillaryLymph
 - TumorOfBreastWithAdheringMetastaticAxillaryLymph
 - TumorOfBreastWithClinicalDetectedMetastatic
 - TumorOfBreastWithDistantMetastasis**
 - TumorOfBreastWithMetastasisIn3OrMoreAxillary
 - TumorOfBreastWithMetastaticInfradavicularLymph
 - TumorOfBreastWithMetastaticInternalLymphatic
 - TumorOfBreastWithMetastaticSupradavicularLymph
 - TumorOfBreastWithMetastaticAxillaryLymph
 - TumorOfBreastWithSentinalDetectedMetastatic
 - TumorOfBreastWithSentinalDetectedMetastatic

Annotations (Right Pane):

Description: TumorOfBreastWithDistantMetastasis

Equivalent To:

- BreastTumorAggregate and (hasPart some DistantMetastasisOfTumorOfBreast)

SubClass Of:

- BreastTumorAggregate
- isRepresentedBy some BreastTNM_M1

Other Class Axioms:

- TumorAggregate and (hasPart some BreastTumor)
- AnatomicalStructure and (isBearerOf some (TissueDignity and (projectsOnto some Malignant)))
- atSomeTime some (isPartOf some Organism)
- Disposition or Function or ImmaterialObject or InformationObject or MaterialObject or Quality or Role or TemporalRegion or ValueRegion
- isRepresentedBy only (EntityAtSomeTime)

Die Qualitätsbewertung von Ontologien ist nicht „standardisiert“

- ▶ Was eine „gute“ Ontologie ist, kann schlecht definiert werden
- ▶ Kriterien für die Evaluation
 - ▶ strukturelle Eigenschaften
 - ▶ Nähe zu „Goldstandard-Ontologien“
 - ▶ Erfüllung von Competency-Questions
 - ▶ inhaltliche Kriterien
 - ▶ Nutzbarkeit
- ▶ Methoden der Evaluation
 - ▶ expertenbasiertes Rating
 - ▶ Metriken

Gute Qualität von Ontologien in vier Schritten sichern

1. Die Zielsetzungen, Anforderungen und Nutzungsszenarien analysieren und die vorhandenen Ressourcen dabei berücksichtigen
2. Vorhandene Ontologien, Terminologien und andere Begriffssysteme wiederverwenden
3. Upper-Level Ontologien und Ontology Design Patterns einsetzen
4. Deskriptions-Logik und Ontologie-Editoren als Werkzeuge zu beherrschen lernen

Literatur

1. Schulz S, Suntisrivaraporn B, Baader F, Boeker M. SNOMED reaching its adolescence: Ontologists' and logicians' health check. *International Journal of Medical Informatics*. 2009 Apr;78:S86–94.
2. Schulz S, Spackman K, James A, Cocos C, Boeker M. Scalable representations of diseases in biomedical ontologies. *J Biomed Semantics* [Internet]. 2011 [cited 2012 Aug 15];2(Suppl 2):S6. Available from: <http://www.biomedcentral.com/content/pdf/2041-1480-2-S2-S6.pdf>
3. Boeker M, Tudose I, Hastings J, Schober D, Schulz S. Unintended consequences of existential quantifications in biomedical ontologies. *BMC Bioinformatics*. 2011 Nov 24;12:456.
4. Schober D, Tudose I, Svatek V, Boeker M. OntoCheck: verifying ontology naming conventions and metadata completeness in Protégé 4. *Journal of Biomedical Semantics* [Internet]. 2012 [cited 2013 May 2];3(Suppl 2):S4. Available from: <http://www.jbiomedsem.com/content/3/S2/S4>
5. Seddig-Raufie D, Jansen L, Schober D, Boeker M, Grewe N, Schulz S. Proposed actions are no actions: re-modeling an ontology design pattern with a realist top-level ontology. In: Herre H, Hoehndorf R, Loebe F, editors. *Proceedings of Ontologies in Biomedicine and Life Sciences (OBML 2011)*, Berlin, Germany [Internet]. Berlin, Germany: *Journal of Biomedical Semantics*; 2012 [cited 2012 Aug 15]. p. S2. Available from: <http://www.jbiomedsem.com/content/3/S2/S2/abstract>
6. Schulz S, Seddig-Raufie D, Grewe N, Röhl J, Schober D, Boeker M, et al. Guideline on Developing Good Ontologies in the Biomedical Domain with Description Logics [Internet]. Freiburg: University Medical Center Freiburg; 2012 [cited 2013 Jul 17]. Available from: <http://www.purl.org/goodod/guideline>
7. Schulz S, Boeker M. BioTopLite: An Upper Level Ontology for the Life Sciences. Evolution, Design and Application. In: Hornbach M, editor. *INFORMATIK 2013 Ontologien in den Lebenswissenschaften*. Bonn: Gesellschaft für Informatik; 2013. pp. 1889–99.
8. Boeker M, Jansen L, Grewe N, Röhl J, Schober D, Seddig-Raufie D, et al. Effects of Guideline-Based Training on the Quality of Formal Ontologies: A Randomized Controlled Trial. *PLoS ONE* [Internet]. 2013 May 7 [cited 2013 May 7];8(5):e61425. Available from: <http://dx.doi.org/10.1371/journal.pone.0061425>
9. Boeker M, Grewe N, Röhl J, Schober D, Schulz S, Seddig-Raufie D, et al. Measuring the Effect of a Guideline-based Training on Ontology Design with a Competency Questions based Evaluation Approach. In: Hornbach M, editor. *INFORMATIK 2013 Ontologien in den Lebenswissenschaften*. Bonn: Gesellschaft für Informatik; 2013. pp. 1783–95.