



Harmonizing SNOMED CT with BioTopLite. An exercise in principled ontology alignment

Stefan Schulz, Catalina Martínez-Costa

Medical University of Graz, Austria



Semantic Interoperability

- ▶ Meaningful exchange of biomedical information requires support by ontologies and terminologies
- ▶ SNOMED CT and others (e.g. OBO Foundry ontologies, WHO classifications) have recognized the need of precise descriptions of the entities denoted by terms and concepts, their ontological nature and the way they are related
- ▶ They increasingly use a formal language, typically description logics (DL) axioms

Problem statement

- ▶ Numerous ontologies have been developed bottom-up in different contexts
- ▶ They do not share an joint upper-level. Key terms (e.g. disorder, animal, drug, situation, condition) have different meanings and often lack explicit definitions
- ▶ Alignment of lexical labels does not guarantee alignment of meaning
- ▶ Interoperability between semantic artefacts is facilitated by
 - ▶ A well-understood and well performing representational language
 - ▶ A top-level layer of shared categories and relations (Top level ontology)

Goal

- ▶ Analyse the ontological structure of the OWL version of SNOMED CT
 - ▶ Upper level concepts (classes)
 - ▶ Relations
 - ▶ Constraints
- ▶ Manually align it with the Upper-Level Ontology BioTopLite2 (BTL2)
- ▶ Check for consistency and performance
- ▶ Assess feasibility of moving to a richer language as claimed by Rector & Brandt (2008)

Rector AL, Brandt S. Why do it the hard way? The case for an expressive description logic for SNOMED. J Am Med Inform Assoc. 2008 Nov-Dec;15(6):744-751

Comparison Upper Level Classes

SNOMED CT

- ▶ Body structure
- ▶ Clinical finding
- ▶ Environment or geogr. location
- ▶ Event
- ▶ Observable entity
- ▶ Organism
- ▶ Pharmaceutical/biologic product
- ▶ Physical force
- ▶ Physical object
- ▶ Procedure
- ▶ Qualifier value
- ▶ Record artefact
- ▶ Situation with explicit context
- ▶ Social context
- ▶ Special concept
- ▶ Specimen
- ▶ Staging and scales
- ▶ Substance

BioTopLite2

- ▶ Disposition
- ▶ Function
- ▶ Immaterial object
- ▶ Information object
- ▶ Material object
- ▶ Process
- ▶ Quality
- ▶ Role
- ▶ Temporal region
- ▶ Value region

Comparison Top Level of Relations

SNOMED CT*

- ▶ Access
- ▶ After
- ▶ Associated finding
- ▶ Associated morphology
- ▶ Associated procedure
- ▶ Associated with
- ▶ Causative agent
- ▶ Component
- ▶ Direct device
- ▶ Direct morphology
- ▶ Direct substance
- ▶ Finding context
- ▶ Finding site
- ▶ Has active ingredient
- ▶ Has dose form
- ▶ Has focus
- ▶ Has intent
- ▶ Has interpretation
- ▶ Has specimen
- ▶ Interprets
- ▶ Laterality
- ▶ Method
- ▶ Occurrence
- ▶ Procedure context
- ▶ Procedure device
- ▶ Procedure site
- ▶ Procedure site - Indirect
- ▶ Role group
- ▶ Specimen source topography
- ▶ Subject relationship context
- ▶ Temporal context
- ▶ Using access device
- ▶ Using device
- ▶ Using substance

Total 66, *most frequent, cover 95%

BioTopLite2

- ▶ at some time
- ▶ includes
 - ▶ has part
 - ▶ has boundary
 - ▶ has granular part
 - ▶ has component part
 - ▶ is bearer of
- ▶ causes
 - ▶ has realization
- ▶ precedes
- ▶ has condition
- ▶ projects onto
- ▶ has participant
 - ▶ has agent
 - ▶ has patient
 - ▶ has outcome
 - ▶ is life of
- ▶ is referred to at time
- ▶ represents

Total 37, inverses not displayed

Mapping process

- ▶ Manual mapping of upper classes and relations
- ▶ Iterative approach:
 - ▶ add map
 - ▶ classify (using DL reasoning)
 - ▶ satisfiability check:
 - ▶ positive: check entailments
if OK proceed to next mapping step
 - ▶ negative: analyse error (explanation tool)
fix error and classify again

Dealing with performance issues

▶ Problem:

- ▶ SNOMED CT is huge (300,000 concepts), but has OWL-EL expressiveness
- ▶ BTL2 has OWL-DL expressiveness

▶ Result:

- ▶ Performance unacceptable in described workflow

▶ Solution

- ▶ Instead of whole SNOMED CT, use random modules
- ▶ Module creation: maximal number of patterns → maximal representativeness (pattern: unique combination of relations + concepts of a given subhierarchy)
- ▶ Module produced with a signature of one concept per patterns → 11,000 concepts

Results: Class mappings

- ▶ Only one equivalence mapping: *Organism*
- ▶ Eight subclass mappings
 - ▶ e.g. *sct:Event* subclassOf *btl2:process*
- ▶ Four complex subclass mappings
 - ▶ e.g. *sct:Finding* subclassOf *btl:process* or *btl:material entity* or *btl:disposition*
- ▶ No mappings:
 - ▶ *sct:Qualifier value*
 - ▶ *sct:Situation with explicit context*
 - ▶ *sct:Social context*
 - ▶ *sct:Special concept*

overly
hetero-
geneous

Results: relation mapping (I)

- ▶ Far more complex
- ▶ Do far only done for subset of relations
- ▶ Only one equivalence: **sct:After** \equiv **bt12:precedes**
- ▶ Context-dependent mapping, e.g.
 - ▶ Findings: **sct:RoleGroup** \equiv **bt12:hasCondition**
 - ▶ Procedures: **sct:RoleGroup** \equiv **bt12:hasPart**
- ▶ Most relations mapped as subrelations to BTL2 relations, with refined domains and ranges, e.g.
 - ▶ **sct:Finding site** subrelation of **bt1:is included in** with domain *bt1:condition* and range *sct:Body structure*;

Results: relation mapping (II)

- ▶ Lossy mapping where exact meaning cannot be reconstructed by BTL2,
 - ▶ e.g. **sct:Procedure site - Indirect**,
sct:Direct morphology
- ▶ Complex relationships: shortcuts for composed expressions, e.g.
 - ▶ **sct:hasFocus**: intent of a procedure to reach a certain goal. Would require model of intentionality
 - ▶ **sct:FindingContext**: information model entity related to *Finding*:
btl2:represents only (*FindingX* or (not *Finding*))

Performance issues

Explanation for: 'Allergic sensitization by patch test (disorder)' EquivalentTo Nothing

1)	'Allergic sensitization by patch test (disorder)' EquivalentTo 'Complication of patch testing (disorder)' and 'Allergic sensitization (disorder)' and ('Role group (attribute)' some ('Causative agent (attribute)' some 'Patch test substance (substance)))	In ALL other justifications
2)	'Complication of patch testing (disorder)' EquivalentTo 'Complication of diagnostic procedure (disorder)' and ('Role group (attribute)' some ('Associated with (attribute)' some 'Patch test (procedure)))	In ALL other justifications
3)	'Patch test (procedure)' SubClassOf 'Type 4 hypersensitivity skin test (procedure)' and 'Test for allergens (procedure)'	In ALL other justifications
4)	'Type 4 hypersensitivity skin test (procedure)' SubClassOf 'Delayed hypersensitivity skin test (procedure)'	In ALL other justifications
5)	'Delayed hypersensitivity skin test (procedure)' SubClassOf 'In vivo test of hypersensitivity (procedure)' and 'Skin test (procedure)' and ('Role group (attribute)' some (('Method (attribute)' some 'Evaluation - action (qualifier value)')) and ('Procedure site (attribute)' some 'Skin structure (body structure)'))	In ALL other justifications
6)	'In vivo test of hypersensitivity (procedure)' SubClassOf 'Clinical immunological test (procedure)'	In ALL other justifications
7)	'Clinical immunological test (procedure)' SubClassOf 'Immunologic procedure (procedure)' and ('Role group (attribute)' some ('Component (attribute)' some 'Immune response, function (observable entity)))	In ALL other justifications
8)	'Component (attribute)' Range 'Substance (substance)'	In ALL other justifications
9)	'Substance (substance)' SubClassOf 'amount of pure substance' or 'compound of collective material entities'	In ALL other justifications
10)	'amount of pure substance' SubClassOf 'collective material entity'	In ALL other justifications
11)	'collective material entity' SubClassOf 'material object'	In NO other justifications
12)	'compound of collective material entities' EquivalentTo compound and ('has component part' only 'collective material entity')	In 1 other justifications
13)	compound SubClassOf 'material object'	In ALL other justifications
14)	'Immune response, function (observable entity)' SubClassOf 'Immunologic function (observable entity)'	In ALL other justifications
15)	'Immunologic function (observable entity)' SubClassOf 'Function (observable entity)'	In ALL other justifications
16)	'Function (observable entity)' SubClassOf 'Observable entity (observable entity)'	In ALL other justifications
17)	'Observable entity (observable entity)' SubClassOf 'information object'	In ALL other justifications
18)	DisjointClasses: disposition, 'immaterial object', 'information object', 'material object', process, quality, role, 'temporal region', 'value region'	In ALL other justifications

Protégé explanation function

- ▶ Debugging of unsatisfiable expressions requires expertise and consumes time
- ▶ Classification time
 - ▶ avg 15 min for modules in which all patterns are represented

Explanation for: 'Allergic sensitization by patch test (disorder)' EquivalentTo Nothing

- 1) 'Allergic sensitization by patch test (disorder)' **EquivalentTo** 'Complication of patch testing
- 2) 'Complication of patch testing (disorder)' **EquivalentTo** 'Complication of diagnostic pro
- 3) 'Patch test (procedure)' **SubClassOf** 'Type 4 hypersensitivity skin test (procedure
- 4) 'Type 4 hypersensitivity skin test (procedure)' **SubClassOf** 'Delayed hypers
- 5) 'Delayed hypersensitivity skin test (procedure)' **SubClassOf** 'In vivo te
Direct (attribute)' **some** 'Skin structure (body structure)'))
- 6) 'In vivo test of hypersensitivity (procedure)' **SubClassOf** 'Clinical i
- 7) 'Clinical immunological test (procedure)' **SubClassOf** 'Immun
- 8) 'Component (attribute)' **Range** 'Substance (substance)
- 9) 'Substance (substance)' **SubClassOf** 'amount of p
- 10) 'amount of pure substance' **SubClassOf** 'coll
- 11) 'collective material entity' **SubClassOf** 'r
- 12) 'compound of collective material entities' **Equ**
- 13) compound **SubClassOf** 'material object'
- 14) 'Immune response, function (observable entity)' **SubCl**
- 15) 'Immunologic function (observable entity)' **SubCla**
- 16) 'Function (observable entity)' **SubClassOf** 'O
- 17) 'Observable entity (observable entity)' **S**
- 18) **DisjointClasses:** disposition, 'immaterial object', 'information object', 'material object', pro

Performance issues

Explanation for: 'Allergic sensitization by patch test (disorder)' EquivalentTo Nothing

1)	'Allergic sensitization by patch test (disorder)' EquivalentTo 'Complication of patch testing (disorder)' and 'Allergic sensitization (disorder)' and ('Role group (attribute)' some ('Causative agent (attribute)' some 'Patch test substance (substance)))	In ALL other justifications
2)	'Complication of patch testing (disorder)' EquivalentTo 'Complication of diagnostic procedure (disorder)' and ('Role group (attribute)' some ('Associated with (attribute)' some 'Patch test (procedure)))	In ALL other justifications
3)	'Patch test (procedure)' SubClassOf 'Type 4 hypersensitivity skin test (procedure)' and 'Test for allergens (procedure)'	In ALL other justifications
4)	'Type 4 hypersensitivity skin test (procedure)' SubClassOf 'Delayed hypersensitivity skin test (procedure)'	In ALL other justifications
5)	'Delayed hypersensitivity skin test (procedure)' SubClassOf 'In vivo test of hypersensitivity (procedure)' and 'Skin test (procedure)' and ('Role group (attribute)' some (('Method (attribute)' some 'Evaluation - action (qualifier value)')) and ('Procedure site (attribute)' some 'Skin structure (body structure)'))	In ALL other justifications
6)	'In vivo test of hypersensitivity (procedure)' SubClassOf 'Clinical immunological test (procedure)'	In ALL other justifications
7)	'Clinical immunological test (procedure)' SubClassOf 'Immunologic procedure (procedure)' and ('Role group (attribute)' some ('Component (attribute)' some 'Immune response, function (observable entity)))	In ALL other justifications
8)	'Component (attribute)' Range 'Substance (substance)'	In ALL other justifications
9)	'Substance (substance)' SubClassOf 'amount of pure substance' or 'compound of collective material entities'	In ALL other justifications
10)	'amount of pure substance' SubClassOf 'collective material entity'	In ALL other justifications
11)	'collective material entity' SubClassOf 'material object'	In NO other justifications
12)	'compound of collective material entities' EquivalentTo compound and ('has component part' only 'collective material entity')	In 1 other justifications
13)	compound SubClassOf 'material object'	In ALL other justifications
14)	'Immune response, function (observable entity)' SubClassOf 'Immunologic function (observable entity)'	In ALL other justifications
15)	'Immunologic function (observable entity)' SubClassOf 'Function (observable entity)'	In ALL other justifications
16)	'Function (observable entity)' SubClassOf 'Observable entity (observable entity)'	In ALL other justifications
17)	'Observable entity (observable entity)' SubClassOf 'information object'	In ALL other justifications
18)	DisjointClasses: disposition, 'immaterial object', 'information object', 'material object', process, quality, role, 'temporal region', 'value region'	In ALL other justifications

Protégé explanation function

- ▶ Debugging of unsatisfiable expressions requires expertise and consumes time
- ▶ Classification time
 - ▶ avg 15 min for modules in which all patterns are represented

Conclusion and Outlook

- ▶ Work in progress
 - ▶ shows feasibility of workflow
 - ▶ supports Rector & Brandt's (2008) arguments for a more expressive representation language
- ▶ Mappings depend on individual judgement by experts in both content and ontology
- ▶ Ontological commitment of SNOMED CT classes still under discussion (e.g. Findings / disorders as "Clinical Life Phases")
- ▶ Multitude of relations is a legacy issue - reduction to much lesser relations possible without loss of precision
- ▶ Satisfiability check with constraining upper-level axioms → potentially useful quality check in content production