Complex knowledge graphs for semantic analysis of clinical narratives

First draft 2020-05-15

NLP group @ IMBI Freiburg

Background

- Need for improved semantic analysis of clinical narratives for
  - Information retrieval
  - Decision support
- Availability of huge amount of data
  - Ontologies (SNOMED CT)
  - Other clinical vocabularies
  - Clinical texts
  - Medicine-related publications (scientific, educational , encyclopaedical)
- Tools and technology
  - NLP tools
  - Neural machine learning approaches (embeddings, transformer-based language models)

Ontology

- Axiomatic descriptions of generalizable truths about entities of a domain, shared by all users of the ontology
- Language-independent (apart from human-understandable labels)
- Can be expressed as graphs
- In medicine: SNOMED CT with the largest coverage
- Examples:
  - SCTID 51868009 – Ulcer of duodenum:
    Class that has all individual duodenal ulcers as members
  - SCTID 43706004 – Vitamin C
    Class that has all individual amounts of Vitamin C as members
  - Axioms:
    - Every ulcer of duodenum is part of the duodenum
    - Every amount of vitamin C is an amount of sugar acid

Reality

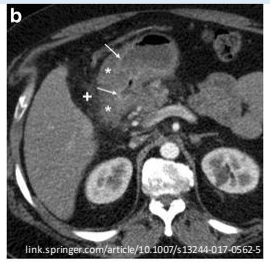
- Everything that happens in a medical or health context, e.g. the Vitamin C tablet Mrs X took this morning or the ulcer in the duodenum of Mr. Y
- Normally referred to by clinical narratives, less often in structured databases, even less frequently annotated with codes (codes often just for billing, with known biases)

Language

- Human language: Vocabulary and grammar used for communication. Highly specialised. Sublanguages like the medical language, with elements from several natural languages. Convergence between sublanguages of different natural languages

Ontology

Reality



Language

„ 05.02.2028 GKCT-Befund:
im Bereich des Pylorus und proximalen Duodenum ödematöse Wandverdickung -> könnte auf ein Ulcus hinweisen“

Ontology

39204006 | Pyloric ulcer (disorder)|

51868009 | Ulcer of duodenum (disorder)|

Finding site some

Finding site some

Associated_morphology some

Associated_morphology some

38848004 | Duodenal structure (body structure)|

SubclassOf

280119005| Pyloric structure of stomach (body structure)|

56734009 | Structure of superior portion of duodenum (body structure)|

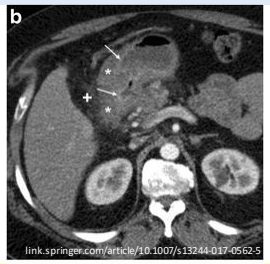
79654002 |Edema (morphologic abnormality)|

263899003 |Thickened (qualifier value)|

56208002 |Ulcer (morphologic abnormality)|

7196007 | Suggestive of (attribute)|

Reality



i24356

i99854

i8800

i93404

i77163

i76897

Language

„ 05.02.2028 GKCT-Befund:
im Bereich des Pylorus und proximalen Duodenum ödematöse Wandverdickung -> könnte auf ein Ulcus hinweisen“

Ontology

Reality

Language

Associated_morphology some

25374005 |Gastroenteritis (disorder)|

56208002 |Ulcer (morphologic abnormality)|

≡

404684003 |Clinical finding (finding)|

Has_interpretation some

Associated_morphology some

7196007 | Suggestive of (attribute)|

79654002 |Edema (morphologic abnormality)|

263899003 |Thickened (qualifier value)|

i99854

i09896

i8800

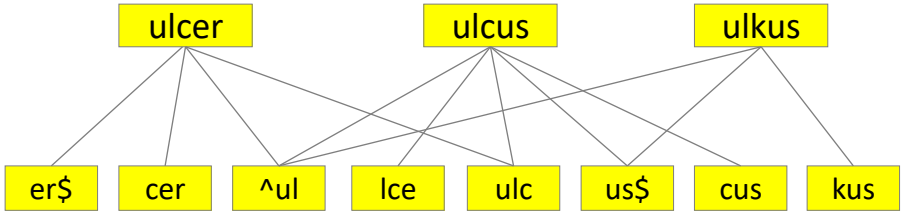
i77163

i98884

„CT interpretation identified findings consistent with peptic gastro-duodenitis, namely oedematous mural thickening“

Graph structures

- Ontology
  - Multi-taxonomy of concepts by transitive is-a relation
  - Other relations between concepts (correspond to existentially quantified OWL object properties)
- Reality (represented by NLP-generated ontology annotations)
  - Sequence of ontology codes representing instantiations of the ontology (real-world objects)
  - Some structure provided by the NLP system (e.g. scope of negation, document sections, attribute-value pairs
  - Hidden relations between instances
  - Interweaving of existential fact statements and predications (including uncertainty statements)
- (Human) Language
  - Vocabulary linked to ontology
  - Grammatical structures and language phenomena
    - anaphora: "im proximalen Dudenum …. Ulcus" means "im proximalen Dudenum …. Ulcus des Duodenum"
    - ellipsis: "mural thickening" means "thickening of the duodenal wall")
  - Cross-lingual similarities
    - Words are linked to word fragments (subwords / morphemes / n-grams)
    - The same fragments are linked to by similar words in different natural languages



Project ideas

- Purpose:
  - Finding similar graphs
  - Enhancing existing graphs
  - Detect and repair weaknesses / errors in NLP output
- Next steps
  - Identify major gaps of current NLP output
  - Check state-of-the-art of Knowledge Graphs
  - Identify interesting research question in that area and assess whether they can be used to improve medical NLP / semantic interpretation
  - Check suitable templates for semantic text content extracts, e.g. FHIR
  - Test suitability of methods (graph embeddings, BERT, etc.)