

Mini Project

Creating a medical termnet



**Bettina, Erik, Robert
Thierry (T)**

Project Overview

Project goals:

- Dataset extension that enables:
 - transitive medical concept hierarchies
 - formalization of semantic attributes and features of the medical concepts
 - lexicalization of basic medical concepts for discovering known and generating unknown complex medical concepts

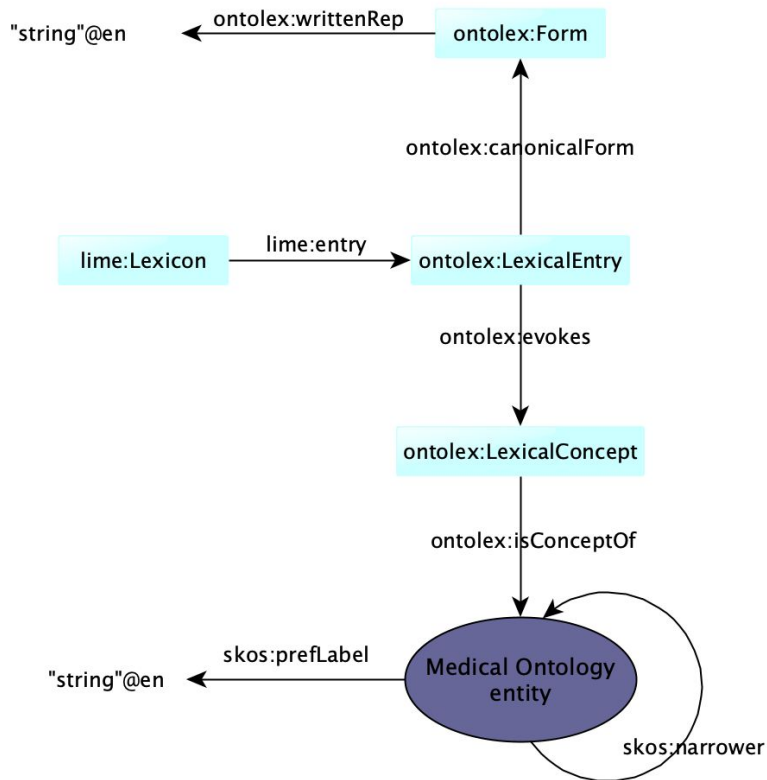
Initial plan: Creating an RDF Wordnet for representing medical domain data

→ *Result: ontolex lexicon Medical Termnet sufficient -> no Wordnet structure applied*

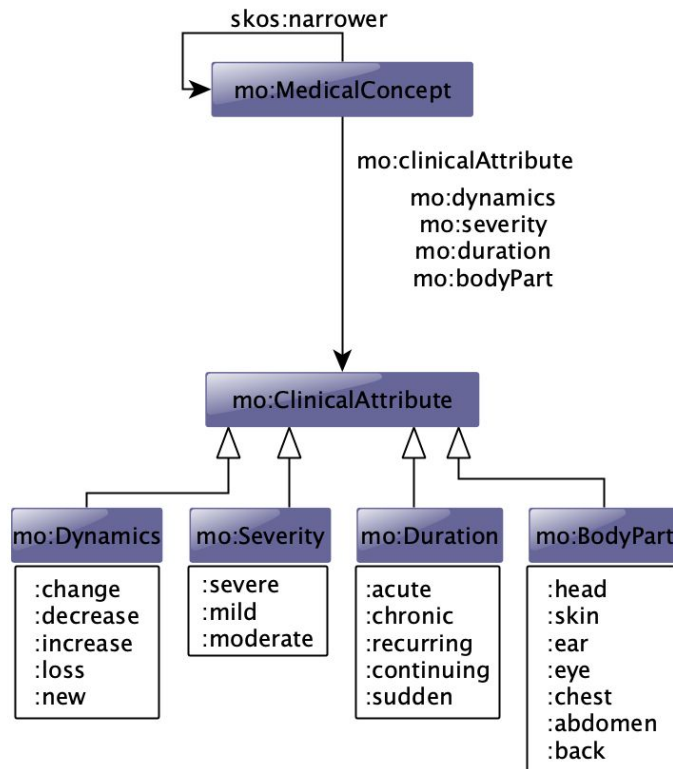
Results:

- Creation of ontolex lexicon: Medical Termnet
- Creation of domain ontology: medical ontology with interrelated concepts
- Application of string-match based entity tagging over medical corpus
- Usage of VocBench 3 for data editing and querying

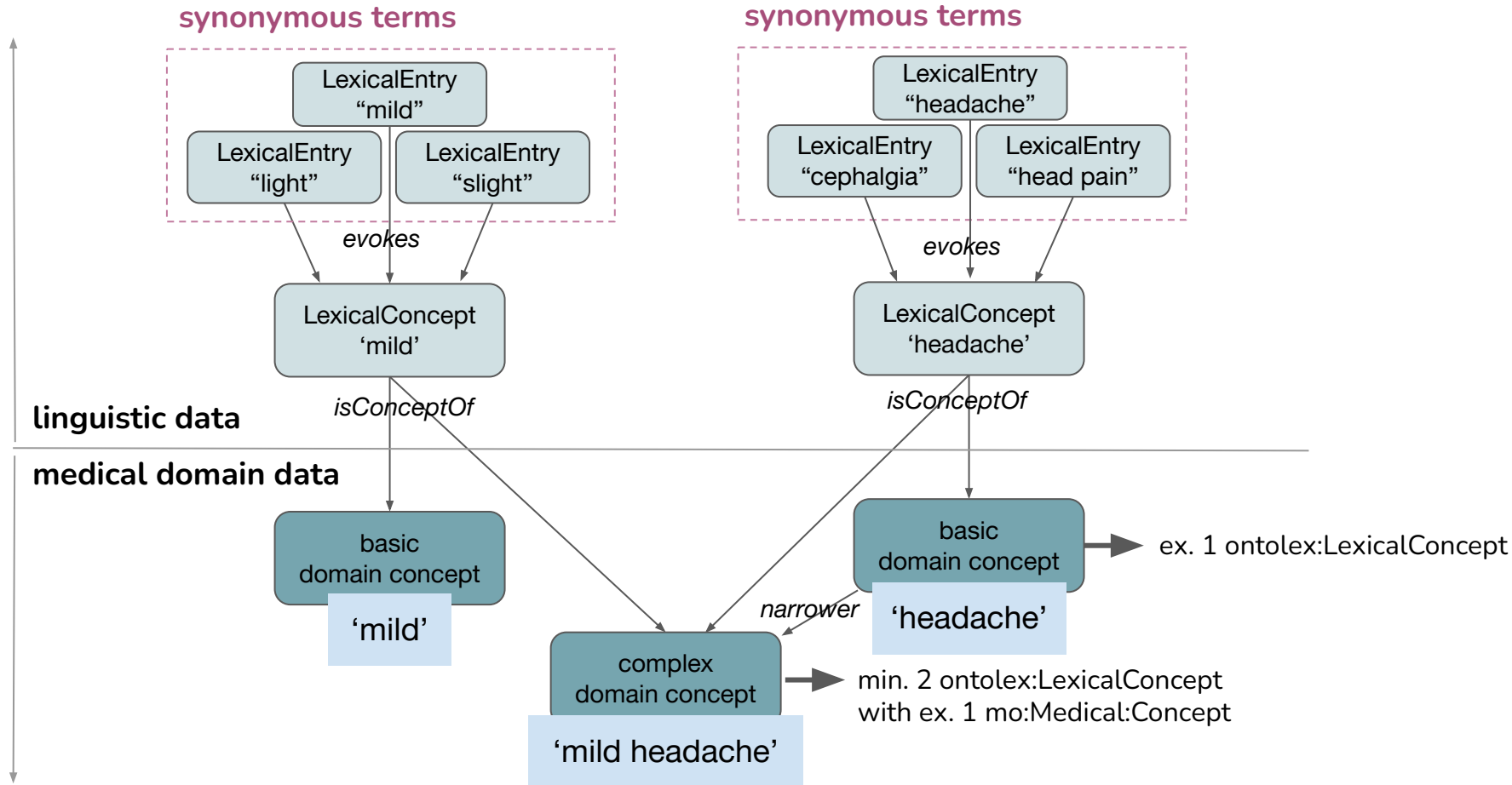
Model: Medical Termnet



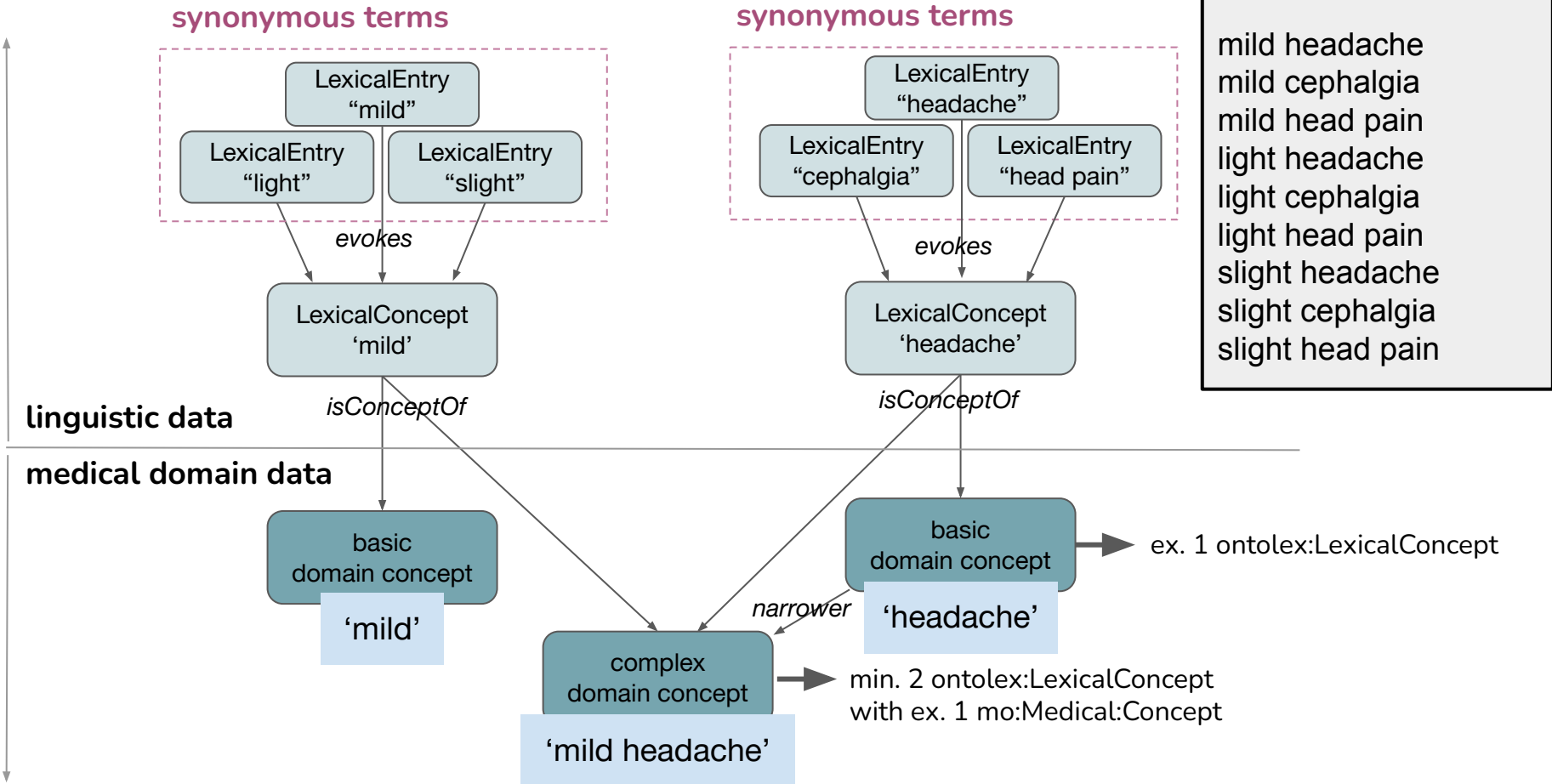
Model: Medical Ontology



Language data-enhanced domain concept identification



Language data-enhanced domain concept identification



Resulting dataset

Initial dataset contained:

- non-interrelated `mo:MedicalConcept` instances that were linked to `ontolex:LexicalEntry` via `ontolex:denotes`
- `ontolex:LexicalForm` instances with `ontolex:writtenRep` term strings

'abdominal pain'

abdominal pain
stomach aches
severe or continuing stomach pain
flank pain
upper right abdominal pain

Resulting dataset:

- creation of `ontolex:LexicalConcept` instances
- creation of data lexicalizing some of the `mo:MedicalConcept` and all the `mo:ClinicalAttribute` concepts
- creation of 17 complex `mo:MedicalConcept` instances
- interrelation of `mo:MedicalConcept` instances with `mo:ClinicalAttribute` instances
- interrelation of `mo:MedicalConcept` instances with `skos:narrower`

q Unnamed query *

☐ Fetch prefix from prefix.cc

```
21 PREFIX terms: <http://purl.org/dc/terms/>
22 PREFIX vann: <http://purl.org/vocab/vann/>
23 PREFIX vartrans: <http://www.w3.org/ns/lemon/vartrans#>
24 PREFIX void: <http://rdfs.org/ns/void#>
25 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
26
27 ▾ SELECT * WHERE {
28     ?s skos:narrower *<http://www.example.org/medical-ontology/Sudden_severe_back_pain> .
```

Submit

Clear



Include inferred (Done in 7 millisec)

s



<http://www.example.org/medical-ontology/Sudden_severe_back_pain>

<http://www.example.org/medical-ontology/Sudden_back_pain>

<http://www.example.org/medical-ontology/Back_pain>

<http://www.example.org/medical-ontology/Pain>

Resulting dataset

Initial dataset contained:

- non-interrelated `mo:MedicalConcept` instances that were linked to `ontolex:LexicalEntry` via `ontolex:denotes`
- `ontolex:LexicalForm` instances with `ontolex:writtenRep` term strings

'abdominal pain'

abdominal pain
stomach aches
severe or continuing stomach pain
flank pain
upper right abdominal pain

Resulting dataset:

- creation of `ontolex:LexicalConcept` instances
- creation of data lexicalizing some of the `mo:MedicalConcept` and all the `mo:ClinicalAttribute` concepts
- creation of 17 complex `mo:MedicalConcept` instances
- interrelation of `mo:MedicalConcept` instances with `mo:ClinicalAttribute` instances
- interrelation of `mo:MedicalConcept` instances with `skos:narrower`

!Dataset not completed but sufficient to be applied as a proof-of-concept!

Discovery of new concepts

Idea

- NLP+ML to discover known and unknown concept-attribute combinations

Steps

- Manually create list of medical symptoms/concepts and attributes
- Train a NER-tagger with entities for CONCEPT, SEVERITY, LOCUS, CAUSE, ...
- Find phrases of multiple adjacent tags
- Create new concepts with [ontolex:isConceptOf](#) relation

Data source: <https://allenai.org/data/cord-19>

- 2020-05-12 (2GB); PDF extracts → 51.868 documents

Implementation & Results

Training

- Simple direct string match for **spaCy NER-Tagger** training data
→ 20.000 annotated sentences ~ 20min

Predicting (NER)

- Filtering out all unknown matches, requiring both a medical concept and attributes
- Combinations of medical attributes without annotated concept → discovery of new/missing concepts; vice versa

Results: ~1,5M sentences

3.827 (u: 467)	found (1 concept + 1 attr., combination 2+)
6.751 (u: 698)	all (+ single occurrences)
7.056 (u: 711)	with known
204.679 (u: 116)	only known from ontology

```
@prefix mo: <http://www.example.org/medical-ontology/> .  
@prefix ontolex: <http://www.w3.org/ns/lemon/ontolex#> .
```

```
mo:Arthralgia ontolex:isConceptOf mo:Chronic_Arthralgia,  
mo:Persistent_Arthralgia,  
mo:Prolonged_Arthralgia,  
mo:Severe_Arthralgia .
```

```
mo:Arthritis ontolex:isConceptOf mo:Chronic_Arthritis,  
mo:Idiopathic_Arthritis .
```

```
mo:Bleeding ontolex:isConceptOf mo:Acute_Bleeding,  
mo:Acute_Hemorrhage,  
mo:Decrease_Bleeding,  
mo:Major_Bleeding,  
mo:Mild_Bleeding,  
mo:Minor_Hemorrhage,  
mo:Moderate_Bleeding,  
mo:Persistent_Bleeding,  
mo:Prolonged_Bleeding,  
mo:SeriouS_Bleeding,  
mo:Severe_Bleeding,  
mo:Severe_Hemorrhage .
```

Initial Investigations

Vocabulary choice

- SKOS
- LexInfo
- OntoLex vartrans and frac module
- Wordnet RDF

Tools

- VocBench 3
- “evoke” (*Exploring vocabularies and the concepts their words evoke*) tool

Project goals

- Semantic interrelation of existing 104 :Symptom instances
- Concept extension of :Symptom concepts to create an exhaustive hierarchy
- Formalization of semantic parameters and features of the medical domain (e.g. by defining custom restrictions)

Optional:

- Programmatic :LexicalEntry data extension of the 246 ‘terms’
- Add more linguistic data