Observational Health Data Sciences and Informatics (OHDSI)

Natural Language Processing (NLP) Working Group

11/14/2018

AGENDA

- Introduction of new members
- Updates on ongoing projects
  - Rules for defining term_exists – led by Stephane Meystre
  - Mapping of CUIs to standard terminology – Juan Banda
  - Mapping of Note Types to LOINC/standard vocabulary – Karthik Natarajan, Ruth Reeves, Jon Duke and Hua Xu– Report type list discussion
  - Landscape Analysis of section identifier systems and proposal of a standard terminology for use – Hua Xu, Karthik Natarajan
  - Examples and rules for term_temporal – led by George Hripsack (Sunny)
  - Standardization of term_modifiers and values – Hua Xu

NOTE_NLP & NOTE

Field	Required	Type
note_nlp_id	Yes	Big Integer
note_id	Yes	integer
section_concept_id	No	integer
snippet	No	varchar(250)
offset	No	varchar(50)
lexical_variant	Yes	varchar(250)
note_nlp_concept_id	No	integer
note_nlp_source_concept_id	no	integer
nlp_system	No	varchar(250)
nlp_date	Yes	date
nlp_date_time	No	datetime
term_exists	No	varchar(1)
term_temporal	No	varchar(50)
term_modifiers	No	varchar(2000)

Field	Required	Type
note_id	Yes	integer
person_id	Yes	integer
note_date	Yes	date
note_datetime	No	datetime
note_type_concept_id	Yes	integer
note_class_concept_id	Yes	integer
note_title	No	varchar(250)
note_text	Yes	RBDMS dependent text
encoding_concept_id	Yes	integer
language_concept_id	Yes	integer
provider_id	No	integer
note_source_value	No	varchar(50)
visit_occurrence_id	No	integer

Normalization of the term_modifier

- 1. define the modifier list;
- 2. define normalized value of each modifier;

term_modifiers

General:
- Negated
- Subject
- Certainty
- Conditional
- Temporal
- discoveryTechnique??
- Confidence??

Disease specific:
- Body location
- Severity
- Course

Medication specific:
- Dose
- Form
- Frequency
- Duration
- Route
- Necessary
- Unit

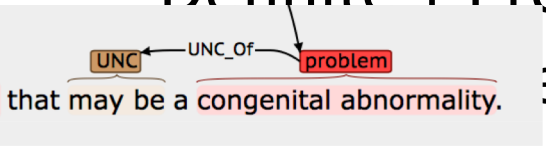
Test specific:
- Value
- Unit
- Reference Range

Procedure specific
- Body location
- Device
- Duration
- Method

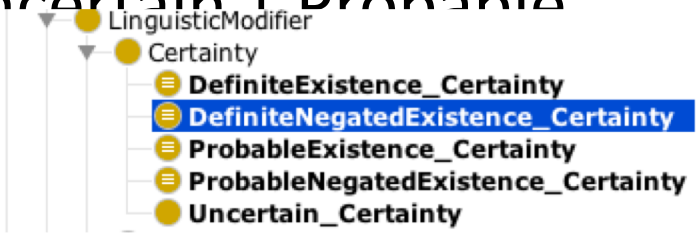
General

- ***Negation***, Boolean; True means negated;
- ***Certainty***, String Enum;
  - Definite | Probable | Uncertain | Probable Negated | Definite Negated

Example in text



Reference

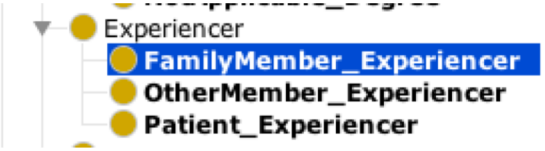


Question: merge negation into certainty?

General

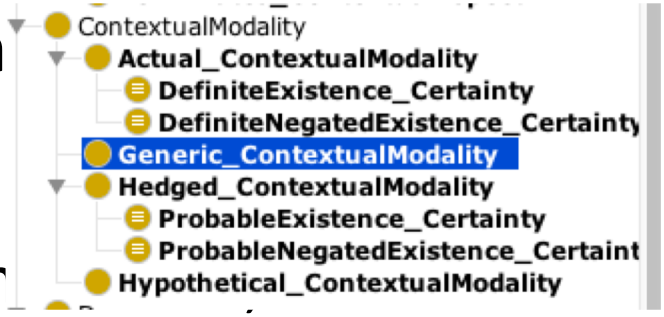
- ***Subject***, String Enum;
  - Question: granularity?
    - *Family | Other | Patient* **Or** *father | mother | roommate | friend | patient…*

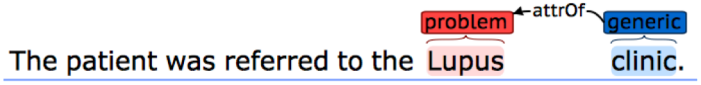
Reference

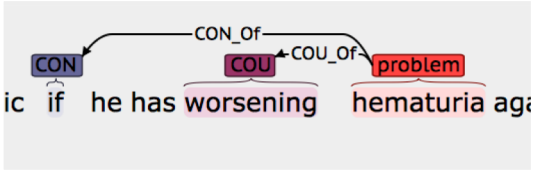


General

- ***Generic***, boolean;
  - True means generic mention
- ***Conditional***, boolean;
  - True when has conditional m

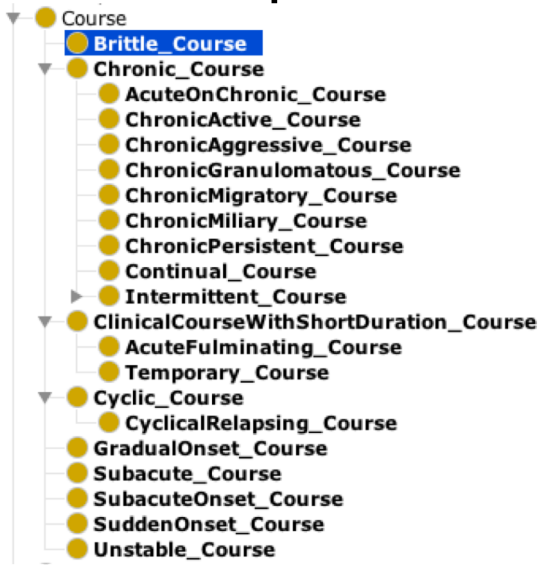






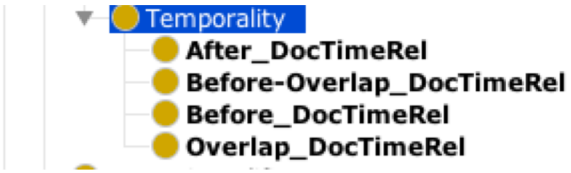
General

- ***Course***
  - unmarked | changed | increased | decreased | improved | worsened | resolved



General

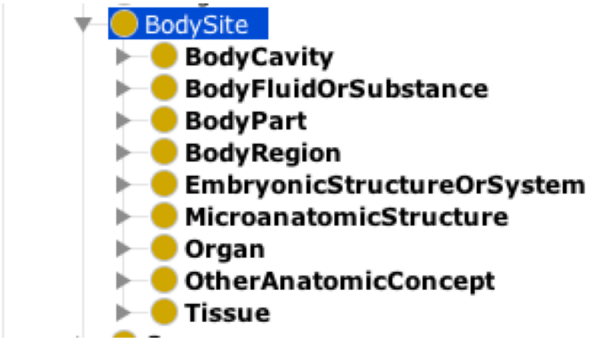
- ***Temporal***, String;
  - Date, Time, Duration, Frequency;

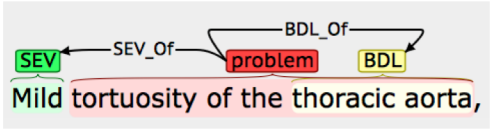


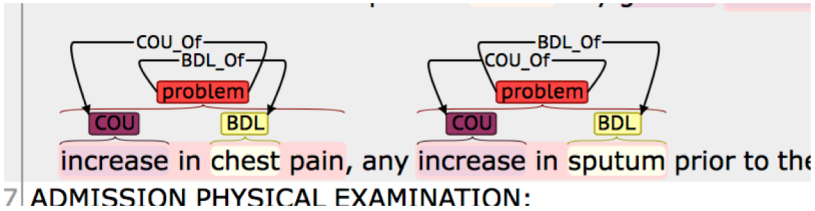
- ***Discovery technique***, String?
- ***Confidence score***, double?

Disease specific

- Body location
  - Map to UMLS CUI | standard
- Severity
  - unmarked | slight | moderate | severe | life fatal

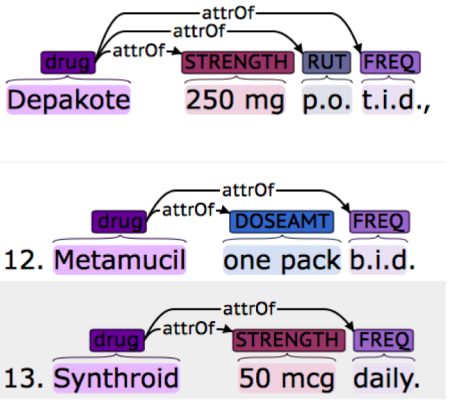






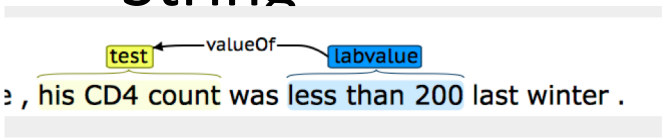
Medication specific

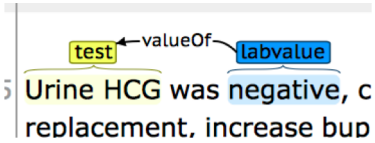
- Medication specific:
  - Dose, double
  - Frequency, String
  - Duration, String
  - Form, String enum
  - Route, String enum
  - Unit, Standard concept id?
  - Necessary
  - Start date?
  - End date?

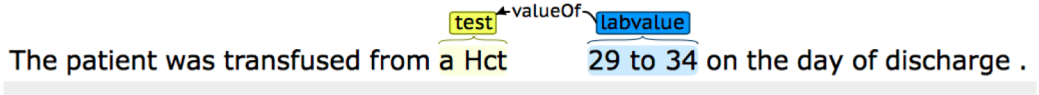


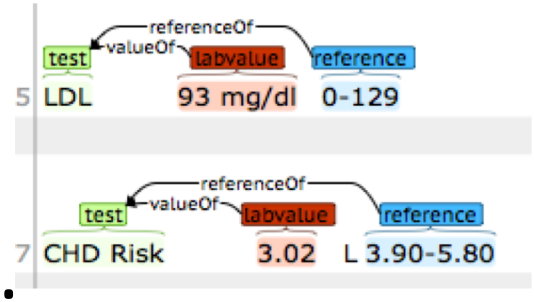
Test specific

- Value, String;
  - Value could be int, double, percentage, range, or String









- Unit, String enum;
- Reference Range, String;

Procedure specific

- Body location
  - Map to UMLS CUI | standard
- Device
  - Map to UMLS CUI | standard vocabulary;
- Duration, String
- Method, String