

**Observational Health Data
Sciences and Informatics (OHDSI)**

**Natural Language Processing (NLP)
Working Group**

06/13/2018

AGENDA

- ❖ Introduction of new members
- ❖ Presentation – Examples and rules for term_temporal – led by George Hripsack (Sunny)
- ❖ 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

OHDSI NLP-WG

term_temporal Implementation

Ning “Sunny” Shang

George Hripcsak

Department of Biomedical Informatics, Columbia University

June 13th, 2018

Definition of term_temporal

- V5.2 (5.2, 1-Feb-2017) Term_temporal is to indicate if a condition is “present” or just in the “past”
 - The following would be past: History = true; Concept_date = anything before the time of the report
- V5.3 An optional time modifier associated with the extracted term. (for now “past” or “present” only). Standardize it later.
- Proposed values
 - Default is nothing (blank), which means possibly in the present
 - PAST means the text asserted occurred in the past and not the present

cTAKES 4.0

109.1.1.1.4.	org.apache.ctakes.typesystem.type.syntax.BaseToken
109.1.1.1.4.1.	org.apache.ctakes.typesystem.type.syntax.ContractionToken*
109.1.1.1.4.2.	org.apache.ctakes.typesystem.type.syntax.NewlineToken*
109.1.1.1.4.3.	org.apache.ctakes.typesystem.type.syntax.NumToken*
109.1.1.1.4.4.	org.apache.ctakes.typesystem.type.syntax.PunctuationToken*
109.1.1.1.4.5.	org.apache.ctakes.typesystem.type.syntax.SymbolToken*
109.1.1.1.4.6.	org.apache.ctakes.typesystem.type.syntax.WordToken*
...	
109.1.1.1.9.	org.apache.ctakes.typesystem.type.syntax.ConllDependencyNode*

109.1.1.1.22.	org.apache.ctakes.typesystem.type.textsem.IdentifiedAnnotation
...	
109.1.1.1.22.2.	org.apache.ctakes.typesystem.type.textsem.DateAnnotation*
109.1.1.1.22.3.	org.apache.ctakes.typesystem.type.textsem.EntityMention
109.1.1.1.22.3.1.	org.apache.ctakes.typesystem.type.textsem.AnatomicalSiteMention*
109.1.1.1.22.4.	org.apache.ctakes.typesystem.type.textsem.EventMention*
109.1.1.1.22.4.1.	org.apache.ctakes.typesystem.type.textsem.DiseaseDisorderMention*
109.1.1.1.22.4.2.	org.apache.ctakes.typesystem.type.textsem.LabMention
109.1.1.1.22.4.3.	org.apache.ctakes.typesystem.type.textsem.MedicationEventMention
109.1.1.1.22.4.4.	org.apache.ctakes.typesystem.type.textsem.MedicationMention*
109.1.1.1.22.4.5.	org.apache.ctakes.typesystem.type.textsem.ProcedureMention*
109.1.1.1.22.4.6.	org.apache.ctakes.typesystem.type.textsem.SignSymptomMention*
109.1.1.1.22.5.	org.apache.ctakes.typesystem.type.textsem.FractionAnnotation*
109.1.1.1.22.6.	org.apache.ctakes.typesystem.type.textsem.MeasurementAnnotation*
...	
109.1.1.1.22.12.	org.apache.ctakes.typesystem.type.textsem.PersonTitleAnnotation*
109.1.1.1.22.13.	org.apache.ctakes.typesystem.type.textsem.RangeAnnotation*
109.1.1.1.22.14.	org.apache.ctakes.typesystem.type.textsem.RomanNumeralAnnotation*
109.1.1.1.22.15.	org.apache.ctakes.smokingstatus.type.SmokerNamedEntityAnnotation
109.1.1.1.22.16.	org.apache.ctakes.typesystem.type.textsem.TimeAnnotation
109.1.1.1.22.17.	org.apache.ctakes.typesystem.type.textsem.TimeMention*

109.1.1.1.34.	org.apache.ctakes.typesystem.type.textspan.Segment*
...	
109.1.1.1.36.	org.apache.ctakes.typesystem.type.textspan.Sentence*
...	
109.1.5.	org.apache.ctakes.typesystem.type.structured.DocumentID*

109.1.6.4.	org.apache.ctakes.typesystem.type.refsem.Event*
109.1.6.4.1.	org.apache.ctakes.typesystem.type.refsem.DiseaseDisorder
109.1.6.4.2.	org.apache.ctakes.typesystem.type.refsem.Lab
109.1.6.4.3.	org.apache.ctakes.typesystem.type.refsem.Medication
109.1.6.4.4.	org.apache.ctakes.typesystem.type.refsem.Procedure
109.1.6.4.5.	org.apache.ctakes.typesystem.type.refsem.SignSymptom
...	
109.1.7.	org.apache.ctakes.typesystem.type.refsem.EventProperties*
...	
109.1.11.	org.apache.ctakes.typesystem.type.refsem.OntologyConcept
109.1.11.1.	org.apache.ctakes.typesystem.type.refsem.UmlsConcept*

109.1.15.2.	org.apache.ctakes.typesystem.type.relation.BinaryTextRelation
...	
109.1.15.2.11.	org.apache.ctakes.typesystem.type.relation.TemporalTextRelation*
...	
109.1.16.	org.apache.ctakes.typesystem.type.relation.RelationArgument*

cTAKES 4.0

- cTAKES temporal modules derive a timeline in
 - EntityMention/EventMention
 - IdentifiedAnnotation class
 - Assertion attribute: historyOf (0 or 1, 1=history of)
 - Subclass of both Event and IdentifiedAnnotation class (DiseaseDisorderMention, Lab(?), MedicationMention, ProcedureMention, SignSymptomMention)
 - Extract the temporal relation between the target event and the document creation time (DCT): “before the DCT”, “after the DCT”, or “overlap the DCT”
 - Temporal relation (TLINK): TemporalTextRelation
 - Detect within-sentence Event-Time relations and Event-Event relations
 - Extracting pairwise events which have temporal relations (e.g. CONTAINS, BEFORE, etc.)

cTAKES rules for EntityMention/EventMention

- EntityMention/EventMention
 - term_temporal = PAST if Eventmention/EntityMention attribute
 - history = true OR
 - history = false AND doctimerel= BEFORE
 - term_temporal = BLANK if Eventmention/EntityMention attribute
 - history = false AND doctimerel = OVERLAP or NULL OR AFTER

attribute	doctimerel	term temporal
history=[false],	doctimerel=[OVERLAP]	
history=[false]		
history=[false],	doctimerel=[BEFORE]	PAST (contradictory)
history=[false],	doctimerel=[AFTER]	
history=[false],	doctimerel=[BEFORE/OVERLAP]	
history=[true],	doctimerel=[OVERLAP]	(contradictory)
history=[true],	doctimerel=[BEFORE]	PAST
history=[true]		PAST
history=[true],	doctimerel=[AFTER]	(contradictory)

cTAKES TemporalTextRelation (TLINK)

- TLINK:
 - arg1: TIMEX, _ref_arg2: event, category attribute
 - category=CONTAINS, BEFORE, BEGINS-ON, ENDS-ON

Arg1 (TIMEX)	Arg2	Sentence	Category attribute
3/26/2006	MRA	5. On 3/26/2006, MRI/MRA of the neck and brain, with and without contrast:	category=CONTAINS
daily	lisinopril	She will continue lisinopril daily.	category=CONTAINS
Discharge	hemoglobin	Discharge hemoglobin 11.4, and hematocrit 33.3.	category=CONTAINS
at that time	vomiting	However, he did not have any vomiting at that time.	category=CONTAINS
Postop	pain	Postop pain was adequately managed with the aid of Acute Pain team.	category=CONTAINS
smoking	three to four months previously	The patient has been smoking up until three to four months previously.	category=ENDS-ON
temperatures	103 for 8-10 days	She had been having temperatures of up to 103 for 8-10 days.	category=ENDS-ON
approximately eight hours.	urine	24 hours later, the Foley catheter was removed and a bladder scan demonstrated 60 cc of residual urine after approximately eight hours.	category=BEFORE

cTAKES rules for TemporalTextRelation (TLINK)

- CONTAINS
 - If date then assume PAST or check the document creation time (DCT)
 - If “at that time” probably PAST; maybe other phrases
 - If “postop” or similar leave BLANK
- ENDS-ON
 - Rare, but appears to be PAST
- BEFORE, BEGINS-ON, etc.
 - Rare, leave BLANK

CLAMP

- CLAMP temporal modules derive a timeline in
 - ClampNameEntityUIMA semanticTag temporal
 - ClampRelationUIMA semanticTag timeof
 - entFrom: TIMEX (ClampNameEntityUIMA semanticTag temporal)
 - entTo: primary entity (ClampNameEntityUIMA semanticTag problem/test/treatment)

entFrom (TIMEX)	entTo	Sentence
April 2008	confusion	HISTORY OF PRESENT ILLNESS: Ms. ABC is a pleasant 60-year-old Caucasian female who was diagnosed in April 2008 with diffuse large B-cell lymphoma after she developed visual saltation, changes, and confusion.
August 31, 2007	Port-A-Cath placement	On August 31, 2007, Port-A-Cath placement.
that time	chest pain	She denied history of chest pain or fevers or cough with purulent sputum at that time.
postoperative day	carcinomatosis	discharge summary, carcinomatosis from appendiceal primary, tablet by mouth, stent placement, postoperative day, ureteral stents, peritoneal carcinomatosis, appendiceal primary, discharge, admitted, insertion, hemicolectomy, splenectomy, peritoneal, appendiceal, postoperatively, carcinomatosis, oncology,
the day of discharge	hemoglobin	On the day of discharge, the patient was pancytopenic with white count of 0.7, ANC of 500, hemoglobin 8.5, hematocrit 24.8, and platelet count 38, 000.

CLAMP rules for ClampRelationUIMA

- If date then assume PAST or check the document creation time (DCT)
- If “that time” probably PAST; maybe other phrases
- If “postoperative day” or similar leave BLANK

MetaMap term_temporal

- Only concepts extracted, term_temporal is BLANK

MedLEE

- MedLEE annotates temporal information in
 - `<status>`, `<timeper>`, `<date>`, `<reltime>`

```

<problem v="cancer" code="UMLS:C0006826_malignant neoplasms" idref="p62">
  <bodyloc v="bladder" code="UMLS:C0005682_urinary bladder" idref="p60">
    <code v="UMLS:C0005682_urinary bladder" idref="p60"/>
  </bodyloc>
  <certainty v="high certainty" idref="p64"/>
  <parsemode v="mode2"/>
  <reltime v="2 year" idref="p68"/>
  <sectname v="report unknown section item"/>
  <sid idref="s1"/>
  <status v="history" idref="p56"/>
  <code v="UMLS:C0006826_malignant neoplasms" idref="p62"/>
  <code v="UMLS:C0699885_carcinoma of bladder" idref="p60 p62"/>
  <code v="UMLS:C0005684_malignant neoplasm of urinary bladder" idref="p60 p62"/>
</problem>
<procedure v="extubation" code="UMLS:C0553891_removal of endotracheal tube" idref="p480">
  <certainty v="high certainty" idref="p478"/>
  <date v="00000206" idref="p484">
    <reltime v="on" idref="p482"/>
  </date>
  <parsemode v="mode2"/>
  <sectname v="report hospital course item"/>
  <sid idref="s21"/>
  <code v="UMLS:C0553891_removal of endotracheal tube" idref="p480"/>
  <cocode v="UMLS:C0336630_endotracheal tube" idref="p480"/>
</procedure>

```

```

<finding v="primary" idref="p301">
  <bodyloc v="appendix"
code="UMLS:C0003617_appendix!UMLS:C1269000_entire appendix"
idref="p299">
    <code v="UMLS:C1269000_entire appendix" idref="p299"/>
    <code v="UMLS:C0003617_appendix" idref="p299"/>
  </bodyloc>
  <certainty v="high certainty" idref="p303"/>
  <parsemode v="mode4"/>
  <sectname v="report hospital course item"/>
  <sid idref="s14"/>
  <timeper v="admission" idref="p305"/>
</finding>

```

MedLEE term_temporal rules

- <status>
 - if value is “history”, “past history” then PAST, if value is “now”, “future”, “postoperative” then leave BLANK
- <timeper>, <date>, <reltime>
 - If date then assume PAST or check the document creation time (DCT)
 - If “discharge”, “admission” or similar leave BLANK