Towards principles of ontology-based annotation of clinical narratives

Stefan Schulz¹,²
Warren Del-Pinto³
Markus Kreuzthaler¹
Lifeng Han³
Sareh Aghaei¹
Goran Nenadic³

¹Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria
²Averbis GmbH, Freiburg, Germany
³Department of Computer Science, University of Manchester, UK
Most information in health records is “locked” in narratives

... in the local languages / sociolects

Paciente G1PO, IG de 38 sem 4 dia(s), TS A+, interna por bolsa roita há mais de 18hs, recebendo penicilina. Evoluí para Parto Eutócico com episiotomia em 27/06/2007 22:24 hs. Nasce RN APGAR 10/10, MASC, 3060 G. Exames: Toxo IGG e IGM neg VDRL neg EQU neg UROC: ausência de crescimento bacteriano. Hemograma 198mil plaq; Hb 13,1; LT 12,5 (75% seg) Em condições de alta, amamentando, útero contraído, lóquios fisiológico, sinais vitais estáveis, FO com bom aspecto. Recebe as orientações abaixo. ORIENTAÇÕES NA ALTA: # AMAMENTAÇÃO EXCLUSIVA POR 6 MESES; # TOMAR AS MEDICAÇÕES PRESCRITAS (SULFATO FERROSO 300MG 3X/DIA POR 90 DIAS, LONGE DAS REFISÇÕES, COM SUCO DE LARANJA; PARACETAMOL 750 MG 6/6HS SE DOR); # ORIENTO ANTICONCEPÇÃO; # RETORNAR À EMERGÊNCIA DESTE HOSPITAL SE FEBRE, SANGRAMENTO AUMENTADO OU OUTRAS INTERCORRÊNCIAS. # NÃO É NECESSÁRIO RETIRAR OS PONTOS. # LAVAR FO 3X/DIA COM ÁGUA E SABÃO DE GLICERINA.

* Anamnese und klinische Symptomatik
* Physikalischer Status
48 jähr.Patient, deutl. reduz. AZ, normaler EZ. Cor: Ht rh, nc, Systolikum mit p.max. über dem Erbschen Punkt mit Fortleitung in die Axila
Pulmo: VA bds., feuchte RGs re>li
Abdomen: BD weich, kein DS
Extremitäten: ausgeprägte Knöchelödeme bds.
Herr DI Max Mustermann wurde aufgrund einer neuerlichen Dyspnieszsymptomatik bei bek. dilat. CMP und hochgrad. MINS zur weiteren Evaluierung stat. vom LKH Fürstenfeld übernommen.
Clinical language: compact, sloppy, contextualised

- Works well for expert-to-expert communication

<table>
<thead>
<tr>
<th>Phenomenon</th>
<th>Example</th>
<th>Elucidation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Telegram style</td>
<td>“left PICA stroke, presented to ED after fall”</td>
<td>Incomplete sentences, sketchy style</td>
</tr>
<tr>
<td>Colloquialisms</td>
<td>“pothole sign”, “snorkel”</td>
<td>Milieu-specific sub-languages</td>
</tr>
<tr>
<td>Ad-hoc abbreviations</td>
<td>“infiltr”</td>
<td>Truncation (&quot;infiltrated mucosa&quot;)</td>
</tr>
<tr>
<td>Ambiguous short forms</td>
<td>“RTA”</td>
<td>“Road traffic accident”, “Renal-tubular acidosis”</td>
</tr>
<tr>
<td>Short forms of regional or local scope</td>
<td>“LDS Hospital”</td>
<td>“Latter-Day-Saints Hospital” (and not “Leak Detection System”)</td>
</tr>
<tr>
<td></td>
<td>“St. p.”</td>
<td>“Status post” = “History of”</td>
</tr>
<tr>
<td>Conventionalized Latin abbreviations</td>
<td>“V mors can dig V dext”</td>
<td>“Vulnus morsum canis digitu quinti dextri” (in some European languages)</td>
</tr>
<tr>
<td>Numeric codes</td>
<td>“45, 46 with crowns”, “VI palsy”, “2-2-2”</td>
<td>Tooth numbers, cranial nerves, dose frequencies</td>
</tr>
<tr>
<td>Spelling errors, typos</td>
<td>“Diabetes”, “Astra-Seneca”, “Hipotireose”,</td>
<td>accidental (quick typing) or systematic (e.g. 2nd language speakers)</td>
</tr>
<tr>
<td></td>
<td>“Esophagus”, “Oesophagus”</td>
<td>e.g. American vs. British English</td>
</tr>
<tr>
<td>Single noun compounds</td>
<td>“Ibuprofenintoxikation”</td>
<td>Non-lexicalized long words (in languages such as German, Swedish)</td>
</tr>
<tr>
<td>Anaphora</td>
<td>(i) “adenoCa rect pN+MX G2 (…). tumor excised in toto”</td>
<td>(i) “Tumor” coreferential to adenocarcinom described in left context</td>
</tr>
<tr>
<td></td>
<td>(ii) “no blood in stomach (…). mult mucosal erosions ”</td>
<td>(ii) “mucosal erosions” refined to “erosions of gastric mucosa”</td>
</tr>
<tr>
<td>Negations</td>
<td>“No evidence of pneumonia”</td>
<td>non-standard, jargon-like</td>
</tr>
<tr>
<td></td>
<td>“Pulmones: nihil”, “metastasenfrei”</td>
<td></td>
</tr>
<tr>
<td>Epistemic contexts</td>
<td>“suspect MI, DD lung embolism”</td>
<td>suspected diagnosis, differential diagnosis</td>
</tr>
<tr>
<td></td>
<td>“h/o Covid-19”, “Streptokokkenangina 06/16”</td>
<td>“history of”</td>
</tr>
<tr>
<td></td>
<td>(i) father: pancreas ca”</td>
<td>(i) family history</td>
</tr>
<tr>
<td></td>
<td>(ii) “refrained from resuscitation”</td>
<td>(ii) plans not executed</td>
</tr>
</tbody>
</table>

- Major interoperability bottleneck for machine processing
Desideratum: making unstructured health record data interoperable

- Using international standards
- Rooted in Applied Ontology principles
- Information extraction via NLP (Natural language processing)

Physical examination on admission revealed purpura of the upper and lower extremities, **swelling of the gums and tonsils**, but no symptoms showing the complication of myasthenia gravis. Hematological tests revealed leucocytosis: WBC count 68,700/µl (blasts 11.5%, myelocytes 0.5%, bands 2.0%, segments 16.0%, monocytes 65.5%, lymphocytes 4.0%, atypical lymphocytes 0.5%), Hb 7.1 g/dl (reticulocytes 12%) and a platelet count of 9.1 × 10^4/µl. A bone marrow aspiration revealed hypercellular bone marrow with a decreased number of erythroblasts and megakaryocytes and an increased number of monoblasts.

**SNOMED CT** large clinical ontology
(350k concepts, > 1M English terms)
Clinical knowledge graph as canonical content representations

- **Primary use:**
  - Document Retrieval
  - Question Answering
  - Content summarization
  - Information extraction
  - Input for decision support
  - Data visualization, navigation

- **Secondary use:**
  - Retrospective research
  - Patient recruiting
  - Training of predictive models
  - Training of decision support systems
  - Quality improvement

---

NLP methods
- Classical pipelines (sequential processing steps), rules, lexicons
- End-to-end approaches: single architecture directly maps input text to the desired output: Deep learning, large language models

Annotated clinical corpora: central resource
- Training, model fine-tuning
- Benchmarking of NLP systems
Document annotation – knowledge acquisition bottleneck

- Requirements
  - Domain expertise
  - Extensive training
  - Motivation

- Problems
  - Inter-annotator variability
  - Annotation fatigue
  - Ambiguities
  - Time constraints

- Success factors:
  - Good tooling
  - Repeated training sessions
  - Adjudication between annotators
  - Quality checks (inter-annotator agreement)
  - Good communication channels
  - **Rigorous annotation guidelines based on clinical standards and ontological principles**

![Medical Students Image](https://depositphotos.com)
Annotation guideline browser, annotation tool

Annotation guideline for semantic annotations of clinical narratives based on SNOMED CT and FHIR

Stefan Schütz1, Ahkida Naz Kapusselsery4, Alexander Beger3, Saeed Aghasi2, Daniel Dürr1, Larissa Hamme2, Kristi Karkesiner2, Markus Krupilzatliner1, NN, NN.

1Medical University of Graz, Austria
2Avertis GmbH, Freiburg, Germany
3Tallinn University of Technology, Estonia

Version 2023-08-14

1. Introduction
2. Background
2.1. Annotation strategies
2.2. Related work
3. Objectives
4. Tools and resources
4.1. Naming and graphical conventions
4.2. Annotation vocabularies
4.2.1. SNOMED CT
4.2.2. HL7 FHIR
4.2.3. LOINC
4.3. INCePTION
5. Basic assumptions and decisions
5.1. Conceptual model
5.2. Principles
5.3. Preferences regarding SNOMED content
5.3.1. Core (focus) concepts
5.3.2. Modifying concepts
5.3.3. Product concepts
5.3.4. Concepts from HL7 value sets
5.3.5. Combination of annotations
5.4. Concepts that require values
5.5. Relations, predicates, and operators

Annotation guideline for semantic annotations of clinical narratives based on SNOMED CT and FHIR

Stefan Schulz1,2, Akhilla Naz Kuppassery3, Alexander Beger1, Sareh Agheel4, Daniel Dur5, Larissa Hammer1, Kristian Kankainen5, Markus Kriezthaler2, NN, NN...

1Medical University of Graz, Austria  
2Averhos GmbH, Freiburg, Germany  
3Tallin University of Technology, Estonia

Version 2023-06-14

1. Introduction 1  
2. Background 2  
2.1. Annotation strategies 2  
2.2. Related work 4  
3. Objectives 5  
4. Tools and resources 6  
4.1. Naming and graphical conventions 6  
4.2. Annotation vocabularies 6  
4.2.1. SNOMED CT 6  
4.2.2. HL-7 FHIR 6  
4.2.3. LOINC 7  
4.3. INCEPTION 7  
5. Basic assumptions and decisions 7  
5.1. Conceptual model 7  
5.2. Principles 7  
5.3. Preferences regarding SNOMED content 10  
5.3.1. Core (focus) concepts 1  
5.3.2. Modifying concepts 12  
5.3.3. Product concepts 12  
5.3.4. Concepts from HL7 value sets 12  
5.3.5. Combination of annotations 13  
5.4. Concepts that require values 14  
5.5. Relations, predicates, and operators 16  

General annotation principles

- Semantic annotation only (no POS, syntactic relations etc.)
- Annotation at two levels
  - Text spans (“entities”) with codes and literals
  - Binary relations with user-friendly predicates (hide complexity from annotators): (i) semantic relations (ii) co-reference annotations
- Annotation vocabularies linked to ontology-based standards
  - SNOMED CT, FHIR, HPO, RxNorm ...
  - multilingual, well-curated, free, ontology-based, compositional (post-coordination)
- Annotation vocabulary determines
  - Annotation spans (subword to multiword): longest match preference
  - Granularity and scope
- Close-to-text annotation
  - no interpretation by annotators
Specific annotation principles (SNOMED CT + FHIR annotation)

- “Core” hierarchies:
  - Clinical finding, Event, Observable entity, Pharmaceutical / biologic product, Procedure, Specimen
  - High proportion of fully defined concepts

- “Supportive” hierarchies
  - Substance, Organism, Body structure, Physical object, Qualifier
  - Primitive concepts
  - HL7-FHIR values sets mapped to SNOMED concepts

- Annotation predicates (binary relations) link “core” concepts to “supportive” concepts, grounded in
  - SNOMED CT object properties or chains thereof,
  - Relational chains of FHIR elements
  - both
Relation annotation vocabulary based on SNOMED CT and FHIR

- Close-to-user predicates
- Mapped to relations or relation chains in underlying standards

<table>
<thead>
<tr>
<th>anno</th>
<th>Domain</th>
<th>Target path</th>
<th>Range</th>
</tr>
</thead>
</table>
| site         | ‘sct:Clinical finding’ | [a] ‘sct:Finding site’  
[b] INV(fhir:Condition.code) || fhir:Condition.body       | ‘sct:Body structure’         |
| site         | ‘sct:Procedure’         | [a] ‘sct:Procedure - Direct’  
| inFamily     | ‘sct:Clinical finding’  | [b] INV(fhir:FamilyMemberHistory.condition) || fhir:FamilyMemberHistory.relationship  
[a] INV(‘sct:Associated finding’)  
|| ‘sct:Subject relationship context’ | ‘sct:Person’                |
| verification status | ‘sct:Clinical finding’ | [b] INV(fhir:Condition.code) || fhir:ConditionverificationStatus  
[a] INV(‘sct:Associated finding’)  
|| ‘sct:Finding context’ | ‘sct:Qualifier value’  
(cf. Tab. 1)
Example 1: “Two level” annotation

1. Text spans, annotated with codes or literals
2. Linkage of text spans by binary predicates

anno:site

397181002 | Open fracture (disorder) | 734143007 | Structure of left femur (body structure) |
Open fracture of left femur

anno:beginTime

The patient had a heart attack on Dec 3, 2021
Example 2: Deep annotation

Annotations exploit the whole depth of the annotation vocabulary
No “entity-type” annotation

Not:

<table>
<thead>
<tr>
<th>Disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unilateral conjunctivitis right</td>
</tr>
</tbody>
</table>
Example 3: Flexible annotation spans / longest match principle

- Annotations spans determined by annotation vocabulary
- Preference given to longest match (precoordinated concepts)
- No determination of spans by NER before annotation

28576007 |Open fracture of femur (disorder)|
The femur exhibited an open fracture
Example 3: Flexible annotation spans / longest match principle

- Annotations spans determined by annotation vocabulary
- Preference given to longest match (precoordinated concepts)
- No determination of spans by NER before annotation

28576007 |Open fracture of femur (disorder)|
The femur exhibited an open fracture

anno:site

76505004 |Thumb structure (body structure)|
The thumb

had an

397181002 |Open fracture (disorder)|
exposed fracture
Example 4: Close-to-text: no interpretation of content

- Only annotate what is explicitly stated, not what might be medically plausible
Example 5: Coreference annotations

- Nominal anaphora

```
A sarcoma was diagnosed. The tumor ...
```

![Diagram showing coreference annotations with entities 424413001 and 108369006 connected by anno:sameAs relation.]

From text to canonical representation

Text

Annotations

Knowledge graph

- Words
  - phrases
  - numeric expressions

- Codes for classes
- Annotation predicates
- literals

- Semantic interpretations
- Classes vs. Individuals
- “Isosemantic” representations
  - (e.g., SNOMED only vs. SNOMED + FHIR)
| Text level | “Suspected breast cancer” | “Mother: breast cancer” | “Diagnosis: breast cancer” |
Conclusion and outlook

- Annotated corpora are essential for training and benchmarking NLP tools, particularly in the current era of deep learning and large language models.

- Semantic resources / ontologz-based standards are crucial:
  - Ontologies (description of entity types): Definitions / Axioms
  - Terminologies (description of natural language): labels, synonyms
  - Information Models (Instance-level templates, link to ontologies and values)

- Clinical free text annotation is a huge and challenging task. Facilitated by
  - Pre-annotations using existing NLP annotators
  - Simple, intuitive set of predicates that map to more complex graph structures in the background

- Adherence to detailed guideline principles
  - might take a long journey
  - Indispensable for high agreement between annotators → canonical clinical content representations
Contact: Stefan Schulz
stefan.schulz@medunigraz.at
http://purl.org/steschu

Comment on our annotation guideline:

Acknowledgements:
- Grant 101057062 “AIDAVA” (funder: the European Commission, HORIZON-HLTH-2021,
- Grant “Assembling the Data Jigsaw: Powering Robust Research on the Causes, Determinants and Outcomes of MSK Disease” (funder: The Nuffield Foundation)
- Grant EP/V047949/1 “Integrating hospital outpatient letters into the healthcare data space” (funder: UKRI/EPSRC).
Semantic equivalences

Clinical Summary > Condition

Condition

verificationStatus | unconfirmed
---|---

39026006 | Suspected gallstones (situation) |

Condition

verificationStatus | unconfirmed
---|---

23591008 | Gallbladder calculus (disorder) |

Condition

verificationStatus | unconfirmed
---|---

313413008 | Calculus finding (finding) |

Condition

verificationStatus | unconfirmed
---|---

3578005 | Structure of body of gallbladder (body structure) |

Condition

verificationStatus | unconfirmed
---|---

56381008 | Calculus (morphologic abnormality) |

Condition

verificationStatus | unconfirmed
---|---

3578005 | Structure of body of gallbladder (body structure) |

EquivalentClasses:

| 1759001 | Disease suspected (situation) |

ObjectIntersectionOf:

| 446523005 | Suspected clinical finding (situation) |

ObjectSomeValuesFrom:

| 56381008 | Calculus (morphologic abnormality) |

ObjectIntersectionOf (ObjectSomeValuesFrom):

| 246960004 | Associated finding (attribute) |

ObjectSomeValuesFrom:

| 408729009 | Finding context (attribute) |

ObjectSomeValuesFrom:

| 408729009 | Finding context (attribute) |

ObjectSomeValuesFrom:

| 408729009 | Temporal context (attribute) |

ObjectSomeValuesFrom:

| 408729009 | Subject relationship context (attribute) |

ObjectSomeValuesFrom:

| 408729009 | Subject relationship context (attribute) |
Open issue: Identity management

126926005 | Neoplasm of breast (disorder) |
NoB34u73axn4

254837009 | Malignant neoplasm of breast (disorder) |
MoBkj88935el

278054005 | Lobular carcinoma of breast (disorder) |
LCBrrp009g65t

713609000 | Invasive carcinoma of breast (disorder) |
ILCB4tz5pplkll

Lobular carcinoma of breast
Invasive Lobular carcinoma of breast