The Role of Ontologies for Sustainable, Semantically Interoperable and Trustworthy EHR Solutions

Semantic enrichment of narrative EHR content

Stefan SCHULZ
Institute of Medical Biometry and Medical Informatics
University Medical Center, Freiburg, Germany
Structure of the talk

- The role of natural language in the electronic health record
  - different types of narratives and different ways of authoring
  - advantages and disadvantages of narrative content
  - need of semantic enrichment of the EHR

- Target representations for semantically enriched EHRs
  - ontologies are not enough
  - context is essential

- Language engineering for semantic EHR enrichment:
  Technical challenges
The importance of narrative EHR content
<table>
<thead>
<tr>
<th>Datum</th>
<th>Krankheits-Geschichte</th>
<th>Datum</th>
<th>Krankheits-Geschichte</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: S. Freud’s handwritten clinical notes
<table>
<thead>
<tr>
<th>Datum</th>
<th>Uhr</th>
<th>Pflegebericht</th>
<th>Verlaufsbeschreibung</th>
<th>Krankenbeobachtung</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.4.18</td>
<td>18:45</td>
<td>Pat. kam mit einer Kompressionsjacke mit</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>keine Verletzungen auffällig</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22.4.18</td>
<td>3:45</td>
<td>Pat. lag in der NOCT mit Komfort</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>geräumig, sie gab ab</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22.4.18</td>
<td>14:00</td>
<td>Pat. st. Pt. Pflegeplan verabschied</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pat. st. Plan abgegeben, ist m'l Beobachtung</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>geschaffen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.4.18</td>
<td>2:30</td>
<td>Pat. schläft seit geraumer Zeit nur mit</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>einem Lombok und einem leinen Schiefer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.4.18</td>
<td>16:00</td>
<td>Pat. hatte 3x Gang bis auf flüssigen Stuhl (Corinna)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weiserg. Li. Plan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.4.18</td>
<td>23:00</td>
<td>Pat. kam 30 Minuten erwischt, hält sich nicht an</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ihre Bettgröße</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25.4.18</td>
<td>03:30</td>
<td>Pat. musste aus der komischen Gepäck</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Vater verstorben an Bronchial-Karzinom, Mutter verstorben an den Folgen einer Pneumonie. Mutter Diabetes mellitus. 5 gesunde Kinder.

**Systemanamnese:**

**Soziale Anamnese:**
Früher Arbeiterin in der Elektronikbrache, dann Hausfrau, verheiratet, lebt mit dem Ehemann zusammen.

**Allergien:** Keine bekannt.

**Medikation bei Aufnahme:**
Ulcogant 1-1-1, Pepdul mit 0-0-0-1, Cellcept 2x1 g, Bayotensin 3x 1, Cynt 0,2 1x1, Ludiomil 50 mg 1x1, Sandimmun 2x150 mg, Clexane 0,4 ml 1x täglich s.c.

**Status bei Übernahme:**

Source: Universitätsklinikum Freiburg, Germany
# Erotilde, 58 anos # HAS # Obesidade

Pcte interna com hist de edema e dor em MID há 3 semanas. Refere que no inicio do quadro apresentava hiperemia local importante, nega febre. Foi avaliada no posto de saúde e iniciado amoxacilina. Fez uso do ATB por 6 dias, com piora da dor, do edema, e surgimento de lesões arredondadas, planas, com bordos bem definido e pequeno ponto escurecido central. Consultou novamente no posto de saude, snedo trocado o ATB para eritromicina, o qual usou por mais 6 dias, com piora das lesões e da dor. Na avaliação inicial a pcte apresentava edema e hiperemia importante de MID, bem como lesões ulceradas, necróticas com bordos bem definidos, sem secreção. Foi realizado ECO doppler que confirmou TVP em MID sendo então iniciada anticoagulação com enoxaheparina. Solicitada consultoria da dermato que realizou biopsia das lesões

Narratives produced by speech recognition
<table>
<thead>
<tr>
<th></th>
<th>Narrative Content</th>
<th>Structured Content</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Electronic Health Record</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ease of production</strong></td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><strong>Quality</strong></td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><strong>Human communication</strong></td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><strong>Machine communication</strong></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td><strong>Data analysis</strong></td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>
how to bridge this gap...?

Narrative Content
- finding reports
- progress notes
- discharge summaries

Structured Content
- terminologies
- classifications
- ontologies
- information models
Semantic enrichment of text using natural language technologies
Semantic enrichment of text using Natural Language Technologies

Narrative Content

How to correctly extract information from medical texts

Structured Content

How to find an interoperable semantic representation formalism
Semantic enrichment of text using Natural Language Technologies

Narrative Content

How to correctly extract information from medical texts

Structured Content

How to find an interoperable semantic representation formalism
Semantic enrichment of text using Natural Language Technologies

Narrative Content

- How to correctly extract information from medical texts

Structured Content

- How to find an interoperable semantic representation formalism
The importance of natural language in the EHR
Semantic enrichment: Target representations
Semantic enrichment: Technical challenges
Semantic enrichment: technical challenges

- Manual annotation / coding
- Automated annotation / coding
- Semantic enrichment at the point of authoring
From narrative to structured content

• Semantic enrichment
  – Annotating text (or other data objects) by in ways that boost the value of the text
  – Semantic identifiers
    • ad-hoc tags
    • controlled terms
    • classes from ontologies and information models

• Strategies
  – Manual annotation / coding
  – Automated annotation / coding
  – Semantic enrichment at the point of authoring
Manual semantic enrichment

- **Examples**
  - MeSH indexing of Medline abstracts
  - Manual assignment of disease and procedure codes to the EHR
Manual semantic enrichment

- Problems
  - Time consuming
  - Requires specific training
  - Motivation gap / bias:
    - undercoding
    - overcoding
    - miscoding
  - limited scope:
    - procedures
    - diseases
your bill is correct, Sir… well, the operation lasted only ten minutes, but then our doctor took two hours finding the right procedure code
Automated semantic enrichment

- Uses natural language processing technology
- Commonly used term: text mining
- Two paradigms
  - Document retrieval
    - Special case: term retrieval
  - Information extraction
Document retrieval scenario

query
expression
Document retrieval scenario

query expression

relevant documents
Document retrieval scenario

query expression

search engine

relevant documents

ranked list of documents considered relevant by the search engine
Term retrieval as a special case of document retrieval

AP: vasculite leucocitoclástica. A pt recebê ciprofloxacín por 5 días e após 2 días de oxaciina. Recebe alta em bom estado geral, com diminuição importante do edema e da dor em MID. Lesões em fase de cicatrização. Revisada a literatura: existe associação de vasculite em áreas de estase, bem como associação com
Term retrieval scenario

AP: vasculite leucocitoclástica. A pte recebeu ciprofloxacín por 5 dias e após 2 dias de oxacicilina. Recebe alta em bom estado geral, com diminuição importante do edema e da dor em MID. Lesões em fase de cicatrização. Revisada a literatura: existe associação de vasculite em áreas de estase, bem como associação com
AP: vasculite leucocitoclastica. A pente recebeu ciprofloxacina por 5 dias e após 2 dias de oxacilina. Recebe alta em bom estado geral, com diminuição importante do edema e da dor em MID. Lesões em fase de cicatrização. Revisada a literatura; existe associação de vasculite em áreas de estase, bem como associação com

ranked list of terms considered matching candidates by the search engine
shadow was pointed out on a routine chest X-ray film, but she had no further examination. 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 leukocytosis: 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 × 10⁴/µl. Further laboratory examination revealed elevated serum lactic dehydrogenase (589 U/l), vitamin B₁₂ (2010 pg/ml) and ferritin (650. 0 ng/ml). Human chorionic gonadotropin and [alpha]-fetoprotein levels were normal. A bone marrow aspiration revealed hypercellular bone marrow with a decreased number of erythroblasts and megakaryocytes and an increased number of monoblasts that were positive for staining by [alpha]-naphthyl butyrate esterase and negative for staining by naphthol ASD chloroacetate esterase. Chest X-ray upon admission revealed a mediastinal mass and an elevated left diaphragm. Computed tomography (CT) of the chest showed a left anterior mediastinal mass. Based on these findings, the patient was diagnosed with a mediastinal tumor accompanied by AMoL. First, in June 1991, the patient was treated with DCMP therapy: daunorubicin (DNR) (25 mg/m², days 1, 2, 3, 4, 6 and 8), cytosine arabinoside (Ara-C) (100 mg/m², days 1-7), 6MP (70 mg/m², days 1-9) and prednisolone (PSL) (20 mg/m², days 1-9), followed by five courses of consolidation chemotherapy [1, DCMP; 2, ID-Ara-C; 3, DCMP; 4, ID-Ara-C; 5, A-triple V: Ara-C, VP-16, VCR, vinblastine (VBL)]. After induction chemotherapy, a hematological examination and bone marrow findings had improved to normal, and complete remission was attained. Chest CT scan after chemotherapy in November 1991 revealed regression of the mediastinal tumor. An invasive thymic tumor was suspected and surgery was undertaken in January 1992. The tumor (50 × 45 × 45 mm), located mainly in the anterior mediastinum, was strongly adhered to the adjacent tissues. Resection of the tumor included the left upper lobe of the lung, the phrenic nerve and pericardium. The histological finding was that the tumor cells have large, vesicular nuclei and prominent nucleoli, but keratinization was unclear. The results of immunohistochemical finding of anti-TdT was negative. From these findings, we diagnosed poorly or moderately differentiated squamous cell carcinoma of the thymus. The postoperative course was uneventful. The patient underwent radiation therapy of the mediastinum and left hilum at doses of 4000 cGy delivered in 10 weeks. She was discharged in March 1992. After the first AMoL remission, the patient suffered a relapse six times and was repeatedly admitted for chemotherapy. During these periods, chest X-ray and CT revealed no recurrence of the mediastinal tumor. During her tenth admission, the patient developed pneumonia during chemotherapy and died in October 1996. No autopsy was performed.
Language processing techniques

• from simple pattern matching...
  "do*" \rightarrow "do", "dog", "done", "doctor", etc.

• to more sophisticated techniques, using
  – lexical knowledge
  – grammatical knowledge
  – domain knowledge
  – empirical knowledge (e.g. annotated corpora)
  – AI approaches
  – statistical approaches
Morphology (Stemmer, Lemmatizer)

Lexicon

Syntax (POS Tagger, Chunker, Parser)

POS/Tree Bank

Semantics (NE Recognizer, Proposition Analyzer)

Proposition Bank

Domain Ontology

Pipeline for NLP Analysis

«An acute infection ended the pregnancy»

example: Udo Hahn, Jena
Possible outputs of sophisticated language processing

- Noun phrases (term candidates)
- Predicate-argument structures
- Classification of named entities
- Attachment of prepositional phrases
- Scope of negations
- Anaphora resolution
- Discourse analysis
- Etc...

Text “understanding” is still the holy grail of computational linguistics
Specific challenges for medical language processing

• High lexical productivity
  – single-word compounds “hyperparathyroidism”
  – acronyms and abbreviations, ambiguous and context-dependent

• Heterogenous document style
  – telegram style
  – enumerations (e.g. lab values)
  – embedded tables

• Low writing quality
  – persisting errors (spelling, punctuation, case, accents…)

• Implicit contexts
Semantic enrichment of text using Natural Language Technologies

Narrative Content

How to correctly extract information from medical texts

Structured Content

How to find an interoperable semantic representation formalism
Naïve approach: content representation by instantiation of ontologies

- Identify term of interest $T$ in a text
- Retrieve a suitable class in the ontology using term retrieval
- Interpret the mention of this term as the reference to a member of this class
- i.e.
  $T$ mentioned in text = there is some referent of $T$ in reality
Paciente interna por quadro de abdome agudo. TC de abdome mostrou aneurisma de aorta com evidencia de extravasamento de contraste. Levado a cirurgia de urgência, sendo realizada aneurismectomia com colocação de prótese. No pós-operatório evoluiu com síndrome da resposta inflamatória sistêmica, com disfunção de múltiplos órgãos e instabilidade hemodinâmica. Apesar do manejo com drogas vasoativas, reposição hídrica e hemodiálise veno-venosa crônica, o paciente apresentou piora progressiva, evoluindo para óbito.
Patient with incisional hernia admitted for herniorrhaphy, but operation was suspended because operation room was urgently needed for liver transplant. Discharged with orientation and rescheduled operation.
### Example: codes and contexts

<table>
<thead>
<tr>
<th>Chunk</th>
<th>Context</th>
<th>Negation</th>
<th>Code (SNOMED)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O cateter</td>
<td></td>
<td></td>
<td>19923001</td>
</tr>
<tr>
<td>foi trocado</td>
<td></td>
<td></td>
<td>103713001</td>
</tr>
<tr>
<td>por disfunção (baixo fluxo)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no mesmo sitio.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Em discussão HYP</td>
<td></td>
<td></td>
<td>235582004</td>
</tr>
<tr>
<td>papilotomia ou colecistectomia com a Gastro</td>
<td></td>
<td></td>
<td>38102005</td>
</tr>
<tr>
<td>mas como não tem cálculo</td>
<td></td>
<td>NEG</td>
<td>56381008</td>
</tr>
<tr>
<td>essa decisão será tomada posteriormente.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Em 31/01/07 apresentou PCR</td>
<td></td>
<td></td>
<td>410430005</td>
</tr>
<tr>
<td>em fibrilação ventricular logo após ter terminado a hemodiálise;</td>
<td>STP</td>
<td></td>
<td>71908006</td>
</tr>
<tr>
<td>recuperado rapidamente não ficando com sequelas.</td>
<td></td>
<td>NEG</td>
<td>362977000</td>
</tr>
</tbody>
</table>
Common contexts in medical documents:

- **Subject of record:**
  - “father had diabetes mellitus”

- **Uncertainty**
  - “…was admitted with suspected diabetes mellitus”

- **Negation**
  - “no diabetes mellitus”

- **Plan**
  - “in the case of ... patient should be checked for diabetes mellitus”
Boundary problem

Ontology

“what is” types of entities by their inherent properties

- Diabetes mellitus
- Pancreas
- Huminsulin Long
  1 Amp. 10 ml 100 I.E./ml
- Glucose serum concentration

Epistemic context

what is known / planned / hypothesized
Ontology vs. Epistemology

Ontology

Ontology

“what is”
types of entities by their inherent properties

OBO

SNOMED CT

ICD-10

Epistemic context

what is known / planned / hypothesized

HL7 V3 templates

openEHR archetypes
Information models

mix ontological with epistemic information

<table>
<thead>
<tr>
<th>Diabetes Mellitus Family History</th>
<th>yes</th>
<th>no</th>
<th>don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grandfather</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grandmother</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Father</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brother / Sister</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusion for semantic representation of clinical narratives

- Clinical narratives combine
  - reference to types of entities (ontology)
  - reference to the state of knowledge of the author (context)
- Ignore context: high risk of false assertions
- Ignore ontology: limited semantic interoperability
- Binding ontologies to information models:
  - topic of research
  - pragmatic solutions
- Using terminological systems that include both: current practice, but problematic
Outlook

• Persisting problem:
  – physicians continue producing text and coded content partly redundantly

• Challenge:
  – semantic enrichment on the fly