Annotation methods to develop and evaluate an expert system based on natural language processing in electronic medical records

Q. GICQUEL, I KERGOURLAY, S GERBIER-COLOMBAN, S CHARIOUT, A BITTAR, F SEGOND, S DARMONI, MH METZGER

MIE, Istanbul
2 September 2014
Introduction

• **The SYNODOS project**
  – collaborative 36-month project launched in October 2012

• **Objective**
  – to develop a generic solution for semantic retrieval of medical data and organize it for use in epidemiological studies or medical decision-making
  – allow medical staff to write their own expert rules independently of their domain of specialization
  – evaluate the performance of the developed solution in extracting information in two domains: hospital-acquired infections and cancer
The SYNODOS consortium

- **Coordination**
  - Lyon 1 (LBBE) : MH Metzger

- **Academic and industrial partners**
  - Industrial: HO2S, Viseo
  - Academic: CiSMeF, Lyon 1

- **Project duration**
  - 1 october 2012 – 30 september 2015

- **Funding**
  - Global costs = 2.0 M€
  - Allocated grant = 0.785 M€
General architecture of the solution
Knowledge representation: populating the database of facts

Metadata, textual data, structured data of a medical record

Surgical report: T0

Discharge summary: T0 +7 days

Consultation letter: J+3 months

“Transition” rules

Vizualisation of the restructured data in the base of facts
Objective of this presentation

to describe how we developed a conceptual model based on manual annotation of medical documents which will be used as “gold standard” for the development and evaluation of the expert system.
Excerpt of the conceptual model
Manual annotation using an application populating the « gold standard » database of facts
<table>
<thead>
<tr>
<th>Medical document after de-identification process (excerpt)</th>
<th>Vizualisation of the manual annotation in MedIndex (excerpt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consultation report [T-12D]</td>
<td>Clinical history of the current episode of care</td>
</tr>
<tr>
<td>Date of birth : [AGE_PATIENT = 73]</td>
<td>Description of clinical symptoms :</td>
</tr>
<tr>
<td>weight loss of 4 kg in 4 months</td>
<td>* weight loss of 4 kg *</td>
</tr>
<tr>
<td>Absence of transit disorder, fecal occult blood testing - , anemia</td>
<td></td>
</tr>
<tr>
<td>[T-19D] colonoscopy : large tumor, impassable, circumferential fairly typical adenocarcinoma .....</td>
<td>Consultation report <em>(type of document)</em></td>
</tr>
<tr>
<td></td>
<td>L: 3 <em>(position of the term in the document)</em></td>
</tr>
<tr>
<td></td>
<td>C0043096/Body Weight decreased <em>(CUI)</em></td>
</tr>
<tr>
<td></td>
<td>Numerical value : 4kg</td>
</tr>
<tr>
<td></td>
<td>in 4 months <em>(temporal label)</em></td>
</tr>
</tbody>
</table>
• **Originality of our annotation method**
  – not only oriented towards identifying key clinical entities in the text or determining the relationships between entities
  – identifies the medical interpretation in the context of a care pathway

• **Objective of this type of annotation**
  – Development and evaluation of an expert system based on natural language processing in electronic medical records.
  – To consider the temporality of the health event in the course of patient’s care.

• **That is why the annotation method developed here includes a manual restructuring of data extracted from medical records**
For more information : http://www.synodos.fr

This project received funding from the National Research Agency (program TecSan 2012)