Using ISO and Semantic Web standards for creating a Multilingual Medical Interface Terminology: A use case for Heart Failure

Elena Cardillo
Fondazione Bruno Kessler

Co-authors:
Maxime Warnier (UCL)
Marc Jamoulle (UCL)
Joseph Roumier (CETIC)
Robert Vander Stichele (UGENT)
Outline

- Context
- Objective
- Approach
  - Building a multilingual Reference Terminology in TMF
  - Building unilingual End-user Lexicons in LMF / Lemon
- Implementation of the use case
- Conclusion
Efficient encoding and retrieval of medical data have become two crucial needs for clinicians.

Classification systems are not well integrated in EHRs and links to them hampered by a translation gap.

Terminological solutions are often naive:
  - Too overwhelming or too restrictive in concepts
  - Not addressing local, cultural differences between vocabularies

Need for a more efficient interface between the human language and the machine language.
Objective

Provide a hybrid healthcare interface terminology capable of

“Systematic collection of healthcare-related phrases (terms) to support clinicians’ entries of patient-related information into computer programs …”

Rosembloom et al, 2006

That includes:

1. A conceptual reference terminology
   - Multidisciplinary, multilingual
   - Restricted to reference concepts
   - Linked to international nomenclatures, thesauri, classifications, ontologies

2. A series of specific end-user lexicons
   - One for each language
   - Bridged to the reference terminology
   - Connected to rich linguistic corpora for NLP
Approach

Interface Terminology

End-user Lexicons  Reference Terminology

Word

Sense 1  Sense 2

(n = 1)

Concept

Term

Semasiologic

Onomasiologic
Architecture

Interface Terminology Service

External Reference Sources
- Linguistic Corpus in L1
- Linguistic Corpus in L2
- Linguistic Corpus in L3

End-user Lexicon L1
- LMF (ISO 24613)
- Technical term

End-user Lexicon L2
- LMF (ISO 24613)
- Admitted term

End-user Lexicon L3
- LMF (ISO 24613)
- Conceptual linking

Multilingual Reference Terminology

External Reference Sources
- Classifications
- Nomenclatures
- Thesauri & Metathesauri
- Ontologies

NLP Connection
LMF/Lemon – TMF Connection
Conceptual linking

01/11/2013
Terminologie et Intelligence Artificielle - TIA 2013, Paris
TMF-based Reference Terminology

• **ISO 16642:**
  - Terminological Markup Framework
  - Resource
  - Concept
  - Language
  - Term
  - Term components

• **ISO 12620 and ISOcat:**
  - Data categories
  - Registry

11/1/2013
Terminologies et Intelligence Artificielle - TIA 2013, Paris
Lemon/LMF-based End-user Lexicons

- ISO 24613 Language Markup Framework (LMF) as a Requirement for the Lexicon
- Use of the Lemon model (LExicon Model for ONtologies)
  - Based on LMF
  - RDF native model
  - Easy to link to data category registries (ISOcat)
  - The sense of a lexical entry can be defined by a reference to an external resource (E.g. TMF-based resource)
  - More powerful and less verbose than the LMF model
  - [http://lemon-model.net/](http://lemon-model.net/) (McCrae et al.)
OWL encoding of the TMF model

• Implementation of the TMF model as a series of OWL-DL ontologies using Protégé:

  › Makes links between the first two;
  › Adds some restrictions and links between classes and individuals;
  › Defines cardinality for each data category;
  › Characterizes properties;
  › Assign domain values

  › Use of `owl:equivalentClass` to link data categories to the corresponding ones on ISOcat.
End-user Lexicon in RDF

- RDF representation of the lexicon with Lemon;
- Links between phrases and their components are provided in the form of an ordered RDF list.
- Properties come from:
  - the lemon model itself, ISOcat; Dublin Core;
- Linguistic properties:
  - POS (only common nouns, adjectives and adverbs in our lexicon);
  - Canonical form (i.e. lemma) with a possible spelling variant;
  - Inflected forms and normative authorization;
  - Grammar gender and number
- Lexical sense/s defined outside the lexicon
  - E.g. in the Reference Terminology (for medical sense);
  - Polysemy and Synonymy can be inferred
Implementation of the use case

• Extraction of relevant concepts and words or phrases from a Belgian bilingual (Dutch and French) guideline on Heart Failure (Scientific Associations of Primary Care Physicians)

• French version as a starting point

• Two different extraction process:
  1. Manual selection of the relevant concepts by domain experts to populate the Reference Terminology
  2. Automated term extraction by a bilingual term extractor to populate the End-user Lexicon
Populating the Reference Terminology 1/2

• 168 clinical concepts manually selected by a domain expert

• Concepts entered as term entries in the Term section together with:
  o a definition in French;
  o a reference to the French original guideline;
  o a preferred term (E.g. “ascite”);
  o one/more admitted terms (E.g. “eau dans le ventre”)
  o Italian translation for each preferred and admitted term (in the Italian language section)
### Populating the Reference Terminology 2/2

#### At the Terminological entry level for each concept:

- **Concept Identifier**
- **Concept Type** (E.g. Symptom, Disease, etc.)
- **Mapping to ICPC and ICD classification systems**
- **Correspondence to SNOMED-CT**
  - Fully Specified Name used as "Standardized Term" in the Term Section
- **Correspondence to UMLS Metathesaurus**
- **Definition in English**
  - UMLS definition if mapping is present
  - Other official definition from standard vocabularies
  - Translation in the other language sections

#### Table:

<table>
<thead>
<tr>
<th>UMLS Code</th>
<th>UMLS Definition</th>
<th>SNOMED-CT Code</th>
<th>SNOMED-CT Term</th>
<th>ICD 10 Code</th>
<th>ICD 10 Term</th>
<th>ICPC 2 Code</th>
<th>ICPC 2 Term</th>
<th>English Description</th>
<th>French Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C003 6916</td>
<td>Sexually Transmit</td>
<td>Disease or Syndrome</td>
<td>CSR/PT</td>
<td>8222 0001 3</td>
<td>Sexually transmitted infectious disease (disorder)</td>
<td>(A50-A64)</td>
<td>Infections with a predominant sexual mode of transmission</td>
<td>Male Genital</td>
<td>Sexuellement transmissible</td>
</tr>
<tr>
<td>C003 6916</td>
<td>Sexually Transmit</td>
<td>Disease or Syndrome</td>
<td>CSR/PT</td>
<td>8222 0001 3</td>
<td>Sexually transmitted infectious disease (disorder)</td>
<td>(A50-A64)</td>
<td>Infections with a predominant sexual mode of transmission</td>
<td>Female Genital</td>
<td>Sexuellement transmissible</td>
</tr>
<tr>
<td>C001 8891</td>
<td>Heart failure</td>
<td>Disease or Syndrome</td>
<td>CSR/PT</td>
<td>8258 9001 4</td>
<td>Heart failure (disorder)</td>
<td>I50</td>
<td>Heart failure</td>
<td>K77</td>
<td>Insuffisance cardiaque</td>
</tr>
<tr>
<td>C020 4714</td>
<td>Acute heart failure</td>
<td>Disease or Syndrome</td>
<td>-</td>
<td>7952 1301 0</td>
<td>Acute heart failure (disorder)</td>
<td>I50</td>
<td>Heart failure</td>
<td>K77</td>
<td>Insuffisance cardiaque</td>
</tr>
<tr>
<td>C205 0127</td>
<td>Heart failure with normal ejection fraction</td>
<td>Disease or Syndrome</td>
<td>-</td>
<td>2881 2110 10</td>
<td>Heart failure with normal ejection fraction (disorder)</td>
<td>I50</td>
<td>Heart failure</td>
<td>K77</td>
<td>Insuffisance cardiaque avec fraction d’éjection préservée (60-75%)</td>
</tr>
<tr>
<td>C026 4723</td>
<td>Low output heart failure</td>
<td>Disease or Syndrome</td>
<td>-</td>
<td>7558 7401 0</td>
<td>Low output heart failure (disorder)</td>
<td>I50</td>
<td>Heart failure</td>
<td>K77</td>
<td>Insuffisance cardiaque avec fraction d’éjection ventriculaire (FEV) diminuée</td>
</tr>
</tbody>
</table>
Manual mapping issues

• Some concepts are too general
• No «Symptom and Complaints» in SNOMED-CT:
  o E.g. drug side effect «dry mouth»;
• Semantic types mismatching:
  o E.g. UMLS Finding ≠ SNOMED-CT Findings;
• Many 1:m mapping to ICPC
  o E.g. «Cardiomiopathy» opened to 3 different ICPC rubrics;

<table>
<thead>
<tr>
<th>External Resources</th>
<th>Reference Term. Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICPC</td>
<td>153</td>
</tr>
<tr>
<td>ICD</td>
<td>131</td>
</tr>
<tr>
<td>SNOMED-CT</td>
<td>161</td>
</tr>
<tr>
<td>UMLS CUI</td>
<td>116</td>
</tr>
</tbody>
</table>
Populating the End-user Lexicon

- Automatic extraction using TexSIS (http://lt3.hogent.be/):
  - French/Dutch glossary of 774 words and phrases
- French terms matched to the French preferred terms in the RT
- 77/168 matched entries entered in the French End-user Lexicon in Lemon
  - Linked to the sense(s):
    - In the Reference Terminology (to the corresponding concept ID)
    - In external resources for other senses (E.g. French DBPedia)
- Admitted terms were added to extend the End-user Lexicon
  - Increase to 138 lexical entries (114/138 phrases)
- A total of 298 lexical entries after decomposition of phrases into single words
Link between the two resources
A total of 16,636 RDF triples representing the Reference Terminology for English, French, and Italian.

Publication of the data as Linked Data on the Data Hub under a CC licence, accessible via SPARQL queries.

Combined use of URL rewriting techniques and SPARQL constructs to allow retrieval of relevant triples.

E.g., http://meriterm.org/heartfailure/25/en/preferred to retrieve the English preferred term «Synus Bradicardia» of the concept with ID «25».

http://datahub.io/dataset/heartfailure
End-user Lexicon in LOD

• A total of 3,400 RDF triples to represent the French End-user Lexicon
• Published on the Data Hub as Linked Data and available also at:
  o [http://meriterm.org/heartfailure/lexicons.rdf](http://meriterm.org/heartfailure/lexicons.rdf)
• Linked to the Reference Terminology, to ISOcat and to the French DBPedia
## Statistics on the data

<table>
<thead>
<tr>
<th>Category</th>
<th>English</th>
<th>French</th>
<th>Italian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Terminology concepts</td>
<td></td>
<td>168</td>
<td></td>
</tr>
<tr>
<td>Preferred terms</td>
<td></td>
<td>168</td>
<td>168</td>
</tr>
<tr>
<td>Admitted terms</td>
<td></td>
<td>168</td>
<td>193</td>
</tr>
<tr>
<td>Standardized terms (Links to Snomed-CT)</td>
<td>161</td>
<td>161</td>
<td>161</td>
</tr>
<tr>
<td>UMLS definitions</td>
<td></td>
<td>116</td>
<td></td>
</tr>
<tr>
<td>Links to ICD-10</td>
<td></td>
<td>131</td>
<td></td>
</tr>
<tr>
<td>Links to ICPC-2</td>
<td></td>
<td>153</td>
<td></td>
</tr>
<tr>
<td>End-user Lexicon entries</td>
<td></td>
<td></td>
<td>298</td>
</tr>
<tr>
<td>Reference Terminology Triples</td>
<td></td>
<td></td>
<td>16,636</td>
</tr>
<tr>
<td>End-user Lexicon Triples</td>
<td></td>
<td></td>
<td>3,400</td>
</tr>
</tbody>
</table>
Conclusions

• Use of a hybrid approach to create a multilingual Interface Terminology in the healthcare domain
• Use of standards models for developing the terminology
  o ISO TMF for the multilingual Reference Terminology
  o Lemon / ISO LMF for unilingual End-user Lexicons
• ISOcat for the choice and labeling of the data categories
• Publication of these resources in the Semantic Web
  o Concepts and terms have a URI in the web and can be easily retrieved, integrated, linked and reused
• Useful for facilitating:
  o Medical data registration and searching of medical information by physicians
  o Translation of technical terms into lay terms for healthcare consumers
  o Translation of medical terms in other languages
Work in progress + Future works

• Use of a Semantic MediaWiki to maintain and extend our terminologies
• Create End-user Lexicon also for English, Italian, and Dutch
• Improvement of the concept/term extraction selection method
• Working on a new use case about Drugs Contraindications
• Links to other multilingual lexi-ontological resources on the web
  o E.g. Babelnet (http://babelnet.org/)
Thanks for your attention ...

Any questions?

[Emails]
cardillo@fbk.eu
www.fbk.eu
meriterm@googlegroups.com