Combining semantic and lexical methods for mapping MedDRA to VCM icons

Jean-Baptiste Lamy
Rosy Tsopra

PT 10058039 Cardiac perforation →
Introduction

Medical terminologies
- Essential for semantic interoperability
- But difficult for Humans!

=> we developed since 10 years VCM, an iconic language for representing medical concepts
- Not as precise as terminologies, but can be used for enriching text or illustrating terms

Requires mapping between icons and terminologies
- Semantic methods for terminologies with a formal semantics (e.g. SNOMED CT [MEDINFO])
- Other terminologies requires more complex methods

Here, we will focus on MedDRA:
- Used for coding adverse effects
- Multiaxial classification without formal semantics

=> lexico-semantic method
VCM
(Visualization of Concept in Medicine)

- An iconic language for medical concepts [BMC]
  - Symptoms
  - Disorders
  - Treatments
  - Exams
  - Adverse effects
- Combinatorial grammars
  - 150 pictograms
  - 5 colors
  - 30 shapes
- => thousands of icons
- A formal semantics (based on an OWL 2.0 ontology) [KBS]
Lexical methods

Design of an OWL ontology

- MedDRA terms with codes, labels, parent-child relations
- Labels are decomposed in words and lemmas
Lexical methods

- **OWL ontology**
  - Association between lemma expressions and VCM concepts

- **Lemma exp.**
  - auricular
    - Ear pictogram

- **Lemma exp.**
  - auricular fibril
    - Heart rhythm pictogram

- **Lemma exp.**
  - coronary
    - Heart pictogram + blood vessel shape
Semantic methods

Based on multiple inheritance through the MedDRA multiaxial classification
Combining both methods

VCM concepts: Disorder, Heart

HLT 10007543 Cardiac disorders NEC

child of

PT 10058039 Cardiac perforation

mapped to

Lemma: cardiac

mapped to

VCM concept: Heart

Lemma: perfor

mapped to

VCM concept: Lesion and perforation

VCM concepts: Disorder, Heart, Procedural complication, Vascular

HLT 10007602 Cardiac and vascular procedural complications

child of
Combining both methods

PT 10058039 Cardiac perforation

VCM concepts: Disorder, Heart, Procedural complication, Lesion and perforation
User interface

- For mapping lemma expressions to VCM concepts
- Python 3
- Use OwlReady2 for accessing the OWL ontology [AIM]
from owlready2 import *

ont_term = get_ontology("http://test.org/meddra_vcm_tal.owl")
ont_align = get_ontology("http://test.org/meddra_vcm_alignement.owl")

with ont_term:

class ConceptTermino(Thing):
    def __repr__(self):
        c = MEDDRA[self.code]
        if c.meddra_type == "PT":
            return "%s (%s %s)" % (c.term, c.code, c.primary_soc_term)
        return "%s (%s)" % (c.term, c.code)

    def _get_concepts_vcm_her(self):
        vcms_her = set()
        for parent in self.parents:
            vcms_her.update(parent._get_concepts_vcm_a_heriter())
        return vcms_her

    def _get_concepts_vcm_tal(self):
        vcms_tal = set()
        expressions = [expression_mot.forme_lemmatise
            for expression_mot in self.expression_mots
            if expression_mot.forme_lemmatise.aligne_avec_concepts_vcm or expression_mot.forme_lemmatise.aligne_avec_concepts_vcm]

        expressions = expressions_prioritaires(expressions)
        for expression in expressions:
            vcms = expression.aligne_avec_concepts_vcm

            vcms_tal.update(vcms)
        return vcms_tal
Results

▶ Application of the methods on the cardiac SOC of MedDRA
  ▶ 634 MedDRA terms (excluding LLT)
    ▶ 212 lemma expressions
      ▶ 123 with 1 lemma
      ▶ 76 with 2 lemmas
      ▶ 13 with more
    ▶ => 212 lemma expressions mapped in lieu of 634 (longer) terms
  ▶ mapped to 114 different VCM icons
    ▶ 541 to a single icon
    ▶ 85 to 2 icons
    ▶ 8 to 3 icons

▶ Evaluation on 50 randomly-chosen terms
  ▶ A medical expert mapped the terms to VCM, blindly (gold standard)
    ▶ For 40 terms, the expert chose exactly the same icons
    ▶ For 6 terms, the generated icons were incomplete or more general
    ▶ For 4 terms, the generated icons were discordant
      ▶ E.g. mycoplasma infections classified as fungal infections
Four main approaches for mapping medical terminologies:

- Manual mapping
  - Long, tedious, and often not reproducible
- Chaining existent mapping:
  - MedDRA → SNOMED CT + SNOMED CT → VCM => MedDRA → VCM
  - But cumules the errors and imprecisions of each mapping
- Lexical approach
  - Difficult with icons
  - Bag of words vs expressions
- Semantic approach
  - Ontology alignment methods
  - Requires terminologies having a formal semantics
Discussion

The proposed method is easier than a manual mapping
- Lemma expressions are shorter than terms, and less numerous

Perspectives
- Extending the methods with other approaches:
  - Learning method: try to learn new lexical mapping from the already asserted ones
  - Chaining method (using SNOMED CT as an intermediate terminology): OntoADR
- Application of the methods to the entire MedDRA terminology
- Integration of VCM icons in pharmacovigilance software
- Reuse of the lemma expressions - VCM concepts mapping
  - For mapping with other terminologies, e.g. ICD10
  - For producing icons from free text
References


[AIM]: Lamy JB. Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. Artif Intell Med 2017;80:11-28