Developing a UMLS-based Ontology of Cardiology Procedures for Cognitive Support in Medical Decision Making

Jorge Biolchini, MD

- Lister Hill Center -
National Library of Medicine
2002
Goal of the research:

- **ontology** of medical procedures in the cardiology domain
- based on knowledge and information found in **UMLS**
- increase **automation** of ontology development
- theoretical background: *KR, ontology, cognition*
Why UMLS?

- **content**: knowledge corpus, knowledge sources
- **comprehensiveness**: more than 60 vocabularies and classifications
- **integration**: of sources and vocabularies
- **repository**: of concepts and relations
- **computability**: data, tables, links, applications (*browsers, NLP, etc*)
Cognition

- Principles
- paradigm shift
- present paradigm
- levels of cognitive processes
Cognition principles:

• Modularity of mental processes
  – structural
  – functional
  – evolutionary

• Parallel distributed multi-level processing
  – autonomy, parallelisms
  – distribution, multi-functionalism
  – multi-level, shifts (*level, direction*), integrations
Paradigm shift:

- setting:
  - laboratory => naturalistic medical
- object:
  - individual => social and technological settings
Expertise-novice paradigm:

• **levels** of training, expertise, competency
• **3** basic levels of competency
• **consistent** results in different professional fields
Parallel and interconnected levels of cognition

• **Factual**: identification, recognition, discrimination, etc.
• **Semantic**: meanings, causes, consequences, etc.
• **Schematic**: classification, variables, composition, etc.
• **Strategic**: plan, action order, values, options, etc.
“Knowledge-based systems emphasize meaning. Instead of processing data as a string of bits, they represent the meaning of data in terms of the real world. They carry on conversations with people in ordinary language, they find important facts before they are requested, and they solve complex problems at expert levels of performance. …

Two fields devoted to knowledge-based systems are cognitive science and artificial intelligence. Cognitive science is a merger of philosophy, linguistics, and psychology with a strong influence from computer science. Artificial intelligence (AI) is the engineering counterpart.”

John Sowa
Knowledge Representation

- **Logic** provides the formal structure and rules of inference.
- **Ontology** defines the kinds of things that exist in the application domain.
- **Computation** supports the applications that distinguish knowledge representation from pure philosophy.
Ontology

• What:
  – formal conceptualization of a particular universe of knowledge about the world

• How:
  – explicit representation of concepts, relations between them, and inference rules about them and the things they represent
Domain ontologies can be used to provide:

- **knowledge support** in aspects of underlying cognitive processes and their inter-relations
- methodology for **connecting** different **databases** through the use of common knowledge structures
- **efficient means for facilitating** professional **communication** about specific domain subjects and tasks
Medical domain ontology field

• Previous researches:
  – Knowledge:
    • elicitation from experts
  – Development:
    • from scratch
  – Foundations:
    • domain, KR
  – Purposes:
    • data collection, database connection, knowledge support, professional communication

• Present research:
  – Knowledge:
    • acquisition from corpus
  – Development:
    • to increase its automation
  – Foundations:
    • domain, KR, and cognition
  – Purposes:
    • same, with emphasis in cognitive properties of medical decision making and medical education processes
Diagram of the methodology phases:

- Concepts with definitions
- Cardiology concepts with definitions
- Mapped concepts
- Semantic groups of concepts
- Formal representations
- Ontology
- Populated categories
- Partonomy and Taxonomy
Selecting concepts with definitions:

• From domain words to domain concepts:
  – lexical-based method

• From seed concepts to domain concepts:
  – hierarchical-based method
Lexical-based method:

1. Concepts with definitions
   - MeSH JDs
   - Anatomy strings
   - Action strings
   - Exclusion strings

2. MRDEF
   - Initial concept set
   - Final concept set

3. Anatomy strings
   - MRDEF

4. MeSH JDs
   - Anatomy strings
   - Action strings
   - Exclusion strings

5. Action strings
   - MRDEF
   - Exclusion strings

6. Exclusion strings
   - Action strings
   - MRDEF

7. Final concept set
   - MRDEF

8. Cardiology concepts with definitions
   - Final concept set
Hierarchical-based method:

1. Seed concepts with definitions
2. Semantic Navigator
3. Hypernym set
4. Hyponym coverage checking

Cardiology concepts with definitions

Concepts with definitions

MRDEF

MRREL
Metathesaurus:
- 2001 edition -

- about 800,000 concepts
- about 1.9 million concept names
- concepts with definitions = 34,095
- = 4.26 % of Metathesaurus concepts
Comparison of sets:
(lexical-based and hierarchical-based)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical:</td>
<td>468</td>
<td>1.37 %</td>
<td>124</td>
<td>0.36 % (0.000155 %)</td>
</tr>
<tr>
<td>Hierarchical:</td>
<td>253</td>
<td>0.74 %</td>
<td>76</td>
<td></td>
</tr>
</tbody>
</table>
Comparison of sets

Selection of concepts

Number of concepts

Concepts: Final Concepts:

- Lexical:
- Hierarchical:
Comparison of sub-sets of lexical-based and hierarchical-based methods:
Diagram of the methodology phases:

1. Concepts with definitions
2. Cardiology concepts with definitions
3. Mapped concepts
4. Semantic groups of concepts
5. Partonomy and Taxonomy
6. Populated categories
7. Ontology
8. Formal representations
From domain concepts to domain semantics:

- Final set
- MetaMap hits set
- POS-NEG set
- Mapped concepts set
- Grouping by Semantic Types
**METAMAP HITS:**

<table>
<thead>
<tr>
<th>Total:</th>
<th>POS:</th>
<th>NEG:</th>
</tr>
</thead>
<tbody>
<tr>
<td>2152</td>
<td>1835</td>
<td>317</td>
</tr>
</tbody>
</table>

85.27%  14.73%
## MAPPED CONCEPTS:

<table>
<thead>
<tr>
<th></th>
<th>POS:</th>
<th>NEG:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>987</td>
<td></td>
</tr>
<tr>
<td></td>
<td>804</td>
<td>183</td>
</tr>
<tr>
<td></td>
<td>81.46 %</td>
<td>18.54 %</td>
</tr>
</tbody>
</table>
MAPPED CONCEPTS:
### Number of Concepts by Semantic Type:

<table>
<thead>
<tr>
<th>Semantic Type</th>
<th>Concepts</th>
<th>Subtype</th>
<th>Concepts</th>
<th>Subtype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional Concept</td>
<td>98</td>
<td>Injury or Poisoning</td>
<td>1</td>
<td>Education Activity</td>
</tr>
<tr>
<td>Therapeutic or Preventive Procedure</td>
<td>91</td>
<td>Sign or Symptom</td>
<td>9</td>
<td>Body Space or Junction</td>
</tr>
<tr>
<td>Qualitative Concept</td>
<td>62</td>
<td>Research Activity</td>
<td>8</td>
<td>Cell</td>
</tr>
<tr>
<td>Spatial Concept</td>
<td>47</td>
<td>Substance</td>
<td>8</td>
<td>Cell Function</td>
</tr>
<tr>
<td>Body Part, Organ, or Organ Component</td>
<td>46</td>
<td>Body Substance</td>
<td>7</td>
<td>Human</td>
</tr>
<tr>
<td>Diagnostic Procedure</td>
<td>44</td>
<td>Pharmacologic Substance</td>
<td>7</td>
<td>Population Group</td>
</tr>
<tr>
<td>Finding</td>
<td>44</td>
<td>Mental Process</td>
<td>7</td>
<td>Neoplastic Process</td>
</tr>
<tr>
<td>Medical Device</td>
<td>37</td>
<td>Laboratory Procedure</td>
<td>6</td>
<td>Classification</td>
</tr>
<tr>
<td>Temporal Concept</td>
<td>34</td>
<td>Congenital Abnormality</td>
<td>5</td>
<td>Group</td>
</tr>
<tr>
<td>Quantitative Concept</td>
<td>31</td>
<td>Acquired Abnormality</td>
<td>5</td>
<td>Environmental Effect of Humans</td>
</tr>
<tr>
<td>Pathologic Function</td>
<td>30</td>
<td>Occupational Activity</td>
<td>5</td>
<td>Group Attribute</td>
</tr>
<tr>
<td>Organ or Tissue Function</td>
<td>28</td>
<td>Phenomenon or Process</td>
<td>5</td>
<td>Conceptual Entity</td>
</tr>
<tr>
<td>Manufactured Object</td>
<td>24</td>
<td>Biomedical or Dental Material</td>
<td>5</td>
<td>Idea or Concept</td>
</tr>
<tr>
<td>Natural Phenomenon or Process</td>
<td>23</td>
<td>Inorganic Chemical</td>
<td>4</td>
<td>Biologically Active Substance</td>
</tr>
<tr>
<td>Intellectual Product</td>
<td>21</td>
<td>Chemical Viewed Structurally</td>
<td>4</td>
<td>Lipid</td>
</tr>
<tr>
<td>Disease or Syndrome</td>
<td>21</td>
<td>Idea or Concept</td>
<td>4</td>
<td>Clinical Attribute</td>
</tr>
<tr>
<td>Spatial Concept</td>
<td>21</td>
<td>Anatomical Structure</td>
<td>4</td>
<td>Patient or Disabled Group</td>
</tr>
<tr>
<td>Body Location or Region</td>
<td>18</td>
<td>Body System</td>
<td>4</td>
<td>Group</td>
</tr>
<tr>
<td>Health Care Activity</td>
<td>16</td>
<td>Indicator, Reagent, or Diagnostic Aid</td>
<td>4</td>
<td>Biologic Function</td>
</tr>
<tr>
<td>Laboratory or Test Result</td>
<td>15</td>
<td>Daily or Recreational Activity</td>
<td>4</td>
<td>Chemical</td>
</tr>
<tr>
<td>Tissue</td>
<td>14</td>
<td>Organic Chemical</td>
<td>4</td>
<td>Embryonic Structure</td>
</tr>
<tr>
<td>Organism Function</td>
<td>13</td>
<td>Anatomical Abnormality</td>
<td>3</td>
<td>Individual Behavior</td>
</tr>
<tr>
<td>Organism Attribute</td>
<td>12</td>
<td>Animal</td>
<td>3</td>
<td>Age Group</td>
</tr>
<tr>
<td>Physiologic Function</td>
<td>12</td>
<td>Element, Ion, or Isotope</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
First third:
Diagram of the methodology phases:

- Concepts with definitions
  - Cardiology concepts with definitions
    - Mapped concepts
      - Semantic groups of concepts
        - Partonomy and Taxonomy
          - Populated categories
            - Ontology
              - Formal representations

Diagram of the methodology phases:
From domain semantics to domain ontology:

1. Grouping by Semantic Types
2. Categories of Semantic Types
3. Categories from schema
4. Final categories
5. Taxonomic categorization
6. Canonical definition
7. Populated categories
From domain semantics to domain ontology:  
(approaches)

• Bottom-up:  
  – comparison of Semantic Types  
  – semantic comparison of concepts  
  – grouping of Semantic Types in categories
From domain semantics to domain ontology: 
(approaches)

- Top-down:
  - basic schema of agents and processes of medical act
  - comparison of schema with Semantic Types and concepts
  - other comparisons
Agents and Processes of a Medical Act

Health Professional

Knowledge

Procedure

Instruments

Disease

Signs and Symptoms

Anatomy

Health Receiver
Comparisons of categorization:

- Original information model
- Semantic Groups of Semantic Types
- MAOUSSC
## Categories of the Ontology:

<table>
<thead>
<tr>
<th>Nuclear:</th>
<th>Auxiliary:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action</td>
<td>Conditional Action</td>
</tr>
<tr>
<td>Anatomical Entity</td>
<td>Method</td>
</tr>
<tr>
<td>Pathology</td>
<td>Phenomenon</td>
</tr>
<tr>
<td>Instrument</td>
<td>Measure</td>
</tr>
<tr>
<td>Material or Energy</td>
<td>Result or Product</td>
</tr>
<tr>
<td>Physiology</td>
<td>Time or Circumstance</td>
</tr>
<tr>
<td>Purpose</td>
<td>Space</td>
</tr>
<tr>
<td>_</td>
<td>Receiver</td>
</tr>
</tbody>
</table>
“A cardiology procedure concept definition is a conceptual entity, which is composed by an action, which: has location in an anatomical entity, and affects a pathology, and affects the physiology, and has a purpose, and uses an instrument, and uses some material or energy, and can be preceded by a conditional action, and has a method, and uses or affects a phenomenon, and uses a measure, and has a result or product, and occurs in a certain circumstance in reference to time, and has some space references, and affects a receiver.”
From partonomy to taxonomy:

• Bottom-up approach
  – levels of specification-generalization
  – represent a compositional partonomy
  – parallelisms in structure for some sub-categories:
    • external and internal consistency
    • restrictions in relationships between sub-classes
Diagram of the methodology phases:

1. **Concepts with definitions**
2. **Cardiology concepts with definitions**
3. **Mapped concepts**
4. **Semantic groups of concepts**
5. **Formal representations**
6. **Ontology**
7. **Populated categories**
8. **Partonomy and Taxonomy**
Populating the categories:

- Human X Automatic processes
- comparison between both methods
- potential generalizability
### Human and Automatic populating methods:

<table>
<thead>
<tr>
<th></th>
<th>Author</th>
<th>POS</th>
<th>NEG</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>797</td>
<td>172</td>
<td>969</td>
</tr>
<tr>
<td>POS</td>
<td>80.75%</td>
<td>17.43%</td>
<td>98.18%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.71%</td>
<td>1.11%</td>
<td>1.82%</td>
</tr>
<tr>
<td></td>
<td>804</td>
<td>183</td>
<td>987</td>
<td></td>
</tr>
<tr>
<td>NEG</td>
<td>81.46%</td>
<td>18.54%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Number of Semantic Types for kinds of result of populating methods:

<table>
<thead>
<tr>
<th>Type of Result</th>
<th>H(+)-A(+)</th>
<th>H(+)-A(-)</th>
<th>H(-)-A(+)</th>
<th>H(-)-A(-)</th>
</tr>
</thead>
<tbody>
<tr>
<td># Semantic Types:</td>
<td>64</td>
<td>44</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
Number of equivalent and different Semantic Types for both methods:

<table>
<thead>
<tr>
<th></th>
<th>H(+)-A(+)</th>
<th>H(+)-A(-)</th>
<th>H(-)-A(+)</th>
<th>H(-)-A(-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H(+)-A(+)</td>
<td>0</td>
<td>43</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>H(+)-A(-)</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>H(-)-A(+)</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>H(-)-A(-)</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>
Diagram of the methodology phases:
Inhibition

Action of reducing the intensity of a body function or rhythm by inducing an inhibiting response to a stimulus.

Role
Concrete

Template Slots

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Cardinality</th>
<th>Other Facts</th>
</tr>
</thead>
<tbody>
<tr>
<td>concept_of_inhibition</td>
<td>Symbol</td>
<td>single</td>
<td>allowed-values=[hypothermia, induction]</td>
</tr>
<tr>
<td>concept_of_reduction</td>
<td>Symbol</td>
<td>single</td>
<td>allowed-values=[reduction - action, prevention]</td>
</tr>
<tr>
<td>has_goal_action</td>
<td>instance</td>
<td>single</td>
<td>classes=[Instrument]</td>
</tr>
<tr>
<td>has_location_action</td>
<td>instance</td>
<td>single</td>
<td>classes=[Anatomical_entity]</td>
</tr>
<tr>
<td>has_approach_action</td>
<td>instance</td>
<td>single</td>
<td>classes=[Anatomical_approach]</td>
</tr>
<tr>
<td>has_circumstance_action</td>
<td>instance</td>
<td>single</td>
<td>classes=[Time_phase]</td>
</tr>
<tr>
<td>has_act_action</td>
<td>instance</td>
<td>single</td>
<td>classes=[Conditional_action]</td>
</tr>
<tr>
<td>diagnoses</td>
<td>instance</td>
<td>single</td>
<td>classes=[Pathology]</td>
</tr>
<tr>
<td>evaluates</td>
<td>instance</td>
<td>single</td>
<td>classes=[Phenomenon]</td>
</tr>
<tr>
<td>affects</td>
<td>instance</td>
<td>single</td>
<td>classes=[Receiver, Pathology, Physiology]</td>
</tr>
<tr>
<td>produces</td>
<td>instance</td>
<td>single</td>
<td>classes=[Result_or_product]</td>
</tr>
<tr>
<td>prevents</td>
<td>instance</td>
<td>single</td>
<td>classes=[Pathology]</td>
</tr>
</tbody>
</table>

Supercategories
- Reduction
Diagram of the methodology phases:

- Concepts with definitions
- Cardiology concepts with definitions
- Mapped concepts
- Semantic groups of concepts
- Formal representations
- Ontology
- Populated categories
- Partonomy and Taxonomy
From categories to relations:

- Partonomy and Taxonomy
- Formal-rationalistic
- Formal-empirical
- Formal and Graphical representations
From categories to relations: (approaches)

• formal-rationalistic:
  – from basic categories
  – associative relationships for each pair
  – Sowa’s model of conceptual structures
  – check for its concrete specific plausibility

• formal-empirical:
  – from original textual definitions
  – semantic relations of UMLS Semantic Network as syntax
  – translate plausible textual relations into formal ones
  – check abstract level with the Semantic Types
Procedure has_act Action

[PROCEDURE] > (ACT) > [ACTION]
Procedure has_location Anatomical_entity

[PROCEDURE] > (LOC) > [ANATOMICAL-ENTITY]
Procedure has_object Pathology

[PROCEDURE] > (OBJ) > [PATHOLOGY]
Procedure has_instrument Instrument

[PROCEDURE] > (INST) > [INSTRUMENT]
Procedure has_goal Purpose

[PROCEDURE] < (AGNT) < [GOAL] > (RSLT) > [PURPOSE]
Procedure has_approach Anatomical_approach

[PROCEDURE] < (AGNT) < [APPROACH] > (LOC) > [ANATOMICAL-APPROACH]
Procedure has_material Material_or_energy

[PROCEDURE] > (MATR) > [MATERIAL-OR-ENERGY]
Procedure has_method Method

[PROCEDURE] > (METH) > [METHOD]
Procedure has_measure Measure

[PROCEDURE] > (MEAS) > [MEASURE]
Procedure has_result Result_or_product

[PROCEDURE] > (RSLT) > [RESULT-OR-PRODUCT]
Procedure has_duration Time_duration

[PROCEDURE] > (DUR) > [TIME-DURATION]
Procedure has_circumstance Time_phase

[PROCEDURE] < (STRT) < [CIRCUMSTANCE] > (PTIM) > [EVENT]
Procedure has_recipient Receiver

[PROCEDURE] > (RCPT) > [RECEIVER]
Procedure uses Instrument

[PROCEDURE] < (AGNT) < [USE] > (INST) > [INSTRUMENT]
Procedure uses Material_or_energy

[PROCEDURE] < (AGNT) < [USE] > (INST) > [MATERIAL-OR-ENERGY]
Procedure uses Method

[PROCEDURE] < (AGNT) < [USE] > (METH) >
[ACT: [SPECIFIC-METHOD] < (AGNT) < [EVENT] > (MANR) > [METHOD-MODIFIER]]
Procedure uses Measure

[PROCEDURE] < (AGNT) < [USE] > (INST) >
[MEASURE: [DIMENSION] > (QTY) > [NUMBER]]