Aligning Representations of Anatomical Knowledge

Songmao Zhang, Ph.D.

Mentor: Olivier Bodenreider, M.D., Ph.D.

Lister Hill National Center for Biomedical Communications

Bethesda, Maryland - USA
General framework

◆ Comparison and evaluation of structures and reasoning potentials in different medical knowledge representation systems

◆ Case study:
  - Foundational Model of Anatomy (FMA)
    - Frame-based structure
  - GALEN common reference model (GALEN)
    - Description logic
An overview
of our alignment project
Generating direct alignment

Concepts

1:1

Lexical similarity

Structural similarity
1:1 concept match

Lexical alignment

FMA: *Fibularis tertius* (synonym: *Peroneous tertius*)
GALEN: *Peroneus Tertius*
1:1 concept match

- Structural alignment: shared hierarchical relationships among concept

FMA

Muscle of leg

Fibularis tertius

Muscle cell

Leg

GALEN

Muscle of Leg

Peroneous Tertius

Myocyte

Leg

is-a

has-part

part-of
Generating direct alignment

Concepts

1:1

Lexical similarity

Structural similarity

n:m

Structural similarity
n:m concept match

FMA

Proximal free limb segment
Middle free limb segment

Arm
Thigh
Forearm
Leg

Left arm Right arm Left thigh Right thigh Left forearm Right forearm Left leg Right leg

GALEN

Extremity Long Part

Arm Thigh Forearm Leg
Generating direct alignment

Concepts

1:1

Lexical similarity

Structural similarity

n:m

Structural similarity

Relationships

Structural similarity
### Associative relationship mapping patterns

<table>
<thead>
<tr>
<th>FMA</th>
<th>GALEN</th>
<th>Frequency (N = 4,070)</th>
</tr>
</thead>
<tbody>
<tr>
<td>branch of</td>
<td>isBranchOf</td>
<td>310</td>
</tr>
<tr>
<td>member of</td>
<td>is-a</td>
<td>42</td>
</tr>
<tr>
<td>nerve supply</td>
<td>part-of – isServedBy</td>
<td>16</td>
</tr>
<tr>
<td>part-of – contained in</td>
<td>isNonPartitivelyContainedIn</td>
<td>10</td>
</tr>
<tr>
<td>contained in</td>
<td>boundsSpace – inverse-isa</td>
<td>2</td>
</tr>
</tbody>
</table>

- branch of: 310 occurrences, 8% frequency
- member of: 42 occurrences, 1% frequency
- nerve supply: 16 occurrences, 0.4% frequency
- part-of – contained in: 10 occurrences, 0.25% frequency
- contained in: 2 occurrences, 0.05% frequency
Generating direct alignment

FMA and GALEN

PMork, etc., Schema matching algorithm for comparing FMA and GALEN

Manual review

MA and NCI

Lexical similarity

Structural similarity

Structural similarity

Structural similarity

Deriving indirect alignment

MA and NCI through FMA as reference

1:1

n:m

Concepts

Relationships
Aligning multiple ontologies of anatomy

Deriving indirect mappings from direct mappings to a reference
Approaches to aligning multiple ontologies

Pairwise alignment

Alignment through a reference
Introduction

- Objective: to investigate the indirect alignment of two anatomical ontologies through a reference ontology
Introduction

◆ Three ontologies of anatomy:
  ● Adult Mouse Anatomical Dictionary (MA)
  ● Anatomy subset of NCI Thesaurus (NCI)
  ● Foundational Model of Anatomy (FMA)

◆ First attempt to automatically derive mappings among ontologies from their alignments to a reference ontology
Materials
Adult Mouse Anatomical Dictionary

◆ Structured controlled vocabulary
◆ 2,404 concepts each identified by one name
  ● Head/neck, Adrenal artery
◆ 259 synonyms
  ● Limb has a synonym Extremity
◆ Directed acyclic graph
◆ Two relationships: is-a and part-of
◆ 38% concepts have no is-a relationship
  ● Knee part-of Hindlimb
◆ 4% concepts have more than one is-a relationship
  ● Hand phalanx is-a Phalanx
  is-a Hand digit bone
NCI Thesaurus

- Standard vocabularies for cancer research
- Anatomy class
- Available in Ontology Web Language (OWL)
- 4,410 concepts each identified by preferred name
  - Abdominal esophagus
- 2,371 synonyms
  - Orbit has a synonym Eye socket
- Every concept has at least one is-a relationship
- 4% concepts have more than one is-a relationship
  - Radius bone is-a Long bone
    - is-a Bone of the upper extremity
- Concepts are connected by a part-of relationship
  - Anatomic structure is physical part of
Foundational Model of Anatomy

◆ Conceptualize the physical objects and spaces that constitute the human body
◆ Frame-based structure in Protégé
◆ 71,202 concepts each identified by preferred name
  - Uterine tube
◆ 52,713 synonyms
  - Uterine tube has a synonym Oviduct
◆ Every concept has one and only one is-a relationship
◆ Seven part-of relationships and their inverses
  - constitutional part of and constitutional part
  - regional part of and regional part
## MA, NCI and FMA

<table>
<thead>
<tr>
<th></th>
<th>MA</th>
<th>NCI</th>
<th>FMA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Underlying data model</strong></td>
<td>Directed acyclic graph</td>
<td>Available in OWL</td>
<td>Frame-based structure in Protégé</td>
</tr>
<tr>
<td><strong>Domain coverage</strong></td>
<td>Mouse anatomy</td>
<td>Human anatomy related to cancers</td>
<td>Human anatomy</td>
</tr>
<tr>
<td><strong>Concepts</strong></td>
<td>2,404</td>
<td>4,410</td>
<td>71,202</td>
</tr>
<tr>
<td><strong>Synonyms</strong></td>
<td>259</td>
<td>2,371</td>
<td>52,713</td>
</tr>
<tr>
<td><strong>Hierarchical relationships</strong></td>
<td><em>is-a, part-of</em></td>
<td><em>is-a, part-of</em></td>
<td><em>is-a, part-of</em> (7)</td>
</tr>
<tr>
<td><strong>Inverses</strong></td>
<td>-</td>
<td>-</td>
<td><em>inverse-isa, has-part</em> (7)</td>
</tr>
</tbody>
</table>
Methods
Three phases

MA

NCI

FMA

Direct alignment MA-FMA

Direct alignment NCI-FMA

Alignment through the FMA

Direct alignment MA-NCI
Generating direct alignment

Concepts

1:1

Lexical similarity

Structural similarity
Phase 1 Direct alignment

Lexical alignment

- Acquiring terms
- Identifying matches (i.e., shared concepts) lexically

Structural alignment

- Acquiring (explicit and implicit) semantic relations
- Identifying matches structurally
Direct alignment Lexical alignment

◆ Comparing two ontologies at the term level
  - Exact match
  - Match after normalization

◆ Preferred names and synonyms are used
  - MA: *Forelimb*
    - NCI: *Upper extremity* (synonym: *Forelimb*)

◆ UMLS synonymy is used to identify additional matches
  - MA: *Profunda femoris artery*
    - NCI: *Deep femoral artery*
Direct alignment Structural alignment

- Acquiring inter-concept hierarchical relationships
  - *is-a*, *part-of*, and their inverses *inverse-is-a* and *has-part*
- Complementing missing inverse relations
- Generating new inter-concept relationships by applying inference rules

Diagram:

- Pyramidal lobe
- Thyroid gland lobe
- Neck

Relationships:

- Pyramidal lobe → Thyroid gland lobe (is-a)
- Thyroid gland lobe → Neck (part-of)
- Pyramidal lobe → Neck (part-of)
Direct alignment Structural alignment

- Reification of *part-of* relationships in the FMA

\[<X, \text{is-a}, \text{Part of } Y>\]

\[<X, \text{part-of, } Y>\]

- Making explicit reified *part-of* relations in the FMA
  - Reified: \(<\text{Heel, is-a, Subdivision of foot}>\)
  - Augmented: \(<\text{Heel, part-of, Foot}>\)
Direct alignment Structural alignment

Positive evidence: structural similarity

MA:

\[ \text{Forelimb} \rightarrow \text{Upper extremity} \]

\[ \text{Upper extremity} \rightarrow \text{Limb} \]

\[ \text{Limb} \rightarrow \text{Hand} \]

\[ \text{Hand} \rightarrow \text{Arm} \]

NCI:

\[ \text{Upper extremity} \rightarrow \text{Limb} \]

\[ \text{Limb} \rightarrow \text{Hand} \]

\[ \text{Hand} \rightarrow \text{Arm} \]
Direct alignment Structural alignment

- Negative evidence: structural conflict

MA:

```
Pericardial cavity
```

FMA:

```
Pericardial cavity
```

```
Pericardial sac
```

```
Pericardium
```

```
part-of
```

```
part-of
```

```
part-of
```
Phase 2 Indirect alignment

- Upper limb
- Forelimb
- Upper extremity

FMA

- Forelimb

MA: direct

NCI: indirect

FMA: direct

NCI: direct
Phase 3 Comparison of two alignments

Direct alignment

Indirect alignment

Shared matches

Specific matches

Specific matches
Results
# Semantic relations acquired

<table>
<thead>
<tr>
<th>Types of hierarchical relations</th>
<th>MA</th>
<th>NCI</th>
<th>FMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicitly represented</td>
<td>2,926</td>
<td>7,250</td>
<td>401,045</td>
</tr>
<tr>
<td>Complemented</td>
<td>2,926</td>
<td>7,250</td>
<td>4,977</td>
</tr>
<tr>
<td>Augmented</td>
<td>-</td>
<td>-</td>
<td>158,282</td>
</tr>
<tr>
<td>Inferred</td>
<td>15,044</td>
<td>45,302</td>
<td>5,553,488</td>
</tr>
<tr>
<td>Total</td>
<td>20,896</td>
<td>59,802</td>
<td>6,117,792</td>
</tr>
</tbody>
</table>
## Three direct alignments

<table>
<thead>
<tr>
<th></th>
<th>MA - NCI</th>
<th>MA - FMA</th>
<th>NCI - FMA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>715 matches</td>
<td>1,353 matches</td>
<td>2,173 matches</td>
</tr>
<tr>
<td>No evidence</td>
<td>62 (8.7%)</td>
<td>66 (4.9%)</td>
<td>205 (9.4%)</td>
</tr>
<tr>
<td>Positive evidence</td>
<td>653 (91.3%)</td>
<td>1,283 (94.8%)</td>
<td>1,958 (90.1%)</td>
</tr>
<tr>
<td>Negative evidence</td>
<td>0</td>
<td>4 (0.3%)</td>
<td>10 (0.5%)</td>
</tr>
</tbody>
</table>
Indirect alignment

- 1,353 matches in direct alignment MA-FMA
- 2,173 matches in direct alignment NCI-FMA
- 703 matches in alignment through the FMA
- 715 matches in direct alignment MA-NCI
Comparison of two alignments

Direct alignment

Indirect alignment

61 49

715 703

654
Comparison of two alignments

654 shared matches

- 583 (89%) supported by structural evidence in three direct alignments
  - {MA: Forelimb, NCI: Upper extremity}
- 65 (10%) received no evidence in at least one of the three direct alignments
  - Chondrocranium in MA, NCI and FMA
- 6 (1%) received negative evidence in one of the three direct alignments
  - Pericardial cavity received negative evidence in MA-FMA
  - Pericardial cavity received no evidence in MA-NCI
  - Pericardial cavity received positive evidence in NCI-FMA
Discussion
Benefits of the indirect alignment

- 49 matches are specific to the indirect alignment between MA and NCI through the FMA
  - 7% of matches in the indirect alignment
  - 45 received positive evidence in both direct alignments MA-FMA and NCI-FMA
  - 2 received no evidence in one of the direct alignments MA-FMA and NCI-FMA
  - 2 received negative evidence in one of the direct alignments MA-FMA and NCI-FMA
Benefits of the indirect alignment

Why are the 49 matches not identified in the direct alignment?

- Additional synonyms by the FMA
  - MA: Integumental system
  - NCI: Integumentary system

- Additional relations by the FMA
  - MA: Hip bone
  - NCI: Pelvic bone
  - NCI: Ischium (syn: Hip)
  - FMA: Hip Bone
  - part-of
  - FMA: Hip
  - isa
Benefits of the direct alignment

◆ 61 matches are specific to the direct alignment
  MA-NCI
  ● Account for 9% of matches in the direct alignment
  ● 53 received positive evidence
  ● 8 received no evidence

◆ 44 matches identified both directly and indirectly, but received no structural evidence in the indirect alignment
  ● 14 received positive evidence in the direct alignment
Benefits of the direct alignment

Why are the 61 matches not identified in the indirect alignment through the FMA?

- Different coverage

**MA:**
- Common iliac artery
- Internal iliac artery
- External iliac artery
- Iliac artery

**FMA:**
- Common iliac artery
- Internal iliac artery
- External iliac artery

**NCI:**
- Common iliac artery
- Internal iliac artery
- External iliac artery
- Iliac artery
Benefits of the direct alignment

- Why are the 14 matches supported in the direct alignment while having no evidence in the indirect alignment through the FMA?
  - Different representation

MA:

Blood vessel

FMA:

Blood vessel

NCI:

Blood vessel

Artery

Vein

Artery

General anatomical term

Vein
Alignment through a reference vs. pairwise alignment

◆ Efficiency of alignment through a reference
  - $n(n-1)/2$ pairwise mappings
  - $(n-1)$ mappings to a reference

◆ Feasibility of alignment through a reference
  - Identified 91% of matches in the direct alignment
  - Identified additional matches not discovered by the direct alignment
Alignment through a reference vs. pairwise alignment

- **Requirements for a reference ontology**
  - Broad coverage
  - Standard representation principles

- **The FMA serves as a reference ontology**
  - Large number of concepts
  - Large number of synonyms
  - Comprehensive relationships among concepts
  - Modeled by a set of declared foundational principles
  - Independent of biomedical applications
Current limitations and future work

◆ 764 matches identified together by direct and indirect alignments between MA and NCI
  ● 32% of MA concepts
  ● 17% of NCI anatomical concepts
◆ Relies heavily on lexical similarity
◆ Only identifies one-to-one concept matches
◆ Fully automatic techniques without validation of the alignments
  ● No manual validation
  ● But cross-validation
Publications


Publications


◆ Songmao Zhang, Olivier Bodenreider, Comparing associative relationships among equivalent concepts across ontologies. Medinfo 2004:459-463

◆ Songmao Zhang, Olivier Bodenreider, Law and Order: Assessing and enforcing compliance with ontological modeling principles. Computers in Biology and Medicine; 2005:(accepted)
Publications

- **Songmao Zhang, Peter Mork, Olivier Bodenreider, Philip A. Bernstein**, Comparing two approaches for aligning representations of anatomy. *Artificial Intelligence in Medicine; 2005:*(submitted)

- **Songmao Zhang, Olivier Bodenreider**, Alignment of multiple ontologies of anatomy: Deriving indirect mappings from direct mappings to a reference. *AMIA; 2005:*(submitted)

- **Olivier Bodenreider, Terry Hayamizu, Martin Ringwald, Sherri De Coronado, Songmao Zhang**, Of mice and men: aligning mouse and human anatomies. *AMIA; 2005:*(submitted)

- **Songmao Zhang, Olivier Bodenreider**, Frame-based structures vs. description logics in representation and reasoning of the Foundational Model of Anatomy. *2005:*(in preparation)

- **Songmao Zhang, Olivier Bodenreider**, Structural alignment of anatomical ontologies. *2005:*(in preparation)
THANK YOU