Aligning relationships in the UMLS
Methods and preliminary results

Olivier Bodenreider
Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA
Acknowledgments

◆ Alexa T. McCray
   Harvard Medical School
   Boston, Massachusetts

◆ Lowell Vizenor
   OntologyWorks, Inc
   Baltimore, Maryland
Overview

◆ The UMLS: A two-level structure
◆ Aligning relationships
  ● Multiple approaches
  ● Preliminary results
◆ Towards an ontology of relationships
The UMLS

A two-level structure
Unified Medical Language System

- **SPECIALIST Lexicon**
  - 360,000 lexical items
  - Part of speech and variant information

- **Metathesaurus**
  - 6M names from over 100 terminologies
  - 1.5M concepts
  - 8M relations

- **Semantic Network**
  - 135 high-level categories
  - 7000 relations among them
Heart

Concepts

Metathesaurus

Semantic Types

Body Part, Organ or Organ Component

Anatomical Structure

Embryonic Structure

Pharmacologic Substance

Disease or Syndrome

Population Group

Semantic Network

Heart

Valves

Heart

Fetal

Tissue Donors

Cardiotoxic Agents

Angina Pectoris

Mediastinum

Saccular Viscus

Esophagus

Left Phrenic Nerve

38

16

5

13

22

49

237

143

Semantic Types

Anatomical Structure

Embryonic Structure

Pharmacologic Substance

Disease or Syndrome

Population Group

Semantic Network
UMLS Metathesaurus
143 source vocabularies
  - 17 languages

Broad coverage of biomedicine
  - 5.9M names
  - 1.4M concepts
  - 8M relations

Common presentation
Metathesaurus Basic organization

◆ Concepts
  ● Synonymous terms are clustered into a concept
  ● Properties are attached to concepts, e.g.,
    ■ Unique identifier
    ■ Definition

◆ Relations
  ● Concepts are related to other concepts
  ● Properties are attached to relations, e.g.,
    ■ Type of relationship
    ■ Source
Integrating subdomains

- Clinical repositories
- Genetic knowledge bases
- Biomedical literature
- SNP databases
- Other subdomains
- SNOMED CT
- OMIM
- MeSH
- Genome annotations
- GO
- FMA
- Anatomy
- NCBI Taxonomy
- Model organisms
- ...
Integrating subdomains

- Clinical repositories
- Genetic knowledge bases
- Biomedical literature
- Genome annotations
- Anatomy
- Model organisms
- Other subdomains

Integrating subdomains in biomedical literature, model organisms, genome annotations, clinical repositories, and other subdomains.
Trans-namespase integration

Addison's disease (363732003)

Other subdomains

Model organisms

NCBI Taxonomy

FMA

Anatomy

GO

Genome annotations

UMLS C0001403

Biomedical literature

Addison Disease (D000224)

Genetic knowledge bases

OMIM

Snomed CT

MESH
Addison’s Disease: Concept

Addison’s Disease

Disease or Syndrome

ADRENAL INSUFFICIENCY (ADDISON’S DISEASE)
ADRENOCORTICAL INSUFFICIENCY, PRIMARY FAILURE
Hypoadrenalisms, Primary
Melasma addisonii
Primary adrenal deficiency
Asthenia pigmentosa
Bronzed disease
Insufficiency, adrenal primary
Primary adrenocortical insufficiency
Addison’s, disease

Maladie d’Addison - French
Addison-Krankheit - German
Morbo di Addison - Italian
Doença de Addison - Portuguese
АДДИСОНОВА БОЛЕЗНЬ - Russian
アジソン病 - Japanese

An adrenal disease characterized by the progressive destruction of the adrenal cortex, resulting in insufficient production of aldosterone and hydrocortisone. Clinical symptoms include anorexia; nausea; weight loss; muscle weakness; and hyperpigmentation of the skin due to increase in circulating levels of ACTH precursor hormone which stimulates melanocytes.
Metathesaurus  Evolution over time

- Concepts never die (in principle)
  - CUIs are permanent identifiers
- What happens when they do die (in reality)?
  - Concepts can merge or split
  - Resulting in new concepts and deletions

Addison's disease
C0001403

Addison's disease, NOS
C0271735

Metathesaurus Relationships

- **Symbolic relations:** ~8 M pairs of concepts
- **Statistical relations:** ~6 M pairs of concepts (co-occurring concepts)
- **Mapping relations:** ~150,000

Categorization: Relationships between concepts and semantic types from the Semantic Network
Symbolic relations

- Relation
  - Pair of “atom” identifiers
  - Type
  - Attribute (if any)
  - List of sources (for type and attribute)

- Semantics of the relationship:
  defined by its type [and attribute]

- Recorded bidirectionally
Symbolic relationships

**Type**

- **Hierarchical**
  - Parent / Child
  - Broader / Narrower than
- **Derived from hierarchies**
  - Siblings (children of parents)
- **Associative**
  - Other
- **Various flavors of near-synonymy**
  - Similar
  - Source asserted synonymy
  - Possible synonymy
Symbolic relationships

- Hierarchical
  - isa (is-a-kind-of)
  - part-of
- Associative
  - location-of
  - caused-by
  - treats
  - ...
- Cross-references (mapping)
Categorize concepts

- High-level categories (semantic types)
- Assigned by the Metathesaurus editors
- Independently of the hierarchies in which these concepts are located

Disease or Syndrome

Diseases

Endocrine Diseases

Adrenal Gland Diseases

Adrenal Gland Hypofunction

Addison’s Disease
Categorization vs. hierarchies

- Semantic type
  - List of all concepts having this semantic type
- Concept
  - List of all descendants
UMLS Semantic Network
Semantic Network

- Semantic types (135)
  - tree structure
  - 2 major hierarchies
    - Entity
      - Physical Object
      - Conceptual Entity
    - Event
      - Activity
      - Phenomenon or Process
Semantic Network

- Semantic network relationships (54)
  - hierarchical (isa = is a kind of)
    - among types
      - Animal isa Organism
      - Enzyme isa Biologically Active Substance
    - among relations
      - treats isa affects
  - non-hierarchical
    - Sign or Symptom diagnoses Pathologic Function
    - Pharmacologic Substance treats Pathologic Function
“Biologic Function” hierarchy (isa)

- Biologic Function
  - Physiologic Function
    - Organism Function
      - Mental Process
    - Organ or Tissue Function
    - Cell Function
    - Molecular Function
      - Genetic Function
  - Pathologic Function
    - Cell or Molecular Dysfunction
    - Disease or Syndrome
      - Mental or Behavioral Dysfunction
    - Experimental Model of Disease
      - Neoplastic Process
Associative (non-isa) relations

- Organism
  - Anatomical Structure
    - Embryonic Structure
    - Congenital Abnormality
    - Acquired Abnormality
  - Anatomical Attribute
    - Laboratory or Test Result
    - Sign or Symptom
  - Fully Formed Anatomical Structure
    - Body System
      - Body Part, Organ or Organ Component
    - Tissue
    - Cell
      - Cell Component
      - Gene or Genome
  - Biologic Function
    - Physiologic Function
    - Pathologic Function
  - Finding
    - Injury or Poisoning
    - Body Location or Region
      - Body Space or Junction
    - Location of
      - Location of
      - Conceptual part of
      - Conceptual part of
      - Adjacent to
    - Location of
      - Location of
      - Conceptual part of
      - Conceptual part of
      - Location of
    - Conceptual part of
      - Conceptual part of
      - Conceptual part of
    - Contains, produces
      - Disrupts
      - Co-occurs with
    - Process of
      - Evaluation of
        - Evaluation of
          - Evaluation of
    - Part of
      - Part of
      - Part of
      - Part of
      - Part of
    - Property of
      - Property of
      - Property of
    - Part of
      - Part of
      - Part of
      - Part of
      - Part of
    - Conceptual part of
      - Conceptual part of
      - Conceptual part of
    - Adjacent to
      - Adjacent to
      - Adjacent to
    - Location of
      - Location of
      - Location of
    - Conceptual part of
      - Conceptual part of
      - Conceptual part of
    - Contains, produces
      - Contains, produces
      - Contains, produces
    - Process of
      - Process of
      - Process of
    - Evaluation of
      - Evaluation of
      - Evaluation of
    - Part of
      - Part of
      - Part of
      - Part of
      - Part of
Relationship hierarchy

- associated with
  - functionally related to
  - temporally related to
  - physically related to
  - spatially related to
  - conceptually related to
Why a semantic network?

- Semantic Types serve as high level categories assigned to Metathesaurus concepts, independently of their position in a hierarchy.

- A relationship between 2 Semantic Types (ST) is a possible link between 2 concepts that have been assigned to those STs.
  - The relationship may or may not hold at the concept level.
  - Other relationships may apply at the concept level.
Relationships can inherit semantics

Semantic Network

- Fully Formed Anatomical Structure
- Biologic Function
- Pathologic Function
- Disease or Syndrome
- Body Part, Organ, or Organ Component
- Adrenal Cortex
- Adrenal Cortical hypofunction
- Metathesaurus

location of

isa
Aligning relationships
Relationships in the UMLS

Metathesaurus
- 139 relationships
  - Thesaural
  - Specified
- No definition
- No organization

Semantic Network
- 54 relationships
- Textual definition
- Organized in 5 hierarchies

- isa
- causes
- due to
- manifestation of
- icd asterisk
- associated finding of
- may treat

- isa
- associated with
  - functionally related to
  - physically related to
  - spatially related to
  - temporally related to
  - conceptually related to
Concepts/Types vs. Relationships

UMLS Semantic Network

Semantic Type a

Concept categorization

Semantic Type b

Semantic Network relationships

Semantic Type c

Concept categorization

Concept 1

Metathesaurus relationship

Concept 2

UMLS Metathesaurus

Lister Hill National Center for Biomedical Communications
Objectives

◆ Elicit the meaning of Metathesaurus relationships

◆ Establish
  ● A correspondence between
    ◆ Each of the 139 relationships in the Metathesaurus
    ◆ And relationships in the Semantic Network
  ● The nature of the correspondence
    ◆ Equivalence, specialization, other

◆ Enrich SN relationships with relationships from the Metathesaurus

◆ Develop an ontology of relationships
Multiple complementary approaches

- **Metathesaurus-centric**
  - Manual elicitation
  - Abstraction at the level of high-level concepts
  - Abstraction at the level of Semantic Types

- **Semantic Network-centric**
Manual elicitation

- Create random samples of a maximum of 50 relations per Metathesaurus relationship
- Review manually
<table>
<thead>
<tr>
<th>Domain</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>causative agent of</strong></td>
<td><strong>causative agent of</strong></td>
</tr>
<tr>
<td>C001544</td>
<td>A0302173</td>
</tr>
<tr>
<td>C002677</td>
<td>A0009901</td>
</tr>
<tr>
<td>C0132649</td>
<td>A0322158</td>
</tr>
<tr>
<td>C011582</td>
<td>A0307611</td>
</tr>
<tr>
<td>C001327</td>
<td>A03627347</td>
</tr>
<tr>
<td>C0007537</td>
<td>A29873568</td>
</tr>
<tr>
<td>C0036874</td>
<td>A2888034</td>
</tr>
<tr>
<td>C0058511</td>
<td>A2185117</td>
</tr>
<tr>
<td>C0056904</td>
<td>A2882629</td>
</tr>
<tr>
<td>C001327</td>
<td>A03627347</td>
</tr>
<tr>
<td>C0003954</td>
<td>A28877625</td>
</tr>
<tr>
<td>C0040515</td>
<td>A3093971</td>
</tr>
<tr>
<td>C0022237</td>
<td>A11807998</td>
</tr>
<tr>
<td>C0434078</td>
<td>A23200991</td>
</tr>
<tr>
<td>C0042016</td>
<td>A2887930</td>
</tr>
<tr>
<td>C0004611</td>
<td>A2917407</td>
</tr>
<tr>
<td>C001327</td>
<td>A69390913</td>
</tr>
<tr>
<td>C0040840</td>
<td>A2887564</td>
</tr>
<tr>
<td>C0314732</td>
<td>A3043411</td>
</tr>
<tr>
<td>C002677</td>
<td>A29828747</td>
</tr>
<tr>
<td>C0360112</td>
<td>A30659640</td>
</tr>
<tr>
<td>C0314732</td>
<td>A3043411</td>
</tr>
<tr>
<td>C0314732</td>
<td>A3043411</td>
</tr>
<tr>
<td>C0014939</td>
<td>A3989775</td>
</tr>
<tr>
<td>C0036875</td>
<td>A28880594</td>
</tr>
<tr>
<td>C0314732</td>
<td>A3043411</td>
</tr>
<tr>
<td>C0448592</td>
<td>A30785694</td>
</tr>
<tr>
<td>C0322643</td>
<td>A33001128</td>
</tr>
<tr>
<td>C0013682</td>
<td>A0301279</td>
</tr>
<tr>
<td>C0036875</td>
<td>A28880592</td>
</tr>
<tr>
<td>C0314732</td>
<td>A3043411</td>
</tr>
<tr>
<td>C0040840</td>
<td>A2887564</td>
</tr>
<tr>
<td>C0314732</td>
<td>A3043411</td>
</tr>
<tr>
<td>C0006167</td>
<td>A3376034</td>
</tr>
<tr>
<td>C001327</td>
<td>A03627347</td>
</tr>
<tr>
<td>C0360108</td>
<td>A03076366</td>
</tr>
<tr>
<td>C0027573</td>
<td>A1162895</td>
</tr>
<tr>
<td>C0445079</td>
<td>A3228891</td>
</tr>
<tr>
<td>C001327</td>
<td>A03627347</td>
</tr>
<tr>
<td>C0045511</td>
<td>A6917107</td>
</tr>
<tr>
<td>C0051208</td>
<td>A28828726</td>
</tr>
<tr>
<td>C0051208</td>
<td>A30828729</td>
</tr>
<tr>
<td>C0043822</td>
<td>A7355448</td>
</tr>
<tr>
<td>C0314754</td>
<td>A30207222</td>
</tr>
<tr>
<td>C0006459</td>
<td>A2887915</td>
</tr>
<tr>
<td>C001582</td>
<td>A3034411</td>
</tr>
<tr>
<td>C0019704</td>
<td>A0301298</td>
</tr>
<tr>
<td>C0052322</td>
<td>A29930980</td>
</tr>
<tr>
<td>C0033609</td>
<td>A2885980</td>
</tr>
</tbody>
</table>
Virus

Anti-psychotic agent

Autonomic agent

Ingestible alcohol

Chronic aggressive viral hepatitis

Sulpiride poisoning of undetermined intent

Lidoflazine allergy

Mental and behavioral disorders

**Infectious agents and Pharmacologic substances**

"Brings about a condition or an effect. Implied here is that an agent, such as for example, a pharmacologic substance or an organism, has brought about the effect"

**causative agent of** _SNOMED CT_ ≡ **causes** _SN_
1b Abstraction at the level of high-level concepts

- Compute the lowest common ancestor(s) for concepts in the domain and range of the relationship, respectively.

Diagram:
- Endoscope
  - access instrument of
- Procedure by method
  - access instrument of (Endoscope, Procedure)
Abstraction at the level of STs

- Compute the distribution of the STs for concepts in the domain and range of the relationship, respectively.

Diagram:
- Medical Device
- Therap. or Prev. Procedure
- Diagnostic Procedure
- Health Care Activity

Access instrument of

Access instrument of (Medical Device, Diagnostic/Therap. or Prev. Procedure)
Identifying candidates in the SN

Medical Device

Therap. or Prev. Procedure

Diagnostic Procedure

"Employs in the carrying out of some activity. This includes applies, utilizes, employs, and avails."

Medical Device

Therap. or Prev. Procedure

Diagnostic Procedure

access instrument of

access instrument of $_{SNOMED\ CT}$ $\equiv \text{inv}(\text{uses}_{SN})$
For a given relationship between two STs in the SN, list all the relationships between concepts categorized by these STs.

\[
\{\text{ingredient of}_*, \text{active ingredient of}_{\text{SNOMEDCT}}, \text{dose form of}_*, \} \equiv \text{ingredient of}_\text{SN}
\]
Results

139 specified relationships (RELA) in the Metathesaurus

Major contributors

- SNOMED CT (62)
- LOINC (15)
- NDF-RT (15)

Most relationships are specific to one vocabulary

- 116 specific to one vocabulary
- 23 found in at least 2 vocabularies
Results  Alignment

◆ 80 relationships (58%) could be aligned
  ● Identical
    - affects, process of, ingredient of
  ● Roughly equivalent to
    - focus of_{SNOMEDCT} \equiv issue in_{SN}
  ● More specific than
    - metabolic site of_{NDFRT} < functionally related to_{SN}
◆ Non aligned (non-semantic relationships)
  ● Lexical relationships (e.g., british form of*, suffix of_{LOINC})
  ● Mapping relationships (e.g., see from_{CRISP}, mapped to*)
  ● Vocabulary management (e.g., replaces_{SNOMEDCT})
Some issues

◆ Manual elicitation
  ● Manual review by experts required

◆ Abstraction at the level of high-level concepts
  ● Lowest common ancestor = root of the terminology
    (heterogeneous set of concepts)

◆ Abstraction at the level of Semantic Types
  ● Multiple STs for a set of concepts
  ● Multiple relationships for a given pair of STs

◆ Semantic Network-centric approach
  ● Multiple relationships in the Metathesaurus between the concepts corresponding to a given pair of STs
Towards an ontology of relationships
Motivation

◆ Consistency checking
  • Relations asserted between STs in the SN should provide constraints for relations between Metathesaurus concepts categorized by these STs

◆ Effective subsumption reasoning
  • Requires explicit equivalence between identical relationships (e.g., \(\text{inv}(\text{due to}) = \text{cause of} \equiv \text{causes}\))
  • Requires explicit subProperty relations
Existing resources

◆ Semantic Network relationships
  ● 54 relationships, not formally defined, organized into a shallow hierarchy
  ● http://semanticnetwork.nlm.nih.gov/

◆ GALEN relationships
  ● Over 500 relationships, semi-formally defined, organized into a hierarchy
  ● http://www.opengalen.org/

◆ OBO relation ontology
  ● 10 relationships, formally defined
  ● http://obofoundry.org/ro/
GALEN relations

- isIngredientOf

Diagram:

- topLinkClass
- Attribute
  - DomainAttribute
    - ConstructiveAttribute
      - InversePartitiveAttribute
        - IsDivisionOf
        - makesUp
          - specificallyMakesUp
          - isIngredientOf
        - isSpecificIngredientOf
        - isMultipleIngredientOf
        - isNotIngredientOf
        - isNotChemicalSubgroupOf
      - inverseProcessPartitiveAttribute
    - InverseStructuralAttribute
      - involves
      - FunctionalAttribute
        - hasReference
        - hasFrameOfReference
        - hasPhysicalMeans
        - hasPersonPerforming
      - records
        - hasEmployer
      - WrapperAttribute
      - TemporalAttribute
      - ModifierAttribute
      - SpecificationLevelAttribute
    - ApplicationAttribute

Table:

<table>
<thead>
<tr>
<th>Name</th>
<th>isIngredientOf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inverse</td>
<td>hasIngredient</td>
</tr>
<tr>
<td>Cardinality</td>
<td>manyMany</td>
</tr>
<tr>
<td>Properties</td>
<td>inverse, transitive</td>
</tr>
<tr>
<td>Transits</td>
<td>makesUp</td>
</tr>
<tr>
<td>Transitive orbits</td>
<td>makesUp</td>
</tr>
</tbody>
</table>

Attribute:  | Pretty: on | Filter: on |
### OBO relations

<table>
<thead>
<tr>
<th>name</th>
<th>transitive</th>
<th>symmetric</th>
<th>reflexive</th>
<th>anti-symmetric</th>
<th>documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>is_a</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td>View detailed summary</td>
</tr>
<tr>
<td>part_of</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td>View detailed summary</td>
</tr>
<tr>
<td>integral_part_of</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td>View detailed summary</td>
</tr>
<tr>
<td>proper_part_of</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>View detailed summary</td>
</tr>
<tr>
<td>located_in</td>
<td>+</td>
<td></td>
<td></td>
<td>+</td>
<td>View detailed summary</td>
</tr>
<tr>
<td>contained_in</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>View detailed summary</td>
</tr>
<tr>
<td>adjacent_to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>View detailed summary</td>
</tr>
<tr>
<td>transformation_of</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>View detailed summary</td>
</tr>
<tr>
<td>derives_from</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>View detailed summary</td>
</tr>
<tr>
<td>preceded_by</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>View detailed summary</td>
</tr>
<tr>
<td>has_participant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>View detailed summary</td>
</tr>
<tr>
<td>has_agent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>View detailed summary</td>
</tr>
<tr>
<td>instance_of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>View detailed summary</td>
</tr>
</tbody>
</table>
Extending the SN relationship hierarchy

- functionally related to
- temporally related to
- physically related to
- spatially related to
- conceptually related to

*indicates associated with associated disease

pharmacokinetics of
References


References

- McCray AT, Nelson SJ.  
  *The representation of meaning in the UMLS.*  

- McCray AT.  
  *An upper level ontology for the biomedical domain.*  
Medical Ontology Research

Contact: olivier@nlm.nih.gov
Web: mor.nlm.nih.gov

Olivier Bodenreider
Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA