Using DL to Support a Very Large Healthcare Terminology: Successes and Challenges

Kent A. Spackman, MD, PhD
Chief Terminologist, IHTSDO
Clinical Professor, Oregon Health & Science University, Portland Oregon USA

Description Logics 2008
Dresden – May 14, 2008
About me

- **Current:**
  - Chief Terminologist for IHTSDO
  - Clinical Professor of Pathology & Medical Informatics, Oregon Health & Science University, Portland, Oregon, USA
- **Background primarily in academic medical informatics, and clinical pathology**
  - MD - Canada. Primary care training.
  - PhD in CS, machine learning, U of Illinois at Urbana-Champaign
  - Specialty training and (former) practice in hematopathology, blood banking, transfusion medicine & coagulation
- **1997-2007 Scientific Director of SNOMED, for the College of American Pathologists**
About SNOMED CT

- **Name:**
  - Systematized Nomenclature of Medicine - Clinical Terms
- **Description:**
  - A work of clinical terminology
- **Main purpose:**
  - Coded representation of meanings used in health information
About IHTSDO

- **International Health Terminology Standards Development Organization**
  - Formed in Denmark, 23\textsuperscript{rd} March 2007
    - Just over a year old
  - Area of standardization:
    - Terminology for interoperability of electronic health information
The purpose of the IHTSDO

- To acquire, own and administer the rights to SNOMED CT and other relevant assets (collectively, the "Terminology Products");
- To develop, maintain, promote and enable the uptake and correct use of its Terminology Products around the world;
- To undertake activities required to achieve these purposes
Status of the IHTSDO

- The IHTSDO is a Danish Association
- The Association is a registered not-for-profit entity in Denmark [23rd March 2007]
- Articles of Association detail the who, what, where and how of the Association
  - http://www.ihtsdo.org/about-us/governance/
- The Association owns the intellectual property
- Intellectual property in SNOMED CT and antecedent works (SNOMED 3.5, RT etc.) transferred to the IHTSDO [26th April 2007]
Status of the IHTSDO

• Members are countries
  – Eligible Members are all voting members of the United Nations
• The Members control the organization and the Articles of Association; [subject to Danish Law]
• Nine Charter [initial] Members:

  Australia, Canada, Denmark, Lithuania, Netherlands, New Zealand, Sweden, United Kingdom, United States of America
Status of the IHTSDO

- Physical office in the IT University in Copenhagen
  - Web site www.ihtsdo.org
- Three year support contract with the College of American Pathologists [First IHTSDO Release July 2007]
- New roles within the IHTSDO [at outset interim]; recruiting permanent posts
- At outset:
  - Chief Executive Officer [Ulrich Andersen]
  - Chief Terminologist [Kent Spackman]
  - Chief Quality Officer [Ed Cheetham]
  - Small number of administrative staff
What’s different now?

- **Articles of association**
- **2.3 Principles**
  - 2.3.1 The Association will seek to govern itself and conduct all of its activities in accordance with principles of **openness**, **fairness**, **transparency** and accountability to its Members.
What’s different now?

• Open collaborative working
• For access to IHTSDO committee documents, work groups, project groups and discussion forums:
  – Send email to support@ihtsdo.org
What’s different now?

- The Association’s work is funded by annual Membership fees paid by the Member nations
  - “Fair share” is determined by World Bank GNI Atlas
- Use in Member territories does not carry an IHTSDO fee
- Use in non-Member territories comes under a single world-wide license (called an “affiliate license”) [more on licensing later]
- SNOMED CT is much more available (no fees for evaluation or for qualifying research projects)
What’s different now?

• Genuine and increased intensity of efforts to cooperate and harmonize with other standards bodies
• Active discussions are ongoing with:
  – HL7, LOINC, IUPAC, WHO, WONCA
• Other contacts have been made or are planned between IHTSDO and:
  – CEN, ISO, IHE, openEHR, RadLex, & others
What’s different now?

• Role of the College of American Pathologists
  – Support Organization, by contract (3 years)
  – Perform maintenance and release
  – NO governance responsibility
  – Name change to reflect changed role and status
    • SNOMED Terminology Solutions (CAP STS)
IHTSDO Model: Organizational Relationships

International release available to all Member Nations

Cooperation between & among countries

Principles of fairness, openness, transparency
Governance

- Each Member Country appoints a representative to the General Assembly
- The General Assembly governs the Association (the SDO).
- Members of the Management Board (MB) are elected by the General Assembly (GA)
- Number of MB Directors is from 3 to 12 (GA determines)
  - Currently nine (one from each Charter Member)
Committees

• Four Standing Committees
  – Content
  – Technical
  – Research & Innovation
  – Quality assurance
• Committee members are elected by the General Assembly
• Worthy of note that Members have nominated and individuals have been elected from outside Member jurisdictions
  – E.g.: Stefan Schulz - Content Committee
• Committee meetings are also open to the public (but only elected committee members have a blanket right to participate)
Working Groups

- **Two types of Working Group** [both are entirely open]
  - Special Interest Group (SIG): [Domain focused e.g. profession (nursing), specialism (mapping)]
  - Project Group: [Task focused]

- **First Global Profession SIG for Pathology and Laboratory Medicine-September** [Special thanks to World Association of Societies of Pathology and Laboratory Medicine]
Special Interest Groups

- **Open** working group meetings + on-line discussion forums
- **Active Special Interest Groups (SIGs):**
  - Concept model SIG
  - Mapping SIG
  - Content-area focused SIGs
    - Pharmacy
    - Primary care
    - Nursing
    - Anesthesiology
    - Pathology and Laboratory Medicine
    - ...
  - ...
Affiliates

- **What is an “affiliate”?**
  - Anyone other than a member nation who has a license (the “Affiliate License”) for use of IHTSDO Terminology Products

- **What is the affiliate license?**
  - The single world-wide license for use of SNOMED CT
Licensing of SNOMED CT

- Single form of end-user license (Affiliate License)
- SNOMED CT is made much more available
- No charge for research purposes
- Permits world-wide use of SNOMED CT
- Affiliate licensees pay:
  - NO fees to IHTSDO for use in any Member nation. All obligations are met by the Member through their IHTSDO membership agreement. Cost-recovery is permitted
  - Charges as set by the IHTSDO for use in non-member nations (for implementation in clinical record systems)
Harmonization Boards - Recent Efforts

- **HL7**
  - Positive discussions, Aug. in Brisbane, Sept. in Atlanta
  - Discussions propose joint endorsement of Terminfo work product as a standard
- **WHO**
  - Positive discussions, August 2007 Brisbane
  - Milestones and high level work proceeding
  - Goal to have a joint harmonization board begin in February 2008
- **LOINC**
  - Positive discussions Sept Atlanta & after
  - Detailed preparatory work proceeding; no milestones yet agreed
- **IUPAC**
  - Detailed preparatory work - October 2007
- **WONCA**
Brief history of the terminologies that went into SNOMED CT

• **SNOMED**
  - developed by the College of American Pathologists (esp. Roger Côté of the University of Sherbrooke, Quebec, Canada)
  - 1979: SNOMED 2, widely adopted by surgical pathology departments worldwide
  - 1993: SNOMED 3, expanded and enhanced, but not widely adopted. Changed meanings of codes from prior version.

• **Read Codes**
  - Developed by James Read, a GP in UK.
  - 1993: Read v2 Adopted by UK NHS for GP electronic records
  - 1992 to 1999: Clinical Terms projects, resulted in Clinical Terms version 3 (CTV3). Greatly expanded and enhanced, but not widely adopted.
Brief History - continued

• 1996 to 2000:
  – CAP and Kaiser Permanente (KP) jointly worked to expand and revise SNOMED 3 to produce SNOMED Reference Terminology (RT) - the first version to be based on a DL foundation

• 1999 to 2002:
  – UK NHS and CAP, along with KP, jointly worked to merge CTV3 and SNOMED RT

- CTV3: 55%
- SNOMED RT: 31%
- Both: 8%
- New: 6%
Purpose of the Terminology

- To represent health information
  - Recorded by clinicians
  - At the level of detail they prefer
    - Not forced into arbitrary categories
    - In coordination with a known information model
- To retrieve and analyze health information
  - According to its meaning, not merely its surface form
  - To enable
    - Decision support for individual cases
    - Population-based aggregation and analysis
Are there SNOMED implementations?

- There is no such thing as a SNOMED Implementation, but
- There are implementations of electronic health record systems that utilize SNOMED CT
Decision support

- **Central to the Value of Semantic Interoperability**
  - Numerous studies document the ability of computerized decision support to decrease costs and improve quality
  - But use is limited
  - One major barrier is lack of standardization
  - Clinical terminology standards help fill this need
    - but we aren’t there yet
Terminology enables decision support:

influenza vaccination

- decision support program criterion:
  - chronic cardio respiratory disorders

- patient record:
  - mild persistent asthma
Terminology enables decision support:

**hemoglobin A₁C interpretation**

- **decision support program asks for:**
  - hereditary anemia due to disturbance of hemoglobin synthesis

- **patient record says:**
  - $A_\gamma \beta^+$ HPFH and $\beta^0$ thalassemia in cis
Terminology enables decision support:

antibiotic therapy

• decision support program asks for:
  – bacterial effusions

• patient record says:
  – tuberculous ascites
State of the art?

• Currently implemented systems are a long way from standardized delivery of semantic interoperability of clinical data
What’s the problem?

- No single barrier
  - Inertia of existing systems
  - Cost of change & lack of clear return for investments in change
- Barriers due to questions about standards:
  - Choice of different standards for same purpose
  - Inadequate coordination between those with different purposes (e.g. terminology vs. information model)
  - Quality, reliability, and implementability
Categories of Standards to Support Interoperability

- Data exchange / messaging
- **Terminology standards**
- Document standards
- Information Model / EHR standards
  - Architecture standards
  - Application standards
What SNOMED is not:

- **SNOMED** is not an attempt to
  - standardize the *language* of health care providers
  - get everyone to speak the same language
- “We are not the language police”
- Language is very changeable, fluid, and context-laden
- Clinicians (and people in general) determine what words mean by how they use them.

- **SNOMED attempts to properly reflect** the meanings given to words and phrases by people
What SNOMED is not:

- **SNOMED is not** an attempt to independently create standard meanings for health professions or scientists
  - It follows existing published standards
  - It seeks to encourage scientific and professional groups to come to consensus and publish standards

- For example:
  - **ISBT** (International Society of Blood Transfusion) publishes terminology for Human Red Blood Cell Surface Antigens.
  - **SNOMED CT** attempts to properly reflect and integrate the standard terminology so that it is usable with all others
What SNOMED is not:

• SNOMED is not a complete knowledge base
  – It represents *terminological knowledge* only
    • Definitional, always necessarily true of each instance
  – It does *not* represent *assertional knowledge*
    • Uncertain, variable from case to case

• Example: Appendicitis
  – Terminological knowledge:
    • Inflammation located in the appendix
  – Assertional knowledge:
    • Associated with anorexia, nausea, abdominal pain initially central but moving to RLQ, rebound tenderness over McBurney’s point, and increased WBC
Building blocks

• Concepts
  – The anchors for meaning

• Descriptions
  – Terms (strings of readable characters) used to express the meanings of the concepts in human language

• Relationships
  – Concept-to-concept links used to express information in computer-processable language
    • First purpose: formal logical meanings
    • Other purposes: tracking retired concepts, representing “facts” that may vary, and supporting post-coordination by suggesting valid qualifiers
Types of concepts (classes)

Current 19 top levels of hierarchies (January 2008 release):

- situation with explicit context
- clinical finding
- procedure
- body structure
- observable entity
- organism
- substance
- pharmaceutical / biologic product
- physical object
- physical force
- event
- environment or geographical location
- social context
- specimen
- record artifact
- stage or scale
- linkage concept
- qualifier value
- special concept
Fully Specified Name (FSN)

- Each concept has a name that has no acronyms and no hidden context

- Example:
  
  CT angiography

  FSN =
  
  angiography by computed tomography with contrast (procedure)

- The word in parens at the end is called the “FSN tag”
### Current 41 FSN tags (January 2008 release):

<table>
<thead>
<tr>
<th>TAG NAME</th>
<th># active</th>
</tr>
</thead>
<tbody>
<tr>
<td>disorder</td>
<td>63564</td>
</tr>
<tr>
<td>procedure</td>
<td>45422</td>
</tr>
<tr>
<td>finding</td>
<td>32559</td>
</tr>
<tr>
<td>organism</td>
<td>27948</td>
</tr>
<tr>
<td>body structure</td>
<td>25627</td>
</tr>
<tr>
<td>substance</td>
<td>23456</td>
</tr>
<tr>
<td>product</td>
<td>19081</td>
</tr>
<tr>
<td>qualifier value</td>
<td>8795</td>
</tr>
<tr>
<td>observable entity</td>
<td>7740</td>
</tr>
<tr>
<td>physical object</td>
<td>4485</td>
</tr>
<tr>
<td>morphologic abnormality</td>
<td>4303</td>
</tr>
<tr>
<td>occupation</td>
<td>4161</td>
</tr>
<tr>
<td>event</td>
<td>3575</td>
</tr>
<tr>
<td>situation</td>
<td>3300</td>
</tr>
<tr>
<td>regime/therapy</td>
<td>2949</td>
</tr>
<tr>
<td>attribute</td>
<td>1126</td>
</tr>
<tr>
<td>environment</td>
<td>1109</td>
</tr>
<tr>
<td>specimen</td>
<td>1052</td>
</tr>
<tr>
<td>assessment scale</td>
<td>884</td>
</tr>
<tr>
<td>cell</td>
<td>609</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TAG NAME</th>
<th># active</th>
</tr>
</thead>
<tbody>
<tr>
<td>geographic location</td>
<td>607</td>
</tr>
<tr>
<td>cell structure</td>
<td>510</td>
</tr>
<tr>
<td>person</td>
<td>370</td>
</tr>
<tr>
<td>navigational concept</td>
<td>346</td>
</tr>
<tr>
<td>ethnic group</td>
<td>262</td>
</tr>
<tr>
<td>tumor staging</td>
<td>214</td>
</tr>
<tr>
<td>record artifact</td>
<td>200</td>
</tr>
<tr>
<td>physical force</td>
<td>171</td>
</tr>
<tr>
<td>religion/philosophy</td>
<td>145</td>
</tr>
<tr>
<td>namespace concept</td>
<td>79</td>
</tr>
<tr>
<td>administrative concept</td>
<td>28</td>
</tr>
<tr>
<td>social concept</td>
<td>26</td>
</tr>
<tr>
<td>life style</td>
<td>21</td>
</tr>
<tr>
<td>racial group</td>
<td>19</td>
</tr>
<tr>
<td>staging scale</td>
<td>15</td>
</tr>
<tr>
<td>link assertion</td>
<td>8</td>
</tr>
<tr>
<td>inactive concept</td>
<td>7</td>
</tr>
<tr>
<td>environment / location</td>
<td>1</td>
</tr>
<tr>
<td>linkage concept</td>
<td>1</td>
</tr>
<tr>
<td>special concept</td>
<td>1</td>
</tr>
<tr>
<td>SNOMED RT+CTV3</td>
<td>1</td>
</tr>
</tbody>
</table>
Largest Hierarchies

- Clinical finding / disorder
- Procedure / regime / therapy
Most important “value” hierarchies

- *Body structure*, cell structure, cell
- Morphologic abnormality
- Qualifier values
- *Substances*
- *Observables*
- *Organisms*
Most important hierarchies based on number of defining relationships that use them as values.
Successes

- Collaborative process for editing
- Conversion of SNOMED 3 to a DL foundation
- Merger with CTV3 without losing DL foundation
- Maintenance of the hierarchies
- Normal forms
- Role grouping (as a compromise)
- Introduction of role hierarchies
Successes

- Collaborative process for editing
- Conversion of SNOMED 3 to a DL foundation
- Merger with CTV3 without losing DL foundation
- Maintenance of the hierarchies
- Normal forms
- Role grouping (as a compromise)
- Introduction of role hierarchies
Collaborative Development

- Fundamental to the initial work on SNOMED RT undertaken by Kaiser Permanente and C.A.P.
Content Development Tools (1998)

- **Ontyx TDE (Ontylog classifier)**
  - Classifies with first-order logic subset
  - Generates change sets for all changes

- **Metaphrase Finder (Lexical Technologies Inc)**
  - Lexical Closure, Suggested Relationships
  - Generates change sets for all changes

- **Galapagos Configuration Tools (KP - Campbell)**
  - Configuration control and conflict identification
  - Interactively resolve conflicts
CMT Development Cycle

Step 1: Acquire CMT
Step 2: Enhance CMT
Step 3: Return changes
Step 4: Identify and resolve conflicts
Onyx Editor
CMT Development Cycle

- Step 1: Acquire CMT
- Step 2: Enhance CMT
- Step 3: Return changes
- Step 4: Identify and resolve conflicts
Metaphrase Suggestions
CMT Development Cycle

1. Acquire CMT
2. Enhance CMT
3. Return changes
4. Identify and resolve conflicts
Conflict Resolution
CMT Development Cycle

Step 1: Acquire CMT

Step 2: Enhance CMT

Step 3: Return changes

Step 4: Identify and resolve conflicts
Configuration Management

- **Change-Set Configuration Management**
  - Supports distributed, concurrent change
  - Provides the foundation for metrics and reproducible, measurable development processes
URU Criteria

1. Definitions should be Understandable by average clinicians, given brief explanations
2. We assess understandability by examining Reproducibility
3. We can ignore distinctions for which we see no Use in health care
Evolutionary Design

- Evolution without pre-ordained design
- Accumulation of desirable features
- Heterogeneity of perspectives

Dealing with Scale
- Participatory consensus-based approach
  - Involve the experts
- Semantics-based concurrency control
  - Description logic foundation
- Configuration management tools
  - Keith Campbell’s “Galapagos” tool set
Successes

• Collaborative process for editing
• **Conversion of SNOMED 3 to a DL foundation**
• Merger with CTV3 without losing DL foundation
• Maintenance of the hierarchies
• Normal forms
• Role grouping (as a compromise)
• Introduction of role hierarchies
# About SNOMED's use of DL

<table>
<thead>
<tr>
<th>SNOMED version</th>
<th>Concept &amp; Role-forming Operators</th>
<th>Role axioms</th>
<th>Language</th>
<th>Role grouping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early work (1996-1999)</td>
<td>$(\cap, \exists R:C)( )$</td>
<td>$\mathcal{EL}$</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>SNOMED RT (2000-2001)</td>
<td>$(\cap, \exists R:C)(+)$</td>
<td>$\mathcal{EL}^+$</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>SNOMED CT (Jan02-Jan04)</td>
<td>$(\cap, \exists R:C)( )$</td>
<td>$\mathcal{EL}$</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>SNOMED CT (Jul04-present)</td>
<td>$(\cap, \exists R:C)(+)$</td>
<td>$R \subseteq S$</td>
<td>$\mathcal{EL}^H^+$</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notation mostly follows Donini in Ch.3 Description Logic Handbook

(+ means right identities were used)
### Using DL for the SNOMED RT hierarchy (2000)

<table>
<thead>
<tr>
<th></th>
<th>P axis</th>
<th>D axis</th>
<th>Others</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stated is-a rels.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>removed</td>
<td>4,214</td>
<td>15,838</td>
<td>390</td>
<td>20,442</td>
</tr>
<tr>
<td><strong>Inferred is-a rels.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>added</td>
<td>15,094</td>
<td>10,557</td>
<td>217</td>
<td>25,868</td>
</tr>
</tbody>
</table>

Number of immediate is-a relationships changed by algorithmic subsumption calculations

Transitive reduction removed many is-a relationships from the D axis because of modeling style
Right identity
(restricted role value maps)

• $R \circ S \subseteq R$
• $xRy \land ySz \rightarrow xRz$

• femurFracture $\subseteq \exists$ site.femur
• headOfFemurFracture $\subseteq \exists$ site.headOfFemur
• headOfFemur $\subseteq \exists$ part-of.Femur

• Allows the automated inference that:
  – headOfFemurFracture $\subseteq$ FemurFracture

• But this isn’t the purpose for which we use right identity in the current release!
Right identity
(restricted role value maps)

- $R \circ S \subseteq R$
- $xRy \land ySz \rightarrow xRz$

- allergyToAspirin $\subseteq \exists$ causativeAgent.aspirinSubstance
- aspirinProduct $\subseteq \exists$ hasActiveIngredient.aspirinSubstance
- allergyToAspirinProduct $\subseteq \exists$ causativeAgent.aspirinProduct

- Allows the automated inference that:
  - allergyToAspirinProduct $\subseteq$ allergyToAspirin
Successes

- Collaborative process for editing
- Conversion of SNOMED 3 to a DL foundation
- **Merger with CTV3 without losing DL foundation**
- Maintenance of the hierarchies
- Normal forms
- Role grouping (as a compromise)
- Introduction of role hierarchies
Pressure to Abandon Strict DL Semantics

- CTV3 was designed to allow navigation by clinicians
- Arbitrary navigation concepts were allowed
  - “Drugs A-F”, “Drugs G-S”, “Drugs T-Z”
- Concerns about how natural to a clinician the hierarchies would look
- Concerns that consistent models were not possible (every doctor is Humpty Dumpty “a word means whatever I choose it to mean”)
How a GP might like to navigate to Influenza A virus from “virus”:

- Virus
- Influenza A virus

It is six levels deep if you try to navigate the is-a hierarchy:

1. Virus
2. RNA virus
3. Enveloped ssRNA virus without a DNA step in life-cycle
4. Enveloped ssRNA virus without a DNA step with multiple-stranded negative-sense genome
5. Family Orthomyxoviridae
6. Genus Influenzavirus A
7. Influenza A virus
Solution

- Separate “reference properties” and “interface properties” of the terminology
- Navigation to find a term is an interface property
- Preferred term selection is an interface property
- Correct definition of meaning is a “reference property”
Avoiding Right Identities by Using SEP Triplets

Liver Structure:
- XM0Ps Liver structure
- T-62000 Liver

Liver Part:
- T-D0535 Liver part

Entire liver:
- 7N330 Liver

Lobe of liver
Successes

• Collaborative process for editing
• Conversion of SNOMED 3 to a DL foundation
• Merger with CTV3 without losing DL foundation
• **Maintenance of the hierarchies**
• Normal forms
• Role grouping (as a compromise)
• Introduction of role hierarchies
New Content Additions

Number of New Active Concepts per Release

- Jul 02
- Jan 03
- Jul 03
- Jan 04
- Jul 04
- Jan 05
- Jul 05
- Jan 06
- Jul 06
- Jan 07
- Jul 07
- Jan 08

The chart shows the number of new active concepts added per release from Jul 2002 to Jan 2008.
Concept Model Change

Number of Attributes with Defining Relationships

Jan 02 - Jul 02
Jan 03 - Jul 03
Jan 04 - Jul 04
Jan 05 - Jul 05
Jan 06 - Jul 06
Jan 07 - Jul 07
Jan 08
Percentage of SNOMED CT concept codes that are “sufficiently defined”
Successes

- Collaborative process for editing
- Conversion of SNOMED 3 to a DL foundation
- Merger with CTV3 without losing DL foundation
- Maintenance of the hierarchies
- **Normal forms**
- Role grouping (as a compromise)
- Introduction of role hierarchies
Definition of Normal Forms

- In original RT work, dual independent modeling required exact agreement on *stated* definition
  - Resulted in unresolved arguments about modeling style
    - State most immediate parent concepts only, and only those relationships that have changed, or
    - State proximal primitives only, and all defining relationships

- Defining a normal form allowed different modeling styles for different purposes or preferences

Successes

- Collaborative process for editing
- Conversion of SNOMED 3 to a DL foundation
- Merger with CTV3 without losing DL foundation
- Maintenance of the hierarchies
- Normal forms
- Role grouping (as a compromise)
- Introduction of role hierarchies
Role Grouping as a Compromise

- Implementers and modelers fear/loathe nesting of expressions
  - Nesting violates simple flat frame-based model

- Reality demands faithful representation

- Role grouping attempted (with partial success) to hide the complexity
  - But it was misunderstood by some in DL community as being a proprietary hack

Need for Role Groups

• When a single concept may have more than one value for a particular attribute
  – for example, “bone fusion with tendon transfer”
    • method = fusion, site = bone, and
    • method = transfer, site = tendon
• And, one attribute-value pair needs to be associated with another.
  – How can we specify that the fusion is done to the bone and not to the tendon? and that the transfer is done to the tendon and not to the bone?
Role Groups as a Solution

- Informally:
  - don’t nest or create sub procedures
  - simply “group” the attribute-value pairs

- Using curly braces as a syntactic marker:
  \{ site=bone, method=fusion\},
  \{site=tendon, method=transfer\}

- Or, in tabular form, use a “group” column:
<table>
<thead>
<tr>
<th>attr</th>
<th>value</th>
<th>group</th>
</tr>
</thead>
<tbody>
<tr>
<td>site</td>
<td>bone</td>
<td>1</td>
</tr>
<tr>
<td>method</td>
<td>fusion</td>
<td>1</td>
</tr>
<tr>
<td>site</td>
<td>tendon</td>
<td>2</td>
</tr>
<tr>
<td>method</td>
<td>transfer</td>
<td>2</td>
</tr>
</tbody>
</table>
Role Grouping Logical Form:  
A Nested Existential Restriction

- \( C \equiv \exists R_{RG}.(\exists R_1.C_1 \sqcap \exists R_2.C_2) \sqcap \exists R_{RG}.(\exists R_3.C_3) \)

- Distributed as three 4-tuples in relationships table:
  
  \[
  \begin{array}{cccc}
  C & R_3 & C_3 & 0 \\
  C & R_1 & C_1 & 1 \\
  C & R_2 & C_2 & 1 \\
  \end{array}
  \]

  - Role group numbers are arbitrary integers, and not designed to be stable across changes in the concept definition
Successes

- Collaborative process for editing
- Conversion of SNOMED 3 to a DL foundation
- Merger with CTV3 without losing DL foundation
- Maintenance of the hierarchies
- Normal forms
- Role grouping (as a compromise)
- Introduction of role hierarchies
Role (attribute) hierarchies

- Selected SNOMED CT attributes have a hierarchical relationship to one another known as “role hierarchies.” In a role hierarchy, one general attribute is the parent of one or more specific subtypes of that attribute. Concepts defined using the more general attribute can inherit concepts modeled with the more specific subtypes of that attribute.
Role hierarchies - procedures

- PROCEDURE DEVICE
  - DIRECT DEVICE
  - INDIRECT DEVICE
  - USING DEVICE
  - USING ACCESS DEVICE

- PROCEDURE MORPHOLOGY
  - DIRECT MORPHOLOGY
  - INDIRECT MORPHOLOGY

- PROCEDURE SITE
  - PROCEDURE SITE - DIRECT
  - PROCEDURE SITE - INDIRECT
Role hierarchies - clinical findings

• ASSOCIATED WITH role hierarchy:
• ASSOCIATED WITH
  – AFTER
  – DUE TO
  – CAUSATIVE AGENT
Challenges & New Directions

• **Quality assurance**
  - Finding errors
    - Model errors
    - Individual concepts
  - Correcting errors
  - Maintaining correctness

• **Tooling for editing and collaborative work**

• **Education of clinician and scientist editors**

• **Technical challenges**
  - Migrating to greater expressiveness
    - Balancing expressivity and classifier complexity
  - Specific issues of implementation
    - Post-coordination, crossovers
Challenges & New Directions

- **Quality assurance**
  - Finding errors
    - Model errors
    - Individual concepts
  - Correcting errors
  - Maintaining correctness

- **Tooling for editing and collaborative work**

- **Education of clinician and scientist editors**

- **Technical challenges**
  - Migrating to greater expressiveness
    - Balancing expressivity and classifier complexity
  - Specific issues of implementation
    - Post-coordination, crossovers
Transitive Closure Stability

Remaining valid is-a's from each release over subsequent release sequence.
Retiring Erroneous Content

Identified and Retired Per Release

- Duplicate
- Ambiguous
Challenges & New Directions

• Quality assurance
  – Finding errors
    • Model errors
    • Individual concepts
  – Correcting errors
  – Maintaining correctness

• Tooling for editing and collaborative work

• Education of clinician and scientist editors

• Technical challenges
  – Migrating to greater expressiveness
    • Balancing expressivity and classifier complexity
  – Specific issues of implementation
    • Post-coordination, crossovers
Tools

- Much better tools are needed to support
  - Collaborative editing
  - Conflict resolution
  - Quality assurance and auditing

- A possible platform for open-source developments:
  - OHT (Open Health Tools)
    - www.openhealthtools.org
Challenges & New Directions

• **Quality assurance**
  - Finding errors
    • Model errors
    • Individual concepts
  - Correcting errors
  - Maintaining correctness

• **Tooling for editing and collaborative work**

• **Education of clinician and scientist editors**

• **Technical challenges**
  - **Migrating to greater expressiveness**
    • Balancing expressivity and classifier complexity
  - **Specific issues of implementation**
    • Post-coordination, crossovers
### Concept & role forming operators & terminological axioms

<table>
<thead>
<tr>
<th>Name of construct</th>
<th>Notation</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primitive concept</td>
<td>$A$</td>
<td>$A^I \subseteq \Delta^I$</td>
</tr>
<tr>
<td>Primitive role</td>
<td>$R$</td>
<td>$R^I \subseteq \Delta^I \times \Delta^I$</td>
</tr>
<tr>
<td>Top</td>
<td>$\top$</td>
<td>$\Delta^I$</td>
</tr>
<tr>
<td>Bottom</td>
<td>$\bot$</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>Conjunction</td>
<td>$C \sqcap D$</td>
<td>$C^I \cap D^I$</td>
</tr>
<tr>
<td>Exists restriction</td>
<td>$\exists R.C$</td>
<td>${x</td>
</tr>
<tr>
<td>Disjunction</td>
<td>$C \sqcup D$</td>
<td>$C^I \cup D^I$</td>
</tr>
<tr>
<td>Negation</td>
<td>$\neg C$</td>
<td>$\Delta^I \setminus C^I$</td>
</tr>
<tr>
<td>Value restriction</td>
<td>$\forall R.C$</td>
<td>${x</td>
</tr>
<tr>
<td>Role composition</td>
<td>$R_1 \circ \ldots \circ R_n$</td>
<td>$R^I_1 \circ \ldots \circ R^I_n$</td>
</tr>
<tr>
<td>Restricted role value maps</td>
<td>$R \circ S \sqsubseteq R$</td>
<td>$x R^I y \wedge y S^I z \rightarrow x R^I z$</td>
</tr>
<tr>
<td></td>
<td>$R \circ S \sqsubseteq S$</td>
<td>$x R^I y \wedge y S^I z \rightarrow x S^I z$</td>
</tr>
<tr>
<td></td>
<td>$R \circ S \sqsubseteq T$</td>
<td>$x R^I y \wedge y S^I z \rightarrow x T^I z$</td>
</tr>
<tr>
<td>Concept definition</td>
<td>$A \equiv C$</td>
<td>$A^I = C^I$</td>
</tr>
<tr>
<td>Primitive concept introduction</td>
<td>$A \sqsubseteq C$</td>
<td>$A^I \subseteq C^I$</td>
</tr>
<tr>
<td>Primitive role introduction (role hierarchy)</td>
<td>$R \sqsubseteq S$</td>
<td>$R^I \subseteq S^I$</td>
</tr>
</tbody>
</table>
Should we add more expressive DL features?

- General concept inclusion axioms
- Transitive roles
- Reflexive roles
- Disjointness axioms
- Value restrictions
- Negation
- Disjunction
- Cyclic definitions
- Number restrictions
General concept inclusion axioms

- Extremely useful feature
- Compatible with a polynomial-time structural subsumption algorithm

- Allows us to say what is true in addition to what is sufficient
  - Gastric ulcer is located in the stomach, and in addition it necessarily involves the gastric mucosa
Transitive roles

• \( xRy \land yRz \rightarrow xRz \)

• Useful for causal/associational chains

• Interaction with role hierarchy is interesting & useful

• Example: Associated-with-after
  – Varicella (chicken pox)
    – An infection with causative-agent = varicella virus
  – Herpes zoster
    – Also has causative-agent = varicella virus, and occurs after varicella
  – Post-herpetic neuralgia
    – Occurs after herpes zoster (therefore occurs after varicella), but is not an infection with causative-agent varicella virus
Reflexive roles

- Plan to introduce reflexive “part-of” as a way of handling “SEP” model evolution

\[
\begin{align*}
\text{proper-part-of} & \subseteq \text{part-of} \\
\epsilon & \subseteq \text{part-of} \\
S & \equiv \exists \text{ part-of} . E \\
P & \equiv \exists \text{ proper-part-of} . E
\end{align*}
\]

Value restriction $\forall R.C$

- Not an intuitive construct
  - $\text{person} \cap \forall \text{hasCar}.\text{Jaguar}$
  - Includes people who have no car, but if they had one it would have to be a Jaguar. . . . Do we encounter this kind of concept in common-sense thinking?

- Creates pernicious interactions with disjunction and negation that tend to make structural subsumption algorithms incomplete

- But it was included in $\mathcal{ALC}$ and $\mathcal{FL}^-$, so languages including it were studied extensively.
Negation $\neg C$

- Head injury without loss of consciousness

$\neg headInjury \sqcap \neg lossOfConsciousness$

$situation \sqcap$

$\exists includes.headInjury \sqcap$

$\neg \exists includes.lossOfConsciousness$
Disjunction $C \sqcup D$

- Some high-level aggregators are naturally disjunctive
- We can address this need partially by using navigation hierarchies
Cyclic definitions, number restrictions

• ? No significant need for these at present