mapping local terminology to SNOMED-CT for eHR in Hong Kong
125676002 person

- 48176007 social context
- 224627004 general categories of person
- 394619001 guardian
- 262043009 partner in relationship
- 438347000 person categorized by
- 410598002 person categorized by
- 428792000 person exposed to disease
- 303119007 person in the community
- 303071001 person in the family
- 303118004 person in the healthcare
- 444018008 person with characteristics
- 419358007 subject of record or other
Edward Tam (person)
Edward Tam

Health Informatics Analyst
eHR Information Standards Office

Mail: edwardtam@ha.org.hk
Phone: +852 3919 2205
123037004 | body structure

- 442083009 | anatomical or acquired
- 91832008 | anatomical organization
- 258331007 | anatomical site notation
- 118956008 | morphologically altered
- 278001007 | nonspecific site
- 361083003 | normal anatomy
- 91722005 | physical anatomical entity
- 21229009 | topography not assigned
- 87100004 | topography unknown
• electronic Health Record (eHR) sharing system

  – project by Hong Kong SAR government
  – territory-wide health record sharing platform
  – both public & private health care providers
  – live-run expected at end of 2014

  – need an information standard for data consistency
• Hong Kong Clinical Terminology Table (HKCTT)
  – information standard for eHR sharing system
  – based on Hospital Authority Clinical Vocabulary (HACVT)
  – accumulated 39700 diagnosis & procedure concepts

  – how to link with SNOMED-CT for interoperability?
ideal for eHR use:

- comprehensive collection of concepts
- varied levels of granularity
• efficient concept searching during info capture
  – synonyms
  – multi-hierarchy relationships
  – definition with attributes
• easy customisation
  – post-coordination
  – easy extension of terminology, while
  – preserving semantic hierarchy
  – good for information retrieval and data analysis
general finding of observation

- appetite problem
- basic learning problem
- bladder problem
- body growth problem
- body odour problem
- body weight AND/OR gastrointestinal system problem
- breast problem
- cardiovascular system problem
- gastrointestinal tract problem
- problem identified with problem manageable
• HACVT ➔ HKCTT

– during the last two decades, accumulated:
  • 55M diagnosis entries
  • 20M procedure entries

– to ensure backward compatibility
• **SNOMED CT**

  – to enable use of SNOMED-CT by health care providers for capturing clinical records and transmitting to eHR

  – to enable use of SNOMED-CT for analysis of eHR records
• HKCTT vs SNOMED CT

– direct implementation of SNOMED-CT?

– mapping HKCTT to SNOMED-CT?
• directly implement SNOMED-CT:
• directly implement SNOMED-CT:
  – need to change working practice & culture!!
  – need data conversion for existing data!!
• map HKCTT to SNOMED-CT:

[Diagram showing data flow between HA, HACVT, eHR, ICD10, HKCTT, SNOMED-CT, and GPs]

To: WHO

Analysis
• map HKCTT to SNOMED-CT:
  
  – preserve working practice & culture
  
  – preserve existing data, need no data conversion
  
  – allowing data capture & analysis using SNOMED-CT

→ best of both worlds
308910008 reference documentation
258049002 protocols
229053005 registers
• basic mapping concepts
  – source
  – target
  – relationships
    • degree of equivalence
    • cardinality of the map
• degree of equivalence

  – exact match / equivalent match
  – approximate match
  – no match
• cardinality of the map
  – one to one
  – one to many
  – many to one
  – many to many
• mapping principles
  – purpose
    • rules and guidelines
  – standards-based
    • authoritative support
    • validation
  – map maintenance
    • organization in charge
    • update with source and targets
• mapping steps
  – develop business case / define use case
  – develop rules
  – plan pilot phase
  – develop full content
    • periodic testing, final QA testing
    • communicate with source / target owners
  – validation
• validation

  – understandable, reproducible, usable
    • map understood by user without manual
    • map developed by process straightforward and reproducible no matter who or what to perform the mapping
    • map usable for use case

  – external validation
    • conducted by entity not involved with development or financial or political interest
• mapping proceeded in **phases:**
  → exact matching
  → post-coordination

• **verification** to ensure quality
procedure by method
• exact matching

  – compare **descriptive meanings** between concepts in HKCTT and SNOMED-CT
  – ignore word form differences like spelling & grammar
• exact matching

– reviewers with clinical knowledge background
– **double-checked** by second reviewer

( HA 7453 ) Deafness of one ear

=  

( SCT 162342008 ) Unilateral deafness

?
• exact matching

  – 46% (18600) concepts exact matched
exact matched (46%)
The SPECIALIST Lexicon

The SPECIALIST Lexicon is an English lexicon (dictionary) that contains many frequently occurring English words. The lexical entry for each word or term includes:

- Syntactic (syntax information)
- Morphological (inflection, derivation, and composition information)
- Orthographic (spelling information)

Currently the SPECIALIST Lexicon contains over 200,000 terms. These terms are a subset of the Language Processing. Words are selected for entry into the Specialized Lexicon for the purpose of using them in the development of the UMLS Semantic Network. Terms in the SPECIALIST Lexicon are often the same as those in the TREC Medline Collection.
The Lexical Tools

The lexical tools are a set of computer programs designed to address the high degree of variability in medical terminology. Some of the primary programs are:

- A lexical variant generator (LVG)
- A normalized string generator (Norm)
- A word Index generator (Wordind)

Together these programs address the high degree of variability in medical terminology.
The Norm Program

The lexical program, Norm, generates the normalized strings for terms. The normalization process involves stripping possessives, replacing words that are considered as a specialization, such as "No Other Specification" or NOS, lower-casing each word, and sorting the words in alphabetic order.

Below is an example of the normalization process for the term Hodgkin's disease:

<table>
<thead>
<tr>
<th>Normalized Term</th>
<th>Original Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hodgkin's diseases, NOS</td>
<td>Hodgkin's disease, NOS-</td>
</tr>
</tbody>
</table>
The Norm Program

The lexical program, **Norm**, generates the normalized strings for terms included in the SNOMED CT vocabulary. The normalization process involves stripping possessives, replacing punctuation with spaces, and converting all words to lowercase. Terms such as "No Other Specification" or NOS are also lower-cased. Each word is then broken into separate strings and sorted alphabetically.

Below is an example of the normalization process for the term Hodgkin's diseases, NOS.

<table>
<thead>
<tr>
<th>Step</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remove genitive</td>
<td>Hodgkin diseases, NOS</td>
</tr>
<tr>
<td>Remove stop words</td>
<td>Hodgkin diseases,</td>
</tr>
<tr>
<td>Lowercase</td>
<td>Hodgkin diseases,</td>
</tr>
<tr>
<td>Uninflect</td>
<td>Hodgkin disease</td>
</tr>
<tr>
<td>Sort words</td>
<td>disease hodgkin</td>
</tr>
</tbody>
</table>

The Norm program is used in systems to:

- Find similar terms
- Map terms to UMLS concepts
- Find lexical variants for a term
• exact match verification

  – the **NORM** program from the **Lexical Tools** from the Unified Medical Language System (UMLS)
  – **perform normalization** on exact matched concept pairs (HKCTT concept & SNOMED-CT concept)
  – compare the two normalized lexical forms, if the same, the concept pair is “exact matched”
### Exact Match Verification

<table>
<thead>
<tr>
<th>NORM</th>
<th>(HA 15217) Macroductyilia (finger)</th>
<th>(SCT 69381005) macroductyilia of fingers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no punctuations</td>
<td>Macroductyilia finger</td>
<td>macroductyilia of fingers</td>
</tr>
<tr>
<td>no stop words</td>
<td>Macroductyilia finger</td>
<td>macroductyilia fingers</td>
</tr>
<tr>
<td>no upper case</td>
<td>Macroductyilia finger</td>
<td>macroductyilia fingers</td>
</tr>
<tr>
<td>no plurals</td>
<td>Macroductyilia finger</td>
<td>macroductyilia fingers</td>
</tr>
</tbody>
</table>

result = **SAME**

macrodactyilia finger

macrodactyilia finger
• exact match verification

  – out of 18600 concepts pairs exact matched,
  – 55% (10100) have same normalized lexical forms
• exact matched but failed lexical verification:

<table>
<thead>
<tr>
<th>HA 13806</th>
<th>Lumbar vertebral disease</th>
<th>SCT 129139009</th>
<th>Disorder of lumbar spine</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA 3494</td>
<td>Pigmented mole</td>
<td>SCT 400096001</td>
<td>Melanocytic nevus</td>
</tr>
<tr>
<td>HA 6130</td>
<td>Plexopathy</td>
<td>SCT 2231001</td>
<td>Nerve plexus disorder</td>
</tr>
<tr>
<td>HA 15325</td>
<td>Breast ectopic</td>
<td>SCT 18166000</td>
<td>Accessory breast</td>
</tr>
<tr>
<td>HA 37863</td>
<td>Hypoglossal pain</td>
<td>SCT 279059006</td>
<td>Hypoglossal neuralgia</td>
</tr>
<tr>
<td>HA 8296</td>
<td>Hypotension</td>
<td>SCT 45007003</td>
<td>Low blood pressure</td>
</tr>
</tbody>
</table>
• exact matched but failed lexical verification:

<table>
<thead>
<tr>
<th>(HA 31631) Ultrasonogram of kidney</th>
<th>(SCT 306005) Echography of kidney</th>
</tr>
</thead>
<tbody>
<tr>
<td>(HA 33095) Control of haemorrhage in pharynx</td>
<td>(SCT 172982005) Hemostasis of pharynx</td>
</tr>
<tr>
<td>(HA 24692) Gastrectomy, partial</td>
<td>(SCT 49209004) Subtotal gastrectomy</td>
</tr>
<tr>
<td>(HA 23917) Chest tapping</td>
<td>(SCT 91602002) Thoracentesis</td>
</tr>
<tr>
<td>(HA 27315) Elbow fusion</td>
<td>(SCT 58494002) Arthrodesis of elbow</td>
</tr>
<tr>
<td>(HA 27879) Excision of Extra Finger</td>
<td>(SCT 90722001) Removal of supernumerary finger</td>
</tr>
</tbody>
</table>
• submission to IHTSDO

  – concepts with no exact match (SNOMED-CT)

  – submit to International Health Terminology Standards Development Organisation (IHTSDO) for addition to the SNOMED-CT international core release
• submission to IHTSDO
  – 12% (4800) concepts submitted to IHTSDO
  – 3% (1100) concepts added to SNOMED-CT core
    → becomes exact match
submitted & exact matched (3%)
<table>
<thead>
<tr>
<th>Code</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>272103003</td>
<td>delayed priority</td>
</tr>
<tr>
<td>1033900000</td>
<td>elective</td>
</tr>
<tr>
<td>25876001</td>
<td>emergency</td>
</tr>
<tr>
<td>394849002</td>
<td>high priority</td>
</tr>
<tr>
<td>88694003</td>
<td>immediate</td>
</tr>
<tr>
<td>394848005</td>
<td>normal priority</td>
</tr>
<tr>
<td>76561005</td>
<td>reclassified</td>
</tr>
<tr>
<td>44408006</td>
<td>reclassified and rescheduled</td>
</tr>
<tr>
<td>64695001</td>
<td>repeat elective</td>
</tr>
<tr>
<td>21282002</td>
<td>repeat emergency</td>
</tr>
<tr>
<td>58334001</td>
<td>rescheduled</td>
</tr>
<tr>
<td>59941001</td>
<td>routine</td>
</tr>
</tbody>
</table>
• priority concept groups

  – concepts in low usage / no usage
  – concepts with special status (inactive, etc)
• priority concept groups

  – 16% (6600) concepts to lower priority mapping plan
to low priority (16%)
• post-coordination
  – concepts with no exact match (SNOMED-CT)
  – represent the concept using SNOMED-CT concepts
  – follow SNOMED-CT editorial guidelines
363787002 observable entity
• post-coordination mapping patterns
  – found common patterns among post-co mapping
- post-coordination

<table>
<thead>
<tr>
<th>mild cystitis</th>
<th>is a</th>
<th>cystitis (disorder)</th>
</tr>
</thead>
<tbody>
<tr>
<td>severity</td>
<td>mild</td>
<td>(qualifier value)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>severe aplastic anaemia</th>
<th>is a</th>
<th>aplastic anemia (disorder)</th>
</tr>
</thead>
<tbody>
<tr>
<td>severity</td>
<td>severe</td>
<td>(severity modifier)</td>
</tr>
<tr>
<td></td>
<td>(qualifier value)</td>
<td></td>
</tr>
</tbody>
</table>
• post-coordination mapping patterns

<table>
<thead>
<tr>
<th>Mild Dx</th>
<th>is a = Dx</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>severity = mild (qualifier value)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Moderate Dx</th>
<th>is a = Dx</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>severity = moderate (severity modifier) (qualifier value)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Severe Dx</th>
<th>is a = Dx</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>severity = severe (severity modifier) (qualifier value)</td>
</tr>
</tbody>
</table>
• post-coordination

<table>
<thead>
<tr>
<th>acute alcoholic psychosis</th>
<th>is a</th>
<th>alcohol-induced psychosis (disorder)</th>
</tr>
</thead>
<tbody>
<tr>
<td>clinical course</td>
<td></td>
<td>sudden onset and/or short duration (qualifier value)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>chronic adjustment disorder</th>
<th>is a</th>
<th>adjustment disorder (disorder)</th>
</tr>
</thead>
<tbody>
<tr>
<td>clinical course</td>
<td></td>
<td>chronic (qualifier value)</td>
</tr>
</tbody>
</table>
• post-coordination mapping patterns

<table>
<thead>
<tr>
<th>Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>acute Dx</td>
<td>is a = Dx, <strong>clinical course</strong> = sudden onset AND/OR short duration (qualifier value)</td>
</tr>
<tr>
<td>chronic Dx</td>
<td>is a = Dx, <strong>clinical course</strong> = chronic (qualifier value)</td>
</tr>
</tbody>
</table>
• post-coordination

<table>
<thead>
<tr>
<th>open trigeminal rhizotomy</th>
<th>is a</th>
<th>trigeminal rhizotomy (procedure)</th>
</tr>
</thead>
<tbody>
<tr>
<td>access</td>
<td>open approach - access (qualifier value)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>percutaneous cardiopulmonary bypass</th>
<th>is a</th>
<th>cardiopulmonary bypass operation (procedure)</th>
</tr>
</thead>
<tbody>
<tr>
<td>access</td>
<td>percutaneous approach - access (qualifier value)</td>
<td></td>
</tr>
</tbody>
</table>
- post-coordination mapping patterns

<table>
<thead>
<tr>
<th>open Px</th>
<th>is a = Px access = open approach - access (qualifier value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>closed Px</td>
<td>is a = Px access = closed approach - access (qualifier value)</td>
</tr>
<tr>
<td>percutaneous px</td>
<td>is a = px access = percutaneous approach - access (qualifier value)</td>
</tr>
</tbody>
</table>
post-coordination

<table>
<thead>
<tr>
<th>transurethral resection of bladder</th>
<th>is a</th>
<th>resection of lesion of bladder (procedure)</th>
</tr>
</thead>
<tbody>
<tr>
<td>surgical approach</td>
<td><strong>transurethral approach</strong> (qualifier value)</td>
<td></td>
</tr>
</tbody>
</table>
post-coordination mapping patterns

<table>
<thead>
<tr>
<th>trans-xxx Px</th>
<th>is a = Px</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>surgical approach</strong> = trans-xxx</td>
<td></td>
</tr>
</tbody>
</table>
- **post-coordination**

<table>
<thead>
<tr>
<th>post infarct dementia</th>
<th>is a</th>
<th>dementia (disorder)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>after</strong></td>
<td></td>
<td>cerebral infarction (disorder)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>aseptic meningitis, post-operative</th>
<th>is a</th>
<th>aseptic meningitis (disorder)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>after</strong></td>
<td></td>
<td>surgical procedure (procedure)</td>
</tr>
</tbody>
</table>
post-coordination mapping patterns

| Dx1, post-Dx2 | is a = Dx1  
|               | after = Dx2 |
| Dx, post-Px   | is a = Dx   
|               | after = Px (procedure) |
**post-coordination**

<table>
<thead>
<tr>
<th>history of disorders of nervous system and sense organs</th>
<th>is a</th>
<th>history of - disorder (situation)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>associated finding</strong></td>
<td></td>
<td>nervous system and sense organ diseases (disorder)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>family history of digestive disorder</th>
<th>is a</th>
<th>family history of disorder (situation)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>associated finding</strong></td>
<td></td>
<td>disorder of digestive system (disorder)</td>
</tr>
</tbody>
</table>
- post-coordination mapping patterns

<table>
<thead>
<tr>
<th>History of Dx</th>
<th>is a = history of - disorder (situation) associated finding = Dx (disorder)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family history of Dx</td>
<td>is a = family history of disorder (situation) associated finding = Dx (disorder)</td>
</tr>
</tbody>
</table>
• post-coordination mapping patterns
  – found 90+ post-coordination mapping patterns
  – **review & refine**, referencing existing SCT core concepts
  – **re-apply** the patterns to post-co similar HKCTT concepts

  – to maintain mapping **quality & consistency**
  – to improve mapping **efficiency**
meetings and conferences

- 229057006 meetings
- 229058001 case conference
- 2494005 nursing conference
- 311459003 team meeting
• mapping team discussion meetings
  – mapping team members
  – bi-weekly discussion meetings
  – to review mapping rules & practices
  – to review post-coordination patterns
• post-coordination

  – 26% (10400) concepts mapped by post-coordination
post-co’ed (26%)
• post-coordination verification

  – 5% (500 / 10400) sample of post-co mappings
  – 90 post-co mapping patterns

  – verification by overseas experts
• post-coordination verification

  – 56% (280 / 500) mappings
  – 60% (54 / 90) mapping patterns
    • expert agrees with HA, or
    • expert advice is taken

  – **re-apply** verified patterns to post-co similar concepts
• mapping results

  – 46% (18600) concepts exact matched
    • 55% (10100) mappings verified (NORM from Lexical Tools)
  – 3% (1100) concepts submitted & added to SNOMED-CT
  – 26% (10400) concepts mapped by post-coordination
    • 56% (280 / 500) mappings verified (overseas expert)
  – 16% (6600) concepts to lower priority
post-co’ed & verified (26%)

exact matched & verified

in progress (8%)

to low priority (16%)

submitted & exact matched (3%)

exact matched (46%)
procedure by method

discussion about a procedure by method

discussion about activity

discussion about behavior

discussion about change

discussion about children

discussion about disorder

discussion about ideal

discussion about options

discussion about preference

discussion about pregnancy

discussion about procedure
the mapping approach

- preserves working practice and culture
- preserves existing data, avoiding data conversion
• the mapping approach

  – enables eHR records to be captured, transmitted, stored, searched, retrieved, analysed using SNOMED-CT

  – enriches HKCTT with semantic definitions & relationship hierarchies, enabling semantic & hierarchical analysis of eHR records in future
• the bonus

  – uncovers problem concepts accumulated but hidden inside the vocabulary table, including duplicated, ambiguous & inconsistent concepts
  – provides easy & efficient method for cleansing
• lessons learnt
  – follow standards
  – use tools to speed up
  – seek authority to fix root problems
  – use priority to adjust progress
  – track patterns, review & re-apply to similar tasks
  – hold discussion meetings to align practices
  – seek expert advices and do external verifications
acknowledgement of feeling
psychodynamic interventions recognition of feeling acknowledging anxiety
Thank you!