SNOMED CT
Coping with concept inactivation

Presenter: Dr Jeremy Rogers, IHTSDO Consultant Terminologist
Date: 10th October 2013, IHTSDO Showcase, Washington DC
Outline

• Recap: how code queries are executed
• The ‘inactive content’ problem
• Two-part solution
  – Substitutions table
  – ‘Role Inclusion Closure’ table
Querying in Ye Olde Worlde
(READ2, ICPC, ICD, OPCS…)

• Direct **lexical** comparison of ConceptID

  C1... Other endocrine gland diseases       E       Respiratory tract
  C10.. Diabetes mellitus                     E02     Plastic operations on nose
  C10E4 Unstable type 1 diabetes mellitus     E02.5 Reduction rhinoplasty

• Pros
  – Doesn’t require any external reference table to compute
  – Quick to execute

• Cons
  – Stuck with a monohierarchy (and duplicate codes)
  – Can’t move concepts if initially put in the wrong place
  – Hierarchy gets full: can’t put in the right place!
  – etc
New World Querying
(CTV3, SNOMED...)

• Identifiers are meaningless
  – to fix probs with meaningful IDs
  – but therefore lexical comparison of IDs won’t work

• Hierarchy stored in IS-A table, not in codes:

<table>
<thead>
<tr>
<th>PARENT</th>
<th>CHILD</th>
</tr>
</thead>
<tbody>
<tr>
<td>X40Gd</td>
<td>X40J1</td>
</tr>
<tr>
<td>X40J1</td>
<td>X40J3</td>
</tr>
<tr>
<td>X40J3</td>
<td>C10..</td>
</tr>
<tr>
<td>C10..</td>
<td>X40J4</td>
</tr>
<tr>
<td>X40J4</td>
<td>Xa4g7</td>
</tr>
</tbody>
</table>

• But IS-A table too slow for subsumption check
  – Recursive calls: \texttt{Xa4g7 subtype-of X40Gd} ?
  – Need derivative: ‘Transitive Closure’ table
‘Transitive Closure’ Table?

- IS-A relationship is logically ‘transitive’:
  \[ \forall a, b, c : a \text{ IS-A } b, \ b \text{ IS-A } c \Rightarrow a \text{ IS-A } c \]

- ‘Closure’ table: list of all relationships inferrable from IS-A axioms

<table>
<thead>
<tr>
<th>PARENT</th>
<th>CHILD</th>
<th>SUPERTYPE</th>
<th>SUBTYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>X40Gd</td>
<td>X40J1</td>
<td>X40Gd</td>
<td>X40Gd</td>
</tr>
<tr>
<td>X40J1</td>
<td>X40J3</td>
<td>X40Gd</td>
<td>X40J1</td>
</tr>
<tr>
<td>X40J3</td>
<td>C10..</td>
<td>X40Gd</td>
<td>C10..</td>
</tr>
<tr>
<td>C10..</td>
<td>X40J4</td>
<td>X40Gd</td>
<td>X40J4</td>
</tr>
<tr>
<td>X40J4</td>
<td>Xa4g7</td>
<td>X40Gd</td>
<td>Xa4g7</td>
</tr>
<tr>
<td>X40J1</td>
<td>X40J1</td>
<td>X40J1</td>
<td>C10..</td>
</tr>
<tr>
<td>X40J1</td>
<td>X40J4</td>
<td>X40J1</td>
<td>X40J4</td>
</tr>
<tr>
<td>X40J3</td>
<td>X40J3</td>
<td>X40J3</td>
<td>C10..</td>
</tr>
<tr>
<td>X40J3</td>
<td>X40J4</td>
<td>X40J3</td>
<td>X40J4</td>
</tr>
<tr>
<td>X40J3</td>
<td>Xa4g7</td>
<td>X40J3</td>
<td>Xa4g7</td>
</tr>
<tr>
<td>C10..</td>
<td>C10..</td>
<td>C10..</td>
<td>C10..</td>
</tr>
<tr>
<td>C10..</td>
<td>X40J4</td>
<td>C10..</td>
<td>X40J4</td>
</tr>
<tr>
<td>C10..</td>
<td>Xa4g7</td>
<td>C10..</td>
<td>Xa4g7</td>
</tr>
<tr>
<td>X40J4</td>
<td>X40J4</td>
<td>Xa4g7</td>
<td>Xa4g7</td>
</tr>
<tr>
<td>X40J4</td>
<td>Xa4g7</td>
<td>Xa4g7</td>
<td>Xa4g7</td>
</tr>
</tbody>
</table>

Similar closure tables can be built for any other transitive relationship e.g. partOf, causedBy, follows etc.
Typical Reporting Architecture
(example from CTV3)

- Query Writer
- Endocrine disorder
- Clinical Data
- Query Library
- Transitive Closure Table 8.5M Rows
- Data Entry

NB Some graph databases, or graph-optimised RDBMS servers, include SQL transitivity extensions such as CONNECT BY that mean you don't have to explicitly build a transitive closure table.
Outline

• Recap: how code queries are executed
• The ‘inactive content’ problem
• Two-part solution
  – Substitutions table
  – ‘Role Inclusion Closure’ table
The ‘inactive content’ problem

• Classical SNOMED CT querying - including by means of ‘transitive closure’ tables - ONLY considers IS-A relationships

• In combination with how SNOMED CT represents concept inactivation ‘out of the box’, this is not sufficient
Before October 2011...

Naïve query for RTI... 

...never found these.

IF you only consider the IS-A relationships

Naïve query for RTI...

...never found these.

Now it won't find this either

IF you only consider the IS-A relationships
Summary of Result
Inactive content in your data

195746005 Recurrent chest infection [V2 H06z2]
308130008 Recurrent chest infection [V3 XaBM8]
363660007 Ambiguous concept
404684003 Clinical Finding
275498002 Respiratory Tract Infection
54150009 Upper respiratory tract infection
312118003 Bacterial upper respiratory infection [V3 XaDcC]
50417007 Lower respiratory tract infection
448739000 Recurrent lower respiratory tract infection

TC 2011a  TC 2011b
252744  250791
2421  522
2257  358
2000  2000
0  100
404684003 Clinical Finding
275498002 Respiratory Tract Infection
50417007 Lower respiratory tract infection
308130008 Recurrent chest infection (inactivated Oct 2011)
448739000 Recurrent lower respiratory tract infection (added current Oct 2011)
Inactive content in your Queries
Before October 2011...

- 363662004 Duplicate concept

- 50448004 Fracture of vertebral column
  - 48522003 Spinal cord disease
  - 269078005 Fracture of spine with spinal cord lesion
    - 208082008 Closed fracture of coccyx with spinal cord lesion
    - 274156000 Fracture of cervical spine with cord lesion
    - 274158004 Fracture of lumbar spine with cord lesion
    - 68733100000100 [UK] Fracture of spine with spinal cord lesion NOS
    - 274157009 Fracture of thoracic spine with cord lesion
    - 263179007 Fracture of transverse process of spine with spinal cord lesion
    - 1734006 Fracture of vertebral column with spinal cord injury
    - 208088007 Open fracture of coccyx with spinal cord lesion
    - 312840005 Spinal fracture with cervical cord lesion
    - 312842002 Spinal fracture with lumbar cord lesion
    - 312841009 Spinal fracture with thoracic cord lesion

- 404684003 Clinical Finding

- 48522003 Spinal cord disease

- 312840005 Spinal fracture with cervical cord lesion

- 274156000 Fracture of cervical spine with cord lesion

- 767161000000101 Open spinal fracture with incomplete cervical cord lesion, C1-4 (+ 3 others)
- 208001004 Closed spinal fracture with anterior cervical cord lesion, C1-4 (+15 others)
- 53868003 Closed fracture of C1-C4 level with anterior cord syndrome (+ 28 others)
Inactive content in your Queries
After October 2011…

363662004 Duplicate concept

312840005 Spinal fracture with cervical cord lesion

767161000000101 Open spinal fracture with incomplete cervical cord lesion, C1-4 (+ 3 others)

404684003 Clinical Finding

48522003 Spinal cord disease

274156000 Fracture of cervical spine with cord lesion

10310006 Open fracture of C1-C4 level with incomplete spinal cord lesion (+3 others)

208001004 Closed spinal fracture with anterior cervical cord lesion, C1-4 (+15 others)

53868003 Closed fracture of C1-C4 level with anterior cord syndrome (+ 28 others)
Summary of Result
Inactive content in your Queries

363662004 Duplicate concept

312840005 Spinal fracture with cervical cord lesion

767161000000101 Open spinal fracture with incomplete cervical cord lesion, C1-4 (+3 others)

1200 100 312840005 Spinal fracture with cervical cord lesion (duplicate Oct 2011)

500 445 274156000 Fracture of cervical spine with cord lesion

400 386 125609005 Open fracture of cervical region with spinal cord injury

48522003 Spinal cord disease

274156000 Fracture of cervical spine with cord lesion

208001004 Closed spinal fracture with anterior cervical cord lesion, C1-4 (+15 others)

53868003 Closed fracture of C1-C4 level with anterior cord syndrome (+28 others)

10310006 Open fracture of C1-C4 level with incomplete spinal cord lesion (+3 others)

404684003 Clinical Finding

SAME_AS

TC 2011a TC 2011b
Concept Inactivations so far..


Inactive Concepts
IHTSDO 2002-2013

Inactive Concepts
UK Edition 2012-13

Pending Move
Moved
Duplicate
Outdated
Erroneous
Duplicate
Retired
Concept Inactivations so far..

Concept inactivation so far in...

UMLS Core Problem List

5191 concepts in original 200908 release
64 are now inactive (1.24% of codes)
Account for 0.75% combined ‘frequency’

65599008 Fetal or neonatal effect of oligohydramnios (disorder)
23294000 Sports injury (morphologic abnormality)
37472003 Fluid volume deficit (finding)
192840004 Benign familial tremor (finding)
284480000 Cellulitis of arm (disorder)
60535003 Adenomatous polyp of cervix (disorder)
275918005 Unstable diabetes mellitus (disorder)
199516000 Known or suspected fetal abnormality (disorder)
66215008 Fetal or neonatal effect of polyhydramnios (disorder)
58193001 Diplegic cerebral palsy (disorder)
386630000 Orchiectomy (procedure)
68983007 Fetal or neonatal effects of maternal complication of pregnancy (disorder)
300889000 Swelling of arm (finding)
417162001 Nasolacrimal duct obstruction (disorder)
267821008 Hypertrophic cicatrix (disorder)
76226003 Tattoo (disorder)
410064000 Non-traumatic subdural hematoma (disorder)
73890002 Fetal or neonatal effect of delivery by vacuum extractor (disorder)
89600009 Secondary cardiomyopathy (disorder)
262951009 Traumatic subdural hematoma (disorder)
262954001 Traumatic subarachnoid intracranial hemorrhage (disorder)
201937003 Traumatic arthropathy of the lower leg (disorder)
73009009 Bloodshot eye (finding)
249782009 Bowing of leg (finding)
61628006 Drug withdrawal syndrome in newborn (disorder)
47874006 Sprain of arm (disorder)
281638009 Hepatitis B contact (finding)
41006004 Depression (finding)
371330000 Fatty liver (disorder)
131016008 Increased thyroid stimulating hormone level (finding)
166829003 Serum cholesterol borderline (finding)
78431007 Influenza due to Influenza virus, type A, human (disorder)
416103000 Elevated erythrocyte sedimentation rate (finding)
50047001 Compound dental caries (disorder)
630790000 Adenomatous polyp of cervix (disorder)
64333001 Preinfarction angina (disorder)
Concept inactivation so far in...

Data from one ED department

- 408,831 episodes
  - 1 code per episode reason for attendance
  - 38 months of data Oct-2008 to Dec-2011
- Code now inactive for 1216 episodes (0.3%)
- 81 codes involved (out of 12,069 used)

Top 20 most frequently used inactive codes:

220 episodes of 267040003 Leg swelling (finding)
130 77299006 Olecranon bursitis (disorder)
118 41006004 Depression (finding)
95 238402004 Cellulitis of leg (disorder)
55 300889000 Swelling of arm (finding)
53 425406006 Hematoma of leg (disorder)
45 211832003 Partial thickness burn of arm (disorder)
39 198881004 Pregnancy complications (disorder)
39 285321000000107 Haematoma of leg (disorder)
38 284480000 Cellulitis of arm (disorder)
34 90821003 Complication related to pregnancy (disorder)
32 211828009 Superficial burn of arm (disorder)
29 47874006 Sprain of arm (disorder)
20 267782008 Cellulitis and abscess of leg (disorder)
18 281842005 Fracture tibial plateau (disorder)
18 102550009 Leg cramp (finding)
15 402261004 Chemical burn (disorder)
14 1508000 Intracerebral hemorrhage (disorder)
13 162345005 Blocked ear (finding)
12 37324003 Superficial injury of leg with infection (disorder)
Outline

• Recap: how code queries are executed
• The ‘inactive content’ problem
• Two-part solution
  – Substitutions table
  – ‘Role Inclusion Closure’ table
Two-part Solution

Part 1: Substitutions Table

• Need to compute the blue arrows
Building a Substitutions Table

• Theoretically, simple lookup of
  – RF1 history relations from sct1_relationships
  – RF2 association reference sets

• In practice, not that simple
Building a Substitutions Table

- Not 1:1 (and often 1:0)
- Necessarily recursive if extensions involved
  - 158497008 [D]Other abdominal and pelvic symptoms
  - 170251007 Child 6 month examination NEC (procedure)
- `SAME_AS` is cyclic for ‘limited status’
- Not all 1:\(\infty\) is explicitly flagged
  - 667611000000104 Myalgia and myositis unspecified (disorder)
- Special handling in some jurisdictions
- Unclear semantics for `WAS_A, MAY_BE`
SNOMED CT Historical Relations

Recursive and cyclic

170251007
Child 6 month examination NEC

170263002
Child 6 month examination

560751000000103
Child 6 month examination NEC

147483009
Child 6 month examination NEC

243788004
Child examination

170263002
Child 6 month examination

560751000000103
Child 6 month examination NEC

243788004
Child examination

170263002
Child 6 month examination

560751000000103
Child 6 month examination NEC
Unflagged ambiguity

- 667611000000104: Myalgia and myositis unspecified (disorder)
  - SAME_AS: 203108005: Myalgia and myositis unspecified (disorder)
  - WAS_A: 75047002: Disorder of skeletal muscle (disorder)
    - WAS_A: 373673007: Disorder characterized by pain (disorder)
  - WAS_A: 395WAS_A: WAS_A: WAS_A: WAS_A
Two-part Solution

Part 1: Substitutions Table

- Need to compute the blue arrows
- Fiddly algorithm (though 2 minutes to execute)
- UKTC Product: UK Substitutions Table
Outline

• Recap: how code queries are executed
• The ‘inactive content’ problem
• Two-part solution
  – Substitutions table
  – ‘Role Inclusion Closure’ table
Using a Substitutions Table

One possible architecture..

Existing and new QueryDefs are updated through the substitutions table before being run.

New data items, including those entered via pre-defined templates, are translated at runtime through the substitutions table before being committed to the EPR.

Existing data items are updated through the substitutions table at each new terminology release, as a batch process.

But there is another way.
‘Role Inclusion Closure’ Table
(aka SNOMED Query Table ?)

Merge the transitive closure and substitutions tables into a single table
• Combinatorial logic of IS-A + history roles
  • \( \forall a, b, c : a \text{ IS-A } b, b \text{ SAME-AS } c \Rightarrow a \text{ IS-A } c \)
  • \( \forall a, b, c : a \text{ SAME-AS } b, b \text{ IS-A } c \Rightarrow a \text{ IS-A } c \)
  • \( \forall a, b, c : b \text{ MOVED-FROM } a, b \text{ IS-A } c \Rightarrow a \text{ IS-A } c \)
  • \( \forall a, b, c : a \text{ IS-A } b, b \text{ MOVED-FROM } c \Rightarrow a \text{ IS-A } c \)
  • \( \forall a, b, c : a \text{ IS-A } b, b \text{ MOVED-FROM } c \Rightarrow a \text{ IS-A } c \)
  • \( \forall a, b, c : a \text{ WAS-A } b, b \text{ IS-A } c \Rightarrow a \text{ IS-A } c \)
  • \( \forall a, b, c : a \text{ IS-A } b, b \text{ WAS-A } c \Rightarrow a \text{ IS-A } c \)
  • \( \forall a, b, c : a \text{ REPLACED-BY } b, b \text{ IS-A } c \Rightarrow a \text{ IS-A } c \)
  • \( \forall a, b, c : a \text{ IS-A } b, c \text{ REPLACED-BY } b \Rightarrow a \text{ IS-A } c \)

• \( \forall a, b, c : a \text{ MAY-BE-A } (b \Delta c), b \text{ IS-A } d, c \text{ IS-A } e \Rightarrow a \text{ MAY-BE-A } d, a \text{ MAY-BE-A } e, a \text{ IS-A } (d \Delta e) \) ?

• More than (single) role transitivity; now ‘role inclusion’
• ‘Closure’ table : list of all descendents inferrable from set of IS-A, SAME-AS, WAS-A, MOVED-FROM, REPLACED-BY and MAY-BE-A axioms

NB Probably something of a challenge to achieve the same logical result using the SQL CONNECT BY extension
The role inclusion closure for all substitutes is a MUCH bigger table.

Substitution Table:

IF 191044006 SAME-AS 73211009 THEN everything that’s known about 73211009 is also true of 191044006.

It has the same descendents.
And the same ancestors.
The role inclusion closure for all substitutes is a MUCH bigger table.
Query Table Advantages

• No change to typical supplier architecture
  – Just use a bigger table where TC should be
• Single transparent fix for inactive content
  in data *and* in queries *and* in bound forms
Outline

• Recap: how code queries are executed
• The ‘inactive content’ problem
• Two-part solution
  – Substitutions table
  – ‘Role Inclusion Closure’ table
• Demo
• Caveats
SNOMED Query Table Demo

Open Source MS Access VBA Application

Simulated data:

30,000 ‘patients’
180,000 coded episodes in GP distro
108k Findings/Disorders
54k Procedures/Therapies
19k Others
850 planted inactive codes
Caveats

• Full RI Table is ‘huge’
  – 33M rows vs 8.5M for standard TC table
    • But only 12M if inactivation before 2003 ignored
• Combinatorial semantics require thought
  – SAME-AS and MOVED-FROM same controversial; true identifier equivalence
  – REPLACED-BY also identifier equivalence?
  – WAS-A : same ancestors, but not descendents
  – Old descendents to be retrieved from RF2??
  – MAY-BE-A with…
    • only one substitute ? (43%)
    • > 1 nominated substitute (max = 23)
      – IS-A lowest common subsumer?

169122008
CTV3: 57... Diagnostic nuclear medicine (& various isotope studies)

Parents
this concept IS A Ambiguous concept

Historical
this concept MAY BE A Brain isotope studies
this concept MAY BE A CSF isotope study
this concept MAY BE A Cystographic isotope studies
this concept MAY BE A Nuclear medicine diagnostic procedure
this concept MAY BE A Nuclear medicine imaging procedure
this concept MAY BE A Nuclear medicine procedure
this concept MAY BE A Placenta isotope study
this concept MAY BE A Plasma radioiron turnover rate
this concept MAY BE A Radioisotope function study of liver
this concept MAY BE A Radioisotope joint imaging
this concept MAY BE A Radioisotope scan of bone
this concept MAY BE A Radioisotope scan of lymphatic system
this concept MAY BE A Radioisotope scan of spleen
this concept MAY BE A Radioisotope study of liver
this concept MAY BE A Radionuclide studies in hematology
this concept MAY BE A Radionuclide studies of heart
this concept MAY BE A Radionuclide study of lung
this concept MAY BE A Radionuclide study of thorax
this concept MAY BE A Radionuclide urinary tract study
this concept MAY BE A Renal isotope studies
this concept MAY BE A Skull isotope studies
this concept MAY BE A Thyroid imaging
this concept MAY BE A Vitamin B12 isotope studies
CTV3 (Jan 2002 – Oct 2010)
X76gD Form of thorax
Xa870 Chest deformity
X76gJ Shield-shaped chest
XM01u Pectus carinatum
XM01w Pectus excavatum
XM01x Barrel chest

SNOMED 2002
- Finding of form of thorax (finding)
  - Chest deformity (finding)
    - Shield-shaped chest (finding)

- Acquired deformity of trunk (disorder)
- Disease of thorax (disorder)
- Acquired deformity of chest (disorder)
  - Acquired pectus carinatum (disorder)
  - Acquired pectus excavatum (disorder)
  - Barrel chest (disorder)
  - Flat chest (disorder)

SNOMED 2006
- Deformity (finding)
  - Finding of form of thorax (finding)
  - Chest deformity (finding)
    - On examination - chest deformity (finding)
      - On examination - ankylosing spondylitis chest deformity (disorder)
      - On examination - chest deformity NOS (finding)
      - On examination - kyphoscoliotic chest deformity (disorder)
      - On examination - localized chest deformity (finding)

SNOMED 2009
- Deformity (finding)
  - Finding of region of thorax (finding)
  - Acquired deformity of chest (disorder)
  - Acquired deformity of rib (disorder)
  - Acquired chest deformity, unspecified (disorder)
  - Acquired chest or rib deformity NOS (disorder)
  - Acquired rib deformity, unspecified (disorder)
  - Pectus excavatum (disorder)
  - Scoliosis (disorder)
  - Acquired pectus carinatum (disorder)
  - Acquired pectus excavatum (disorder)
  - barrel chest (disorder)
  - On examination - barrel chest (disorder)
  - Flat chest (disorder)
  - Kyphoscoliosis of thoracic spine (disorder)
  - Kyphosis of thoracic spine (finding)
  - Postural deformity of thoracic spine (finding)
  - Scoliosis of thoracic spine (finding)
  - Deformity of thoracic vertebrae (disorder)
  - Congenital kyphosis (disorder)
  - Kyphosis in skeletal dysplasia (disorder)
  - Kyphoscoliosis of thoracic spine (disorder)
  - On examination - chest deformity (finding)
  - On examination - ankylosing spondylitis chest deformity (disorder)
  - On examination - barrel chest (disorder)
  - On examination - localized chest deformity (finding)
  - On examination - scoliosis (finding)
  - Shield-shaped chest (finding)
A naïve query for 30185001% will return 56 descendents... but miss 36 inactive codes.
What happens if the inactive concept is in the QueryDef?

After October 2011...

363660007
Ambiguous concept

301865001
Chest deformity

448186009
Deformity of thoracic structure

448488004
Deformity of chest wall

162923001
O/E - chest deformity

162929002
Kyphoscoliotic chest deformity (disorder)

162930007
Ankylosing spondylitis chest deformity (disorder)

162931006
Localized chest deformity (finding)

249673007
Shield-shaped chest (finding)

249683006
Bowed clavicle (finding)

298492007
Deformity of thoracic spine (finding)

298498006
Deformity of joint of thoracic spine (finding)

24228002
Barrel chest

88855001
Acquired deformity of chest

360396000
Deformity of breast

7921007
Congenital deformity of clavicle

NEW CODES NOT PREVIOUSLY RETURNED AT ALL

253445003|Common atrioventricular valve chordae too short (disorder)|
253446002|Common atrioventricular valve chordae too long (disorder)|
448497000|Congenital deformity of mitral valve annulus (disorder)|
447924006|Tortuosity of pulmonary artery (disorder)|
448186009|Deformity of thoracic structure (disorder)|
448185008|Acquired deformity of thoracic structure (disorder)|
448488004|Deformity of chest wall (disorder)|
449466007|Acquired deformity of chest wall (disorder)|
298735007|Deformity of rib
Thank You!

jeremy.rogers@hscic.gov.uk