



1 match search >>

ICD32000-00

is a ICD  
is a 84604002|Sigmoid colectomy (procedure)

# Exploiting Fast Classification of SNOMED CT for Query and Integration of Health Data

Michael Lawley  
Principal Research Scientist  
Australian e-Health Research Centre  
June 2008

# Goal

- Better tool support for consistent querying of health data from multiple sources

<u>Surgery</u>
PersonID
DateDiagnosis
DateSurgery
Procedure
Surgeon
ProcedurePerformed
<b>Procedure</b>
Right hemicolectomy
Ext right hemicolectomy
Left hemicolectomy
Sigmoid colectomy
Total colectomy
Subtotal colectomy
Proctocolectomy
High AR
Low AR
Ultralow AR
APR
Hartmanns
Other

<u>Procedure</u>
MRN
EpisodeNo
ProcedureCode
<b>ProcedureCode</b>
<b>Surgical Procedures</b>
32003-00 (Sig colectomy)
32000-00 (Sig colectomy)
32003-01 (Right hemicolectomy)
32000-01 (Right hemicolectomy)
32005-01 (Ext right hemicolectomy)
32004-01 (Ext right hemicolectomy)
32006-00 (Left hemicolectomy)
32006-01 (Left hemicolectomy)
32005-00 (Subtotal colectomy)
32004-00 (Subtotal colectomy)
32012-00 (Total colectomy with ileorectal anastomosis)
32015-00 (Total colectomy with proctocolectomy +/- ileoanal reservoir/ pouch procedure)
32051-00
32051-01
32030-00 (Hartmann's procedure)
32024-00 (High anterior resection)
32025-00 (Low anterior resection)
32026-00 (Low anterior resection)
32028-00 (Ultra low anterior resection)
32099-00 (Other procedure on rectum)
32108-00 (Other procedure on rectum)
32039-00 (Abdomino-perineal excision)
32029-00 (Formation of colonic J pouch)

# Querying

- Need to *consistently* query existing health data
- Data uses controlled vocabularies with unspecified (or special-cased) semantic relationships
  - 32003-00 Sig colectomy with *anastomosis*
  - 32003-01 Right hemicolectomy
  - 32012-00 Total colectomy
- Queries of data need to account for these implicit relationships
  - e.g., *find procedures involving a colectomy*
- Leads to semantics being embedded in the queries
  - bad

# Semantics embedded in queries

- ```
SELECT S.*  
FROM Surgery S  
WHERE S.procedure = '32003-00'  
OR S.procedure = '32003-01'  
OR S.procedure = '32012-00'  
...;
```
- ```
SELECT S.*  
FROM Surgery S, ProcedureCodes C  
WHERE S.procedure = C.code  
AND C.text LIKE '%colectomy%';
```

# Formal Ontology

- Use an ontology to make the relationships explicit
  - Approach taken by SNOMED CT
- Make these relationships available for querying
  
- `SELECT S.*`  
`FROM Surgery S, Ontology O`  
`WHERE O.ancestor = 23968004`  
`AND S.procedure = O.descendant;`

# How to construct the ontology?

- Could start from scratch, but
  - may need to explicitly encode lots of additional relationships
  - requires good modelling expertise (as well as domain knowledge)
- May be easier to build on (extend) an existing base
  - SNOMED CT
    - ~400,000 concepts
    - ~1,500,000 relationships
    - provides a path to future integration, (esp. in Australian context)
  - size can also be a problem
  - contains errors
  - Need to be careful that extension doesn't corrupt base
    - conservative extension

# Issue of scale

- **Benefits**

- breadth
- rich relationship structure
- established concept model

- **Costs**

- sheer size is hard to manage
- standard tools struggle or can't cope
  
- when classification takes ~45min you don't do it often
- but...

# New algorithm

- Polynomial-time algorithm published by Baader, Lutz, and Suntisrivaraporn
  - "CEL"
  - LISP, Linux only
  - ~30min
- Our implementation
  - "snorocket"
  - Java, tested on Windows, OS X, Linux (RedHat, ubuntu), Solaris
  - ~1min
- `/usr/bin/time`  
`/System/Library/Frameworks/JavaVM.framework/Versions/1.6/Home/bin/java -Xmx2G`  
`-server -jar target/snorocket-1.1-SNAPSHOT-jar-with-dependencies.jar --krssFile`  
`../ontologies/snomedct-stated-2007-07-30.krss`  
`66.27 real 65.39 user 2.16 sys`
- unoptimised incremental version of algorithm gives sub-second results



# Feasibility

- Essential approach has been trialed
  - see "*Experiences Mapping a Legacy Interface Terminology to SNOMED CT*", Geraldine Wade and S. Trent Rosenbloom
  - 2002 terms mapped to combination of single and post-coordinated concepts
  - about 75% were equivalencies (20% of these were to single concepts)
- General issues identified
  - one term may be used for multiple concepts
  - many relationships may not be explicitly represented/have a corresponding term
  - complete concept may be composite and representationally split across multiple columns (and tables)

# Simple case

- examine existing terms
- look at their use in the data
  - use is the determinant for semantics
- find corresponding SNOMED CT concepts
  - or model as post-coordinated expressions
- are they equivalent or primitive?
- classify, look for problems
  - snorocket supports *bottom*,  $\perp$ , and thus disjointness constraints
  - found problems with Skeletal Dysplasia terminology in REAMS
    - used "eye" for both *optic nerve* and *bony orbital structure*
- when done, export the classified extension

# Complex Information Model

Patient	Date	Status	...	Procedure	Laterality
...	...	...	...	...	...

Patient	Date	Status	...	Procedure	Laterality	Concept
...	...	...	...	...	...	...

- with new concepts defined as
  - $\exists$  associatedProcedure. P  $\sqcap$
  - $\exists$  laterality. L  $\sqcap$
  - $\exists$  procedureContext. S
- this set of new concepts forms another extension
- when classified, subsumption queries are still a simple join

# Future

- Evaluation projects
  - current
    - ANZICS Intensive Care Unit terms
  - future
    - Community Health
    - Patient Safety
- Tool support
  - editor specifically for building such extensions
    - real-time feedback via fast incremental classification
    - build in the concept model
    - pinpointing to assist debugging?
  - querying integrated with (extended) ontology
- Problems
  - Negation
  - Free text (NLP)

# Demo?

# Questions?

snorocket evaluation licence  
email [michael.lawley@csiro.au](mailto:michael.lawley@csiro.au)