Retrieval of Clinical Information using the SNOMED CT Relationship Network - Possibilities and Pitfalls

Ole Terkelsen MD, PhD, BSCS
SNOMED CT Implementation Showcase 2014
Use of (clinical) data?

- **Observe**: Some of the content in the following presentation might seem overtly oversimplified or even naive*)
- **Warning**: it also contains a few reflections over human verbal logic and set logic
- ... and I will try to say as little as possible about information models

*) showing a lack of experience, wisdom, or judgment
Current relevance in Denmark?

• Two (of five) regions are going to implement Epic’s EHR system (at least partly) …

• … and the current decision is *somehow* to use SNOMED CT in that framework

• The population in The Capital Region and Region Zealand is 2.2 mill. (of 5.5 mill. in Denmark)
What is SNOMED CT?

• ... actually – and what are the advantages?

• SNOMED CT possesses two major important features (as you all know)
  – Terms that represent the lexical description of each concept
  – The relationship network that represent the logical characteristics of each of the concepts
    • i.e. how the concepts (might) be related
What is SNOMED CT?

• The relationship network is constructed as a **unidirectional graph** (and I’ll come back to that)
  – This is also accountable for some limitations
  – It is among other things difficult to handle negations
  – e.g. Not pregnant (finding)
    • is a: Finding related to pregnancy (finding) (!)
    • many negations in SNOMED has, however, been handled
  – negations should be handled in the information model!

• If you want to use SNOMED for logic (!)- you will have to use pure **set logic:**  \( \forall \exists \neg \subset \)
What is SNOMED CT?

• The concepts is represented by at least two terms
  – fully specified
  – prefered
  – (synonym(s))
• often constructed as motivated terms
  – (mini definitions)
• not more about terms . . . but remember that language can trick you!
• not more - except some examples . . .
What is SNOMED CT?

• Provides care *without* prejudicial behaviour (procedure)
• Hypertension *without* albuminuria AND *without* oedema in the obstetric context
• Nutritional edema *without* dyspigmentation of skin AND/OR hair (disorder)
• Severe manic bipolar I disorder *without* psychotic features (disorder)
• (this is also SNOMED ...
SNOMED CT - relationships

just mentioning . . .

<table>
<thead>
<tr>
<th>Appendectomy</th>
<th>Bacterial meningitis</th>
</tr>
</thead>
<tbody>
<tr>
<td>is-a Operation on appendix</td>
<td>is-a Infective meningitis</td>
</tr>
<tr>
<td>is-a Partiel excision of large intestine</td>
<td>is-a Bacterial infection of central nervous system</td>
</tr>
<tr>
<td>procedure-site  Appendix structure</td>
<td>finding-site Meninges structure</td>
</tr>
<tr>
<td>method Excision - Action</td>
<td>associated-morphology Inflammation</td>
</tr>
<tr>
<td></td>
<td>pathological process Infectious disease</td>
</tr>
<tr>
<td></td>
<td>Causative-agent Bacterium</td>
</tr>
</tbody>
</table>

(fully defined)

The use of attribute relations follow specific rules (description logics)
The architecture of a concept based terminology

A concept based terminology

Disorder

Tumour

Lung disease

Cancer

Benigne tumor in throat

Throat cancer

Lung cancer

Throat disease

Inflammation

Pneumonia

Tonsillitis
The architecture of SNOMED CT!

A polyhierarchal terminology

One concept can have more than one supertype
The architecture of SNOMED CT!

The "is a" relations always points "upwards"
The architecture of SNOMED CT!

If the "is a" relation is used in "reverse" you can aggregate information (count) from any point (concept) downwards.
The architecture of SNOMED CT!

If the "is a" relation is used in "reverse" you can aggregate information (count) from any point (concept) downwards.
The architecture of SNOMED CT!

If the "is a" relation is used in "reverse" you can aggregate information (count) from any point (concept) downwards.

Count "tumours"
The architecture of SNOMED CT!

If the "is a" relation is used in "reverse" you can aggregate information (count) from any point (concept) downwards.

Count "lung diseases"
... decision support, audit, epidemiology...

• SNOMED CT (benefits) - represents logical semantic relationships between concepts. This allows consistent retrieval of clinical information for a wide range of purposes, including decision support, audit, epidemiology, research, service management, billing and statutory reporting. 

• Does this comply to e.g.:
  – Top level aggregations?
  – Medium level usage?
  – The detailed clinical (?) level?
... decision support, audit, epidemiology...

• Top level: With some examples

• What "medium" level?
  – might correspond to e.g. clinical databases?

• (Detailed clinical (?) level out of scope)
“Top level” examples from data in the National Danish Patient Register (NPR)

• The following slides shows a few examples of aggregation of coded information using SNOMED’s relationship network

• The data originate from the NPR that comprises information about outpatients and admitted patients in Denmark (11 mil. entries/year)

• The ICD-10-DK codes are mapped “in reverse” to SNOMED ...
  – The “aggregation points” are SNOMED CT concepts shown in *italics*
Data from NPR – ”aggregated” with SNOMED CT

SNOMED CT concept in italics

All types of pneumonia and viral pneumonia
all admitted patients, distributed by age
Data from NPR – "aggregated" with SNOMED CT

SNOMED CT concept in italics

All types of complications to diabetes
Outpatients and admitted patients, distributed by age
Data from NPR – ”aggregated” with SNOMED CT

*SNOMED CT concepts in italics*

**Hereditary and congenital disease**

Admitted patients, distributed by age
Top level ...

• Quite advanced “top level” aggregations seems to work ...
• ... and could replace the national statistical reporting to WHO ...
• ... who (!) only receives calculated statistics from nations or Classification Centers ...
• ... based on ICD – used as a tool ...
• ... so – why map from SNOMED to ICD when you can use SNOMED “directly” – as a tool ??
What "medium" level?

- Are SNOMED “useful” in e.g. clinical quality databases?
- Objective: We wanted to extend the methods mentioned above for analysis of data from individual databases and maybe across databases (contains lots of data !) – Could we take it to a lower level?
- Denmark: 60+ nationwide clinical databases (including cancer) some for 10+ years
- many (most) of the indicators are the same in other countries
- SNOMED could be tested in cancer databases because of their importance and because SNOMED could be expected to be “perfect” in the relevant areas ...
What "medium" level?

• We started out with colon cancer and gynaecological cancers ...
• In the questionnaire for colon cancer the first entry was: is the tumour situated in colon or rectum
• The conceptual intent is to discriminate between
  • Primary adenocarcinoma of colon
  • ... and ...
  • Primary adenocarcinoma of rectum
What "medium" level?

• We found (in 2009) the following in SNOMED:
  • Adenocarcinoma of large intestine (disorder)
    – Adenocarcinoma of appendix (disorder)
    – Adenocarcinoma of cecum (disorder)
    – Adenocarcinoma of sigmoid colon (disorder)
    – Adenocarcinoma of rectum (disorder)
      • Adenocarcinoma of rectosigmoid junction (disorder)
  • But not colon ! and no “primary”
What "medium" level?

- Primary adenocarcinoma of colon is one of the most common cancers and a major "killer."

**Subdivisions**
- Cecum
- Appendix
- Colon
- Rectum
- Anal canal

![Diagram of the colon and rectum](image)
What "medium" level?

- Pragmatism often wins and we could live without “primary” – but not without “colon” ...
- ... and we found comparable “flaws” for gynaecological cancers ...
- ... and approximately (only) 1/3 of the concepts we needed in the databases
- I even modelled the colon cancer area in Protégé Owl – and had the intend to submit it to NRC – but other priorities came up
- and - frankly - it was easier to continue business as usual ...
What "medium" level?

• ... then the following year (SNOMED evolves) the following turned up:

• Adenocarcinoma of large intestine (disorder)
  – Adenocarcinoma of appendix (disorder)
  – Adenocarcinoma of cecum (disorder)
  – **Primary** adenocarcinoma of *colon* (disorder)
  – Adenocarcinoma of sigmoid colon (disorder)
  – Adenocarcinoma of rectum (disorder)
    • Adenocarcinoma of rectosigmoid junction (disorder)

• Suddenly a “primary” and “colon” ...

• ... but not “primary” for rectum (?)
What "medium" level?

- Out of curiosity I then looked at the “associated morphology” for the adenocarcinomas in the mentioned concepts
- expecting: Adenocarcinoma, no subtype (morphologic abnormality)
- with the legacy code: M-81403
- ordinarily used by surgical pathologists
What "medium" level?

• ... but found three concepts (of 140 adenocarcinomas):
  – Adenocarcinoma, no subtype (morphologic abnormality) = M-81403
  – Malignant adenomatous neoplasm - category (morphologic abnormality)
  – Primary malignant adenomatous neoplasm (morphologic abnormality)

• which just ads to the inconsistency if you e.g. want to use the supporting axes for reasoning logic decision support etc. ...

• The common adenocarcinoma in colon and rectum looks like this: and only one concept is needed
What "medium" level?

• The “Adenocarcinoma of large intestine (disorder)” is just an example – but it is quite common to see comparable issues in other “important” areas of SNOMED ...

• Why talk about the mentioned “flaws” ? because data-entry is very tightly connected to retrieval and analysis !

• You will have to be pragmatic to a certain degree (as when you use e.g. ICD)
And in the end ...  The Beatles, Abbey Road

• As we have heard in this conference and before: The possibilities are there ...  
• ... and there are the usual pitfalls in the hierarchies and texts also known from other classifications ...  
• ... and some additional related to the complexity of SNOMED  
• But ...  
• ... the extent of gaps (and flaws) in the content including relationships imposes a great deal of work before SNOMED becomes really useful and beneficial and accomplish value that cannot be achieved easier otherwise ...
Getting Better - all the time ...  The Beatles, Sgt Pepper

• Another possibility: Ask for the right content – you will have to make it yourself ...
• ... and clinical quality databases is a good place to start ...
• ... keep it simple ...
• ... and remember: if you want to use SNOMED for aggregation, reasoning, decision support, research etc. use set logic and avoid “concepts” with negations, AND/OR etc. and use e.g. lexical exclusions etc. with care ...