Report on implementing and extending the SNOMED CT Diagramming Guideline using Graphviz.

Dr. Edward Cheetham | NHS Digital, UK

Introduction

The SNOMED CT Diagramming guideline [1] ('the guideline') *"...defines a recommended form for diagrams representing SNOMED CT concepts..."*. Such diagrams are an established feature of SNOMED CT browsers, publications and presentations. The guideline currently includes diagram element templates for a number of desktop applications.

Graphviz [2] is open-source graph visualization software. It is used extensively in plugins for many development and analysis environments, and bindings to multiple languages have been produced [3].

SNOMED CT users may wish to take advantage of Graphviz's features, and its use to generate graphs of ancestor and subtype relations is relatively straightforward. Using the same technology to replicate individual concept definition diagram elements and their layout is more challenging, however doing so enables a single diagramming technology to support the production of more complex yet consistent views of SNOMED CT data.

While Graphviz can specify the diagram elements as well as their combination and layout rules, all but the simplest ancestor and subtype diagrams can require considerable data pre-processing (e.g. [5]).

This ePoster describes how the individual elements of the guideline can be represented using Graphviz (to complement the current guideline templates) and shows how an approximation to the concept definition layout can be achieved. Basic element features are then reused in several Graphviz-generated diagram variants. It is intended that the examples presented will assist others wishing to apply Graphviz to SNOMED CT, serve as an inspiration for diagram variant ideas, and be considered by the SNOMED CT community for future versions of the guideline if felt worthwhile.

Methods

Concept definition diagram elements

Graphviz layout programs take, as input, instructions using the 'dot' syntax [4]. This syntax is used to define the properties of elements and how they are combined.

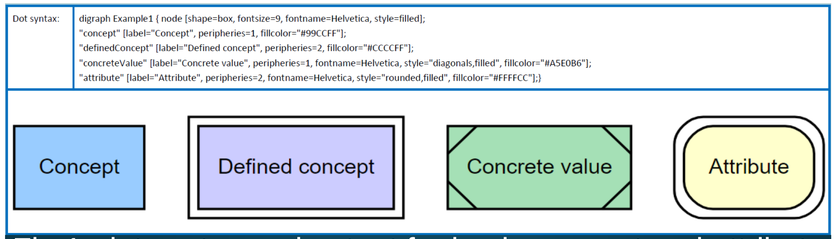

Fig 1: dot syntax and output for basic concept and attribute classes

Figures 1 and 2 show how the individual elements specified in section 4 of the guideline can be represented using the dot syntax and subsequently rendered.

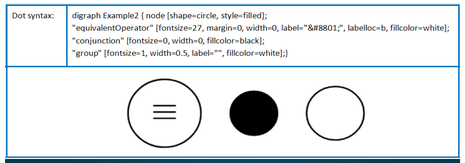

Fig 2: dot syntax and output for relational operator, conjunction and group elements.

Combination and layout

Some Graphviz layout algorithms allow absolute specification of position, however the work presented in this ePoster uses automatic graph layout features (mostly using the ‘dot’ layout engine, but also ‘fdp’ and ‘patchwork’ for specific variants).

Automatic layouts risk deviation from the guideline form, but benefit from allowing experimentation with more complex definition diagrams, combination of concept definitions and related graphs, and the application of additional Graphviz features to SNOMED CT data.

snomedexpo.org

SNOMED CT EXPO 2022

Sept 29-30, 2022 × Lisbon, Portugal

ePosters sponsored by: tpp

Report on implementing and extending the SNOMED CT Diagramming Guideline using Graphviz.

Dr. Edward Cheetham | NHS Digital, UK

Methods

Dot syntax:

```
digraph Example3 {rankdir=LR; node [shape=box, fontsize=9, fontname=Helvetica, style=filled];
"definedConcept"->"equivalentOperator";
"equivalentOperator"->"conjunction1";
"conjunction1"->"group";
"group"->"conjunction2";
"conjunction2"->"attribute";
"attribute"->"concreteValue";
"childConcept"->"parentConcept" [arrowhead=onormal];
"childConcept" [label="Child concept", peripheries=1, fillcolor="#99CCFF"]; "parentConcept" [label="Parent concept", peripheries=1, fillcolor="#99CCFF"];
"definedConcept" [label="Defined concept", peripheries=2, fillcolor="#CCCCFF"];
"concreteValue" [label="Concrete value", peripheries=1, fontname=Helvetica, style="diagonals,filled", fillcolor="#A5E0B6"];
"attribute" [label="Attribute", peripheries=2, fontname=Helvetica, style="rounded,filled", fillcolor="#FFFFCC"];
"equivalentOperator" [shape=circle, fontsize=27, margin=0, width=0, label="&#8801;", labelloc=b, fillcolor=white];
"conjunction1" [shape=circle, fontsize=0, width=0, fillcolor=black]; "conjunction2" [shape=circle, fontsize=0, width=0, fillcolor=black];
"group" [shape=circle, fontsize=1, width=0.5, label="", fillcolor=white];}
```

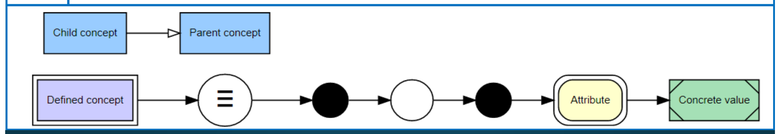

Figure 3: dot syntax and output for basic edges joining diagram elements (edge rows in **bold**)

In addition to the **nodes** shown in figures 1 and 2, Graphviz diagrams also represent **edges**. Figure 3 shows examples of edges relevant to basic concept diagrams as specified in the guideline.

Diagram variants

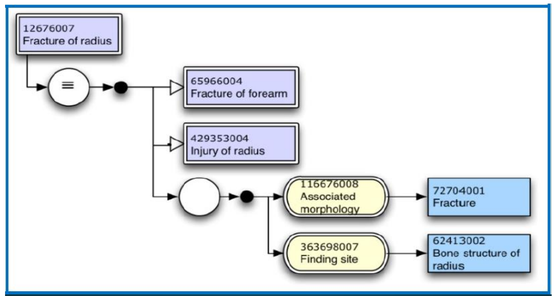

Figure 4: Sample diagram as shown in the SNOMED CT Diagramming guideline

Concept definition diagrams

Figure 4 shows an example concept definition diagram from the guideline, using several of the node and edge elements already introduced. Figure 5 shows a sequence of Graphviz diagrams that attempt to emulate the layout seen in Figure 4, introducing features that progressively improve the similarities.

Diagram ❶ includes the ***splines=ortho*** and ***rankdir=LR*** attributes. These produce right angled (rather than gently curved) edges, and set the edges to flow from left to right.

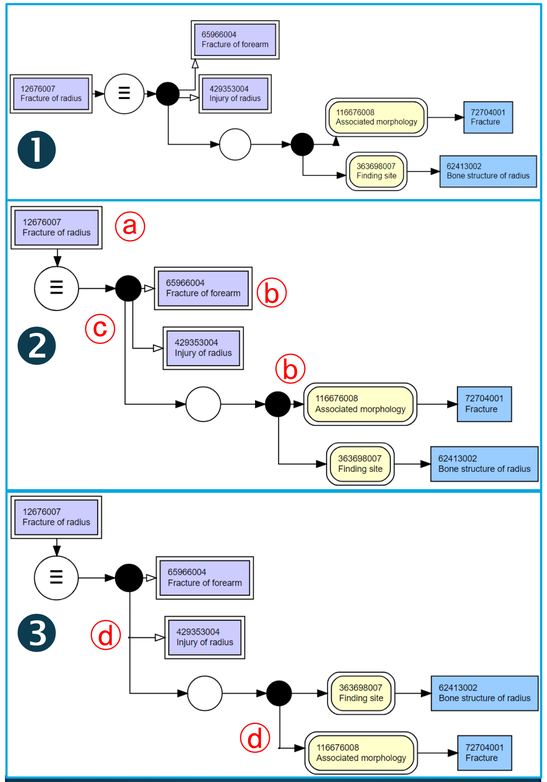

Fig 5: Graphviz-based concept definition variants (details in text)

Diagram ❶ creates a clear layout, but there is scope for improvement. In Diagram ❷ we see the focus concept now sitting above the relational operator (by adding ***constraint=false*** to this relationship ⓐ) and relationship targets positioned in-line and then below each conjunction node (by adding a ***weight=50*** entry to each 'top' relationship ⓑ). It is noticeable (at ⓒ in Diagram ❷) that multiple edges passing between conjunctions and their targets run parallel. This becomes less acceptable as edge numbers increase (e.g. multiple parents, multiple grouped roles) but can be addressed as shown in Diagram ❸ (at ⓓ). Multiple edges are merged at build time, and joined through intermediate nodes. These are represented using tiny ***shape=circle*** or ***shape=point*** nodes, aligned using ***{rank=same; ...}*** constraints, and minor drifts in layout managed with invisible edges (at the bottoms of groups) and by explicitly naming port positions by compass point assignments (esp. ***':sw'***).

Diagram ❶ is less satisfactory as a guideline-conformant concept definition diagram, however the approach used is topologically comparable and is well-suited for more complex diagrams involving multiple focus concepts and nested definitions (such as Figure 6, diagram❷ and Figure 7).

Note: space limits prevent inclusion of full dot files, but complete examples corresponding to those in the ePoster can be found in the Gist at reference [6]

snomedexpo.org

SNOMED CT EXPO 2022
Sept 29-30, 2022 ✕ Lisbon, Portugal

ePosters sponsored by: tpp

Report on implementing and extending the SNOMED CT Diagramming Guideline using Graphviz.

Dr. Edward Cheetham | NHS Digital, UK

Diagram variants

Basic element features (such as labels, colour and shape) can be reused in diagram variants and extensions optimised for other purposes, as in the following examples:

Extended and multi-focus definitions: Many concept definitions include values that are themselves richly modelled, often including deeply-nested definitions. Graphviz syntax makes it easy to generate diagrams that retain consistency with single focus definition diagrams, but also show nested definition expansions or definitions of concept sets. Figure 6 shows the concept definition of 416558007 |Exophthalmos due to thyroid eye disease| (❶) and one option for its expansion (❷). The latter include definition details of referenced values (the original focus and recursively referenced values are in pink). The expanded diagram also merges values where referenced by multiple attributes and shows subgraph relationships between values. Figure 7 is a multi-definition diagram for a set of increasingly specific ventricular septal defect procedures (highlighted in pink).

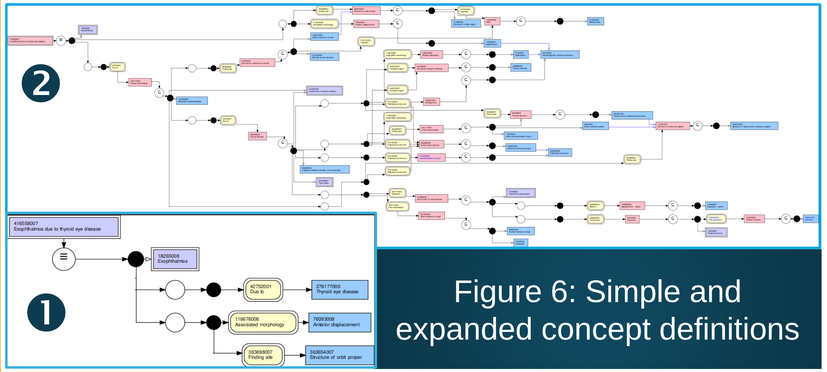

Figure 6: Simple and expanded concept definitions

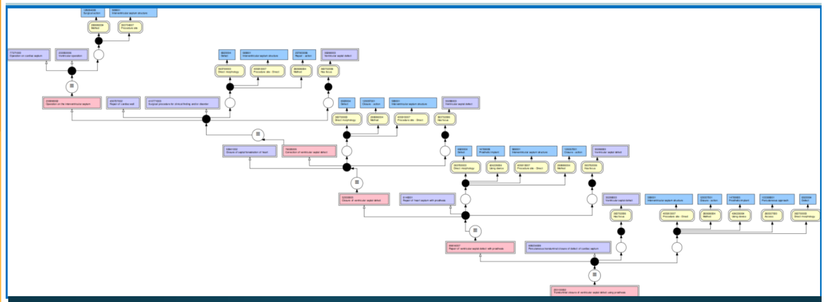

Figure 7: Multi-focus concept definitions

The diagram uses a ‘bottom-top’ (***rankdir=BT***) flow for each definition (indicating increasing specificity to the bottom) and unlike Figure 6 diagram ❷, shared values are *not* merged.

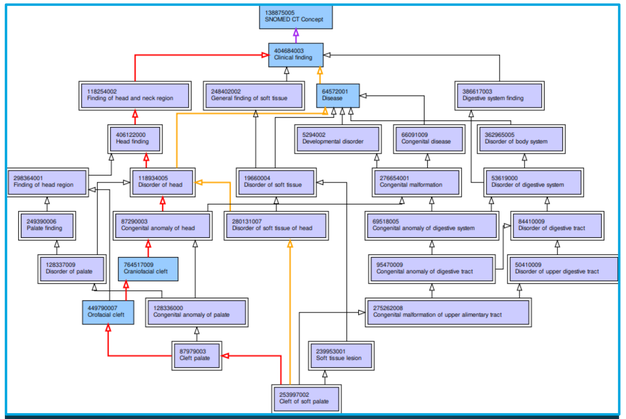

Figure 8: Simple ancestor set

Ancestor, descendant and structured sets: Graphviz is designed to draw graphs, and is therefore well suited to draw ancestor, descendant and (given suitable pre-processing) structured display of arbitrary set members. Figure 8 shows the ancestry of 253997002 | Cleft of soft palate |. Defined and primitive concept conventions as well as subtype arrows from the guideline are used, and in this example sample longest (red) and shortest (orange) paths to root are also shown.

Historical associations: Inbound and outgoing historical associations can be used to augment ancestor and descendant diagrams. Figure 9 augments the ancestry of 253997002 | Cleft of soft palate | with incoming historical 'relationships’ related to the focus. These are distinguished by colour and labelled using corresponding association reference set metadata. As well as incoming historical associations from inactive concepts (here POSSIBLY EQUIVALENT TO, SAME AS and WAS A), 'REFERS TO CONCEPT' ‘relationships’ resulting from description reassignment can be shown between active concepts.

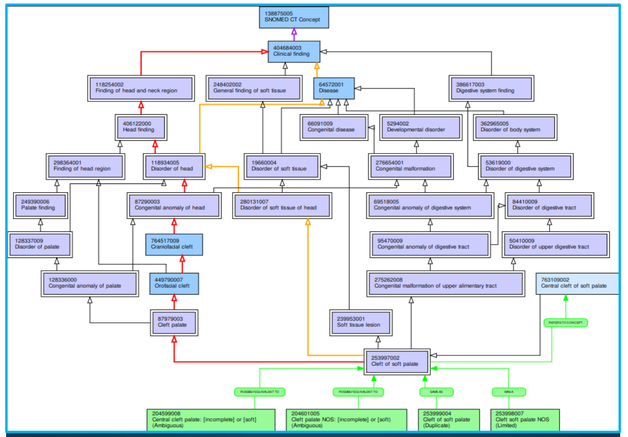

Figure 9: Ancestor set and focus-related historical associations

snomedexpo.org
SNOMED CT EXPO 2022
Sept 29-30, 2022 ✕ Lisbon, Portugal
ePosters sponsored by: tpp

Report on implementing and extending the SNOMED CT Diagramming Guideline using Graphviz.

Dr. Edward Cheetham | NHS Digital, UK

Diagram variants

Treemap descendants: Using the Graphviz 'patchwork' layout a treemap view of each concept's immediate children can be generated. This complements a hierarchical or browser view by rapidly conveying each child's descendant count. Figure 10 shows such a Treemap-style child view for concept 276654001 |Congenital malformation|. Guideline-conformant colours help distinguish primitive and sufficiently-defined children. Descendant (and child) numbers are also shown.

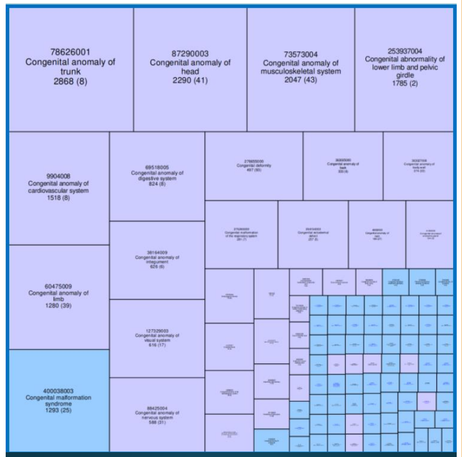

Figure 10: Treemap descendants

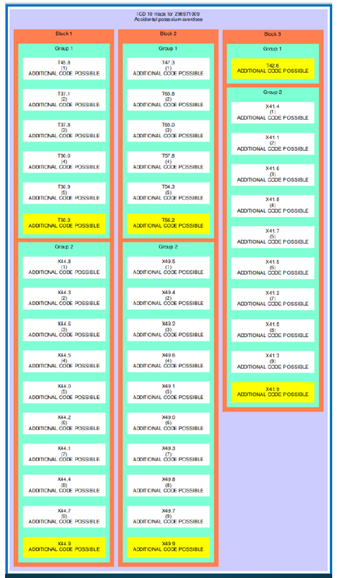

Figure 11: UK Complex cross-maps

Complex cross-maps: UK classification cross-maps use a combination of blocks, groups and target code choices to combine and rank candidate map targets. Map data is often difficult to read if presented as a sequence of table rows, however it is possible, using the Graphviz 'subgraph' and 'cluster' mechanisms, to produce two-dimensional nested diagrams that neatly summarise the map details. Figure 11 shows the UK ICD 10 cross maps for 296971009 |Accidental potassium over-dose|. The layout rapidly conveys the number of cross map ‘block’ choices available, how many ‘groups’ are required within each block, and what the ICD 10 target codes may be (including defaults in yellow) within each group. Echoing the guideline colour convention, the background of Figure 11 indicates that 296971009 is fully-defined.

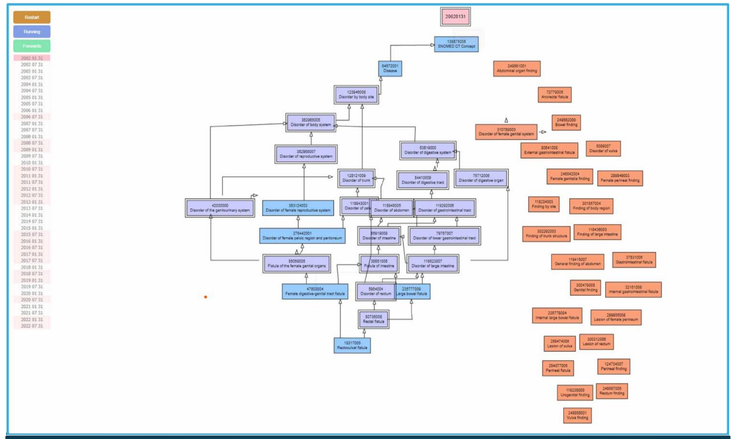

Figure 12: Animated ancestor evolution

D3-enabled animation: It is possible to combine Graphviz’s layout capabilities with the animation and transition features of D3.js [7]. Figure 12 shows such an animation, revealing the nature and extent of change to a concept’s ancestry over time (here 18317005 | Rectovulval fistula | at 6 month intervals between 2002 and 2022). An interactive version of this animation can be found at [8].

Conclusions

This ePoster showcases the application of Graphviz to SNOMED CT data. It shows how automatic layouts can be tailored to reproduce guideline-conformant concept definition diagrams, and how these diagrams can be extended and combined, introducing additional (guideline-consistent) diagram types to reveal further details of SNOMED CT definitions, structure and evolution. It is hoped that the examples provided will assist members of the SNOMED CT community to understand and use Graphviz, and if felt valuable, to consider features of the diagram variants for incorporation into future versions of the guideline.

References

[1] https://confluence.ihtsdotools.org/display/DOCDIAG/Diagramming+Guideline
[2] https://graphviz.org/
[3] https://graphviz.org/resources/
[4] https://graphviz.org/doc/info/lang.html
[5] https://bitbucket.org/cheethame2017/sct2/src/master/
[6] https://gist.github.com/edcheet/3c910debf741cd8c1f0cfc828c00f60e
[7] https://d3js.org/
[8] https://bl.ocks.org/edcheet/6b84ba58b86bca24994447a0259f7fdf

snomedexpo.org

SNOMED CT EXPO 2022
Sept 29-30, 2022 × Lisbon, Portugal

ePosters sponsored by: tpp