

Visualising OMOP concept relationships with omopcept

Andy South

University College London Hospitals, UK

Background

The OMOP standardised vocabularies are a hugely valuable resource comprising more than 10 million concepts, from over 130 vocabularies, and more than 28 million valid relationships¹. This is quite a large number to get ones head around. The sheer number of concepts and relationships presents both challenges and opportunities.

Data visualisation applied to health data has a history extending back to Florence Nightingale, John Snow and others. More recently there is plenty of research around the effectiveness of data visualisation and guidelines on good practice^{2,3}. Data visualisation offers a means of communicating concept relationships to a range of audiences including clinicians, researchers, patients, data engineers and OMOP vocabulary designers.

As I started to look into visualising OMOP concept relationships I realised that an ability to subset concepts and relationships effectively is key to the creation of intelligible visualisations. To this end I have created an R package `omopcept`⁴ for accessing, querying and visualising the OHDSI/OMOP standardised vocabularies.

Methods

`omopcept`⁴ is an open-source R package, hosted on Github,. It is tidyverse compatible and designed to be used in dplyr pipelines and doesn't require Java. The package stores standardised vocabularies locally (efficiently as parquet files) so it can work offline and to maintain speed. There are options to process vocabulary files downloaded from Athena. `omopcept` is designed to be user friendly and quick to operate. Function names are concise and all lower case so they can be typed quickly and mostly start with `omop_` to work well with RStudio code completion.

`omopcept` uses the R package `arrow` so that it doesn't have to read all of the concepts into R's memory (which would be slow). Using `arrow` a connection is opened to a parquet file, initially querying is done on the connection, then only those results that satisfy the query need to be read into memory. There are functions to search concepts by names or codes and to query the relations, ancestors and descendants tables.

Visualisation options are offered through the `omop_graph()` function that accepts tables of relations, ancestors or descendants.

Results

omopcept can be used to produce visualisations of hundreds or even thousands of concepts on large canvases or display screens. I will show visualisations with hundreds of concepts in my talk.

Equally, visualisations of a few tens of concepts and their relationships can be produced. Figure 1 shows the result of recursively searching from a single SNOMED concept to relations that are either 'Is a' or 'Subsumes'. The resultant visualisation shows how the SNOMED concepts 'Accident caused by a sardine tin lid' & 'Accident caused by a corned beef tin lid' are descendants of 'Accident caused by a tin lid' which, in turn, is a descendant of 'Accident caused by a sharp-edged object', itself a descendant of 'Accident caused by nonpowered hand tools and implements'.

Arguments to `omop_graph()` allow the colours and sizes of nodes and text to be set to represent attribute values Figure 2. shows the default option of colouring nodes and text by domain to show the immediate relations of the concept 'Multiple myeloma'.

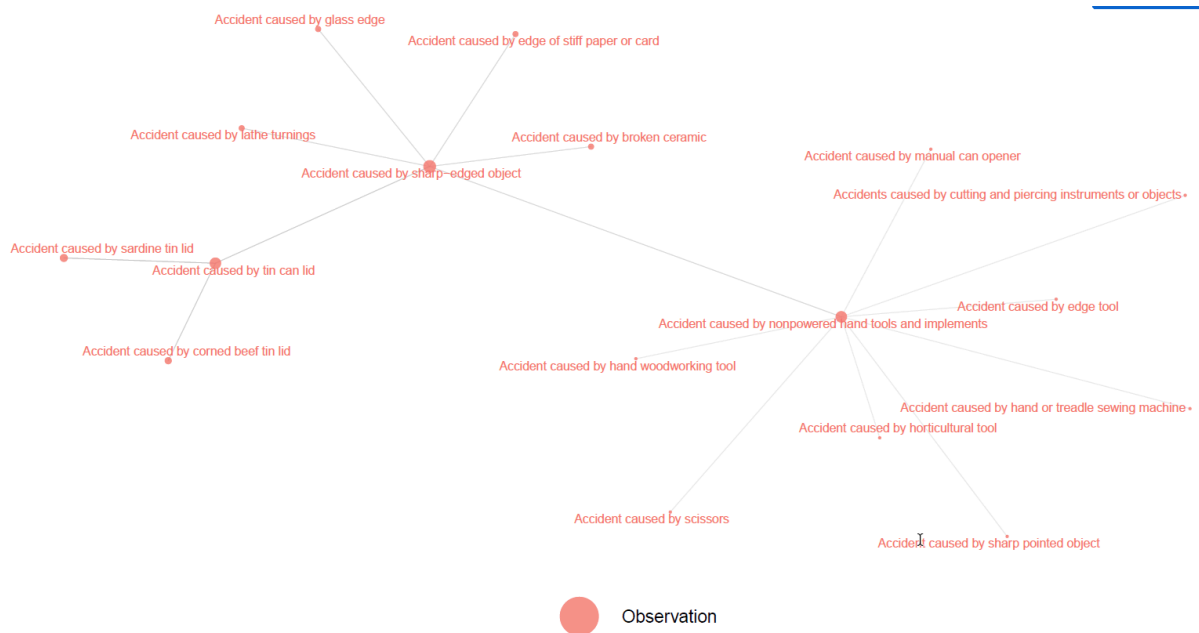


Figure 1. Example of a small network produced by `omop_graph()` within the `omopcept` package. Arrived at by first querying the relations of a single SNOMED concept. Using code `omop_relations_recursive(4055050L, r_ids = c('Is a', 'Subsumes'), num_recurse = 4) |> omop_graph(plot=FALSE)`



Figure 2. Example of a slightly larger network produced by `omop_graph()` within the `omopcept` package. Arrived at by first querying the relations of a single SNOMED concept. Using code

```
mms <- omop_names("multiple myeloma", exact=TRUE, d="Condition", v="SNOMED", s="S")
```

```
mmsr1 <- omop_relations(mms$concept_id)
```

```
omop_graph(mmsr1, width=40, height=30, legendtxtsize=10, legendcm=1)
```

Conclusion

omopcept is already a powerful tool for visualising concept relationships. I look forward to seeing how people use it and to adding new functionality based on user feedback.

References

1. Christian Reich, Anna Ostropelets, Patrick Ryan, Peter Rijnbeek, Martijn Schuemie, Alexander Davydov, Dmitry Dymshyts, George Hripcsak, OHDSI Standardized Vocabularies—a large-scale centralized reference ontology for international data harmonization, *Journal of the American Medical Informatics Association*, Volume 31, Issue 3, March 2024, Pages 583–590, <https://doi.org/10.1093/jamia/ocad247>
2. Alberto Cairo 2016 *The Truthful Art: Data, Charts, and Maps for Communication*. 400 pp
3. Tamara Munzner 2014. *Visualization Analysis and Design*. A K Peters Visualization Series, CRC Press, 404 pp
4. omopconcept : an R package to access OMOP conCEPTs (no cons!) and flexible tidyverse compatible R functions for querying and visualising. Accessed June 21, 2024 <https://github.com/andysouth/omopconcept>