Solubility Data in Bioclipse #4: Finding ChEBI IDs (Again, but better)

Egon Willighagen ⬤

Citation

Keywords

Sparql, Chebi, Bioclipse

Abstract

Those who carefully analyzed the second SPARQL query in Solubility Data in Bioclipse #3: Finding ChEBI IDs will have noticed the use of owl:sameAs.

Copyright

chem-bla-ics

Those who carefully analyzed the second SPARQL query in Solubility Data in Bioclipse #3: Finding ChEBI IDs will have noticed the use of *owl:sameAs*. Those who did not, here's the query again:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX ons: <http://spreadsheet.google.com/plwwufp30hfq0udnEmRD1aQ/onto#>
PREFIX rdfonm: <http://rdf.openmolecules.net/#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>

SELECT DISTINCT ?title ?chebi WHERE {
  ?solvent a ons:Solvent .
  ?solvent dc:title ?title .
  ?solvent owl:sameAs ?same .
  ?same rdfonm:chebiid ?chebi
}
```

This syntax is a bit clumsy, considering we said *?solvent* and *?same* are the same thing. Fortunately, there are tools that do take this into account. One such tool for Jena (which I use in Bioclipse) is Pellet. I just commited code for Bioclipse to use Pellet, which simplifies the above query to:

```
PREFIX ons: <http://spreadsheet.google.com/plwwufp30hfq0udnEmRD1aQ/onto#>
PREFIX rdfonm: <http://rdf.openmolecules.net/#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>

SELECT DISTINCT ?title ?chebi WHERE {
  ?solvent a ons:Solvent .
  ?solvent dc:title ?title .
  ?solvent rdfonm:chebiid ?chebi
}
```

The key thing here to understand, and I know this is rather abstract, is that the RDF document we build for the ONS Solubility data does not define the relation between Solvent and ChEBI identifiers, but using RDF we know this to be true. Only because the system now understands the `owl:sameAs` relation.

Now, Pellet does not stop there, and there are many more statements we can make. Even better, anyone can plug in such relations. Any database can define *owl:sameAs* and other relations, so that we can transparently browse the internet for chemistry in a semantically meaningful way.

I also know that the above is rather technical. For those chemists who have not stopped reading yet, what I would like to hear from you is what data you would like to see linked. It does not really matter what, because we can do it all (given Open Data).