chem-bla-ics

Boiling points in Wikidata

Egon Willighagen

Published August 12, 2023

Citation
Willighagen, E. (n.d.). In *chem-bla-ics*. chem-bla-ics. https://doi.org/10.59350/xr6k8-z4480

Keywords
Rdf, Wikidata, Chemistry

Copyright
Copyright © Egon Willighagen 2023. Distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

chem-bla-ics

Some days ago, I started added boiling points to Wikidata, referenced from Basic Laboratory and Industrial Chemicals (wikidata:Q22236188), David R. Lide's 'a CRC quick reference handbook' from 1993 (well, the edition I have). But Wikidata wants pressure (wikidata:P2077) info at which the boiling point (wikidata:P2102) was measured. Rightfully so. But I had not added those yet, because it slows me and can be automated with QuickStatements.

I just need a few SPARQL queries to list to which statements the qualifiers needs to be added. Basically, all boiling points which has the book as a reference and that do not have the pressure info. First, there are values with 'unknown value', which results in blank nodes (by the time you read this, they likely are already fixed):

```
SELECT ?cmp ?bp ?pressure WHERE {
  ?cmp p:P2102 ?bpStatement .
  ?bpStatement prov:wasDerivedFrom/pr:P248 wd:Q22236188 ;
    ps:P2102 ?bp .
  ?bpStatement pq:P2077 ?pressure .
  FILTER (contains(str(?pressure), "http://"))
}
```

So, to get the list for which I want to write the QuickStatements which does not have any P2077 qualifier yet, I use this query:

```
SELECT ?cmp WHERE {
  ?cmp p:P2102 ?bpStatement .
  ?bpStatement prov:wasDerivedFrom/pr:P248 wd:Q22236188 ;
    ps:P2102 ?bp .
  MINUS { ?bpStatement pq:P2077 ?pressure }
}
```

At the time of writing, this lists 54 boiling points.

I can the WDQS create CSV-styled QuickStatements with:

```
SELECT (SUBSTR(STR(?cmp),32) AS ?qid) ?P2102 ?qal2077 WHERE {
  ?cmp p:P2102 ?bpStatement .
  ?bpStatement prov:wasDerivedFrom/pr:P248 wd:Q22236188 ;
    ps:P2102 ?P2102 .
  MINUS { ?bpStatement pq:P2077 ?pressure }
  BIND ("101.325U21064807" AS ?qal2077)
}
```

Here, the SPARQL variables double as QuickStatement instructions. Finally, note to use of "U21064807" which is the Wikidata item for kilopascal (wikidata:Q21064807).

I also need to "add" the boiling point again, to make sure QuickStatements knows which statement to add the qualifier to. I think this can be done better, but not sure how to target

statements directly. This is not fool proof: I noted that this approach ignores the situation where there are two statements with the (exact) same boiling point, but different error margins. But that I will monitor and where needed correct manually.