chem-bla-ics

Tagging Molecules: a mashup of Connotea and RDF

Egon Willighagen

Published September 19, 2007

Citation
Willighagen, E. (2007). Tagging Molecules: a mashup of Connotea and RDF. In *chem-bla-ics*. chem-bla-ics. https://doi.org/10.59350/dvdc0-ps191

Keywords
Rdf, Connotea, Inchi

Abstract
Using the InChI and the new rdf.openmolecules.net website, it is now possible to tag molecules. And if you use Connotea for that, your tags will even show up on the rdf.openmolecules.net website. For example, at the time of writing, methane was tagged with alkanes and gas.

Using the InChI and the new rdf.openmolecules.net website, it is now possible to tag molecules. And if you use Connotea for that, your tags will even show up on the rdf.openmolecules.net website. For example, at the time of writing, methane was tagged with *alkanes* and *gas*.

The trick I use, is that the rdf.openmolecules.net gives every molecule a unique HTTP URL. This simply web2.0 approach offers an enormous amount of possibilities. The simplest application is that you can tag your molecules with a set label, such as *my-tomato-set*; after all, Connotea is account based. In this way, you can do open notebook QSAR studies (though the activities would still be missing).

The aforementioned example, however, give two classifications. *Methane is an alkane* and *Methane is a gas* (at room temperature). Not very well determined semantics, but it is web2.0, not the semantic web.

Interestingly, given some loosely defined semantics, use Connotea to link a molecule to a certain publication. For example, I can define *Estrone is cited in the article with PubMed ID 15659855* using the tag *pmid:15659855*. I'm sure using the DOI would work too, using the tag doi:10.1107/S0108768104028344. I have not used these informal semantics in the rdf.openmolecules.net website yet, but if there is such interest, I can have such functionality hacked in minutes.

BTW, did anyone see Gene Ontology terms being used in social bookmarking services? For example, seeing a link to the PDB database with a tag *go:0008152*? Would be a bit cryptic, and, really, in this case rather minimalistic on information.

What's next?

Now comes the tedious task of converting the QSAR data sets I used in my PhD research with these tags. It's really something I wanted to do for a while now. Next on my TODO list is the Greasemonkey script that adds the tags from Connotea to PubChem.