chem-bla-ics

Including SMILES, CML and InChI in blogs

Egon Willighagen

Published December 10, 2006

Citation

Willighagen, E. (n.d.). In *chem-bla-ics*. chem-bla-ics. https://doi.org/10.59350/wpaxr-93167

Keywords

Cml, Inchi, Cb, Microformat

Copyright

Copyright © Egon Willighagen 2006. 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.

The blogs ChemBark and KinasePro have been discussing the use of SMILES, CML and InChI in Chemical Blogspace (with 70 chemistry blogs now!). Chemists seem to prefer SMILES over InChI, while there is interest in moving towards CML too. Peter commented.

Any incorporation of content other than images and free text requires some HTML knowledge, but this can be rather limited. It is up to us chemoinformaticians to write good documentation on how to do things; so here is a first go.

Including CML in blogs and other RSS feeds

I blogged about including CML in blogs last February, and can generally refer to this article published last year: *Chemical markup, XML, and the World Wide Web. 5. Applications of chemical metadata in RSS aggregators* (PMID:15032525, DOI:10.1021/ci034244p). Basically, it just comes down to putting the CML code into the HTML version of your blog content, though I appreciate the need for plugins.

Including SMILES, CAS and InChI in blogs

Including SMILES is much easier as it is plain text, and has the advantage over InChI that it is much more readable. Chris wondered in the KinasePro blog on how to tag SMILES, while Paul did the same on ChemBark about CAS numbers.

Now, users of PostGenomic.com know how to add markup to their blogs to get PostGenomic index discussed literature, website and conferences. Something similar is easily done for chemistry things too, as I showed in Hacking InChI support into postgenomic.com (which was put on lower priority because of finishing my PhD). PostGenomic.com basically uses microformats, which I blogged about just a few days ago in Chemo::Blogs #2 , where I suggested the use of `<span class="chemicalcompound">asperin</span>`.

And this is the way SMILES, CAS and InChI's can be tagged on blogs. The `<span>` element is HTML code to indicate a bit of similar content in HTML, and can, among many other things, be formatted differently than other text. However, this can also be used to add semantics in a relatively cheap, but accepted, way. [Microformats](http://microformats.org/ are formalized just by use, so whatever we, as chemistry bloggers, use will become the de facto standard. Here are my suggestions:

- for SMILES: `<span class="smiles">CCO</span>`
- for CAS registry numbers: `<span class="casnumber">50-00-0</span>`
- for InChI: `<span class="inchi">InChI=1/CH4/h1H4</span>`

The RDFa alternative

The future, however, might use RDFa over microformats, so here are the RDFa equivalents:

- for SMILES: `<span class="chem:smiles">CCO</span>`

chem-bla-ics

- for CAS registry numbers: `<span class="chem:casnumber">50-00-0</span>`
- for InChI: `<span class="chem:inchi">InChI=1/CH4/h1H4</span>`

which requires you to register the namespace `xmlns:chem="http://www.blueobelisk.org/chemistryblogs/"` somewhere though. Formally, the URN for this namespace needs to be formalized; Peter, would the Blue Obelisk be the platform to do this? BTW, this is more advanced, and currently does not have practical advantages over the use of microformats.