Why chemistry-rich RSS feeds matter…

Egon Willighagen ⓘ

Citation

Willighagen, E. (2008). Why chemistry-rich RSS feeds matter… In *chem-bla-ics*. chem-bla-ics. https://doi.org/10.59350/cbnqa-gdc49

Keywords

Rss, Chemistry

Abstract

Peter wrote up an item on Nick's CrystalEye's RSS feed, and I have been enthusiastic about chemistry-enriched RSS feeds for some time. CMLRSS has the chemical data inline in the RSS; see DOI:10.1021/ci034244p, the use of CMLRSS in Chemical blogspace described here and here , and the CMLRSS support in Bioclipse .

Copyright

chem-bla-ics

Peter [wrote up an item](#) on Nick's [CrystalEye's RSS feed](#), and I have been enthusiastic about chemistry-enriched RSS feeds for some time. CMLRSS has the chemical data inline in the RSS; see DOI:[10.1021/ci034244p](#), the use of CMLRSS in Chemical blogspace described [here](#) and [here](#) , and the [CMLRSS support in Bioclipse](#) .

Nick's RSS feed does not put the chemistry inline, but does link to the raw CML file:

```
<entry>
  <title>No title supplied</title>
  <link rel="enclosure" href="http://wwmm.ch.cam.ac.uk/crystaleye/summary//acs/inocaj/200
  <!-- much more, that I skipped for brevity -->
</entry>
```

The example shown by Peter was nicely chosen: something is wrong with that example. It uncovers a bug in the pipeline, that could have been uncovered by a simple agent monitoring the RSS feed. That is why this technology is important! It allows pipelining of information between services.

Anyway, before you read on, check the [structure in the example](#) yourself *(Bis(pyrimidine-2-carboxylato-K2N,O)copper(II))*.

Done? Checked it? You saw the problem, right? Good.

I have scanned the CIF source, but that does not seem to contain the problem. It nicely shows a general limitation of commonly used chemoinformatics tools: the lack of proper atom typing (a problem I have been looking into for the [Chemistry Development Kit](#); see [Atom Typing in the CDK](#) and [Evidence of Aromaticity](#) .).

You will have noted that the 2D diagram in Peter's blog is charged. I checked the [complete CML source code](#) for the CrystelEye entry, and that contains the charges on the two oxygens bound to the cupper too. However, the copper is not charged. That leads to a rather unlike situation; that is, that crystal structures will about attract the whole laboratory to itself in a blink of an eye: there is nothing to balance the double-negative charge! It is conveniently summarized in this bit of the CML:

```
<formula formalCharge="-2" concise="C 10 H 6 Cu 1 N 4 O 4 -2">
  <atomArray elementType="C H Cu N O" count="10.0 6.0 1.0 4.0 4.0"/>
</formula>
```

Now, I also checked the *raw* CML; that seems to be unaffected too. So, the bug must be somewhere in the software that converts the raw CML into complete CML. And, before the InChI calculation, because that one is wrong too. A agent scanning the RSS feed, would have detected this. Someone interested in writing up a grant proposal on this?

BTW, the system is not awfully wrong: the negative charge on the acidic carboxyl groups is to be expected. But if the bond between the oxygen and the carbon would have been coordinating,

chem-bla-ics

not covalent, and the copper would have been +2, then it was fine. Because many chemoinformatics tools do not have really support for dative bonds, a covalent bond could be drawn, but then the oxygens should be uncharged... right, not? :)

Oh, and surely, one can do much, much more with those feeds. I blogged about that earlier in Automatic Classification of thousands of Crystal Structures .