

Tagging, thesauri or ontologies?

Egon Willighagen 

Published December 10, 2007

Citation

Willighagen, E. (2007). Tagging, thesauri or ontologies?. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/smfvd-pdy70>

Keywords

Ontology

Abstract

Controlled vocabularies, hierarchies, microformats, RDF.

Copyright

Copyright © Egon Willighagen 2007. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

chem-bla-ics

Controlled vocabularies, hierarchies, microformats, RDF. [Nico Adams](#) pointed me to [this excellent video](#):

Ein Fehler ist aufgetreten.

JavaScript kann nicht ausgeführt werden.

It's a really nifty piece of work, which goes into the differences between thesauri, controlled vocabularies, and, as such, ontologies, and social tagging systems. Both have their virtues; it is fuzzy logic versus ODEs all over again. Whether one is better than the other only depends on the problem at hand. For example, can you imagine social tagging in atom typing prior to performing force field calculations? Or, an 150-term ontology to annotate the scientific content of your literature archive?

More from where they come from...

The video appears to be made by the [Digital Ethnography](#) group, which has made several [more movies](#). Certainly something I'm going to check out over the winter holidays (I guess I am quite a bit more religious about ODOSOS than about gods).

Nico wrote: *As long as we appreciate that there may be more than one top node....* I am not entirely sure, but if he refers the thesauri, which are, a particular form of ontologies, where basically the only relations that can be found are **is-a** or **is-parent-of**, resulting in a hierarchy of controlled terminology with one top node (such as the Gene Ontology). Ontologies can and should be much richer if we really want to take advantage of our information technologies, just like we do with any graph mining. Why mould reality in a tight hierarchy?

Chemical ontologies

Peter has [not seen the movie yet] (<https://blogs.ch.cam.ac.uk/pmr/2007/12/09/ontologies-in-physics-and-chemistry/>), but replied with a recent comment he had on [CML](#):

Ebs and Michael had reviewed CML and questioned why the key concepts were atoms, molecules, electron, substances, whereas they suggested it would have been better to start from reactions. I think that's a very clear difference in orientation between endurants and perdurants. Although chemists publish reactions, most of the emphasis is on (new) substances and their properties. CML is designed to map directly onto the way chemists seem to think - at least in their public communication - e.g. through documents. Of course we can also do reactions in CML, but even there the emphasis is often on the components.

The suggestion by Ebs and Michael is indeed quite surprising: ontologies tries to capture knowledge and expressed this an a small set of terms, each of which with an accurate and non-overlapping meaning (orthogonal, if you wish). Now, the terms carbon, nitrogen, oxygen, and the other 104 elements are quite accurate and rather different from each other, at least from a chemical point of view. Sure, bonding is more difficult, and let's not start about aromaticity. But to question atoms, bonds or electrons as key concepts??