Ontologies and classification of chemicals: can they help each other?

Andrew Buxton
UDC Editorial Team, UK
Classes of chemicals

- By chemical composition
  - e.g. compounds of nitrogen and oxygen
- By uses
  - e.g. dyes, narcotics
- By physical properties
  - e.g. solids, electrical conductors
- By chemical properties
  - e.g. stable in air
Advantages of a synthetic classification

- Retrieval by part of the structure
e.g. chlorine compounds, alcohols
- Don’t need to list every chemical in CAS Registry
- Accommodate new substances
## Synthetic classification

<table>
<thead>
<tr>
<th>NAME</th>
<th>FORMULA</th>
<th>UDC</th>
<th>A synthetic classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferric chloride</td>
<td>FeCl$_2$</td>
<td>546.722’131</td>
<td>eFe o+2 eCl</td>
</tr>
<tr>
<td>Ferrous chloride</td>
<td>FeCl$_3$</td>
<td>546.723’131</td>
<td>eFe o+3 eCl</td>
</tr>
<tr>
<td>Toluene (methylbenzene)</td>
<td>C$_6$H$_5$.CH$_3$</td>
<td>547.533</td>
<td>a6 m1 c1 m1</td>
</tr>
<tr>
<td>Dimethylbenzene</td>
<td>C$_6$H$_4$(CH$_3$)$_2$</td>
<td>547.534.2</td>
<td>a6 m1 c1 m2</td>
</tr>
<tr>
<td>Dinitrobenzene</td>
<td>C$_6$H$_4$(NO$_2$)$_2$</td>
<td></td>
<td>a6 m1 (eN eO m2) m2</td>
</tr>
</tbody>
</table>
Morphine
Ontologies

“An ontology is a controlled vocabulary that describes objects and the relations between them in a formal way, and has a grammar for using the vocabulary terms to express something meaningful within a specified domain of interest” (Jermey & Browne, 2004: 94).

EXAMPLE

ChEBI - Chemical Entities of Biological Interest
Example relations in ChEBI

acetone is a methyl ketone
acetone is a propanone

sodium chloride is a inorganic sodium salt

caffeine monohydrate has part caffeine

warfarin has role rodenticide
warfarin has role anticoagulant
Graphical display of hierarchy in ChEBI

It is now possible to exclude obsolete terms from the auto-completion list when searching for terms. Simply uncheck the include obsolete terms box and any terms that have been marked as obsolete will no longer be returned in the search results.

**August 2006:**
SCOPes element added to WSDL
The SCOPes element has been added to the WSDL service declaration to improve interoperability with PHP and other languages that require it.

**July 2005:**
New webservice deployed
As announced in the previous news item, the OLS webservice has been redployed to make it compliant with the latest WSDL specifications and also improve its interoperability. If you are experiencing problems with the new webservice code, please do not hesitate to contact us for support.

**June 2006:**
Notification of WSDL change
The OLS webservice will be undergoing change that will not be backwards-compatible and might break existing code at the end of June. The current WSDL defines multiple operations with the same name but with different parameters. While such method overloading was allowed under previous WSDL specifications, it will no longer be supported.

Furthermore, to maximize interoperability, the WSDL will change from a RPC-style service to a document/wrapped style service. As such, users of the OLS webservice are encouraged to update their code to use the new WSDL, which will be published on June 20th.
How can classifications help ontologies?

1. Providing a ready-made hierarchy (problems with a synthetic classification?)

2. Providing a notation

3. Providing terms from other disciplines
How can ontologies help classifications?

1. Providing hierarchies

<table>
<thead>
<tr>
<th>chemical entity</th>
<th>54 Chemistry</th>
</tr>
</thead>
<tbody>
<tr>
<td>molecular entity</td>
<td>547 Organic chemistry</td>
</tr>
<tr>
<td>main group molecular entity</td>
<td>547.4 Multivalent acyclic compounds. Acyclic compounds with mixed functions</td>
</tr>
<tr>
<td>p-block molecular entity</td>
<td>547.45 Aldehyde alcohols. Ketone alcohols</td>
</tr>
<tr>
<td>carbon group molecular entity</td>
<td>547.454 Carbohydrates</td>
</tr>
<tr>
<td>organic molecular entity</td>
<td>547.455 Simple sugars or monosaccharides</td>
</tr>
<tr>
<td>natural product</td>
<td>547.455.6 Hexoses</td>
</tr>
<tr>
<td>carbohydrate</td>
<td>547.455.62 Aldohexoses with six carbon atoms</td>
</tr>
<tr>
<td>sugar</td>
<td>547.455.623 Glucose</td>
</tr>
<tr>
<td>monosaccharide</td>
<td></td>
</tr>
<tr>
<td>aldose</td>
<td></td>
</tr>
<tr>
<td>aldohexose</td>
<td></td>
</tr>
<tr>
<td>glucose</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hierarchy in ChEBI</th>
<th>Hierarchy in UDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aldohexose with six carbon atoms</td>
<td>547.455.623 Glucose</td>
</tr>
</tbody>
</table>
Differences found between ChEBI and UDC

1. UDC is lacking some recently produced chemicals and types of chemicals
2. UDC chemistry is lacking some biologically important substances
3. ChEBI divides compounds between p-block and transition elements.
4. UDC divides organic compounds into those with 1 functional group and those with >1
5. UDC divides inorganic compounds of an element firstly by valency
How can ontologies help classifications?

2. Providing a subject index
   e.g. synonyms, specific compounds not in the schedules

3. Providing access by roles, e.g. narcotics, herbicides

4. Access via alternative hierarchies

5. Graphical tools
Conclusions

- Both classifications and ontologies have their uses
- Making links between classifications and ontologies, such as through CAS registry numbers or perhaps including UDC numbers in collaboratively-produced ontologies, offers the possibilities of mash-ups that could combine the best features of both.