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

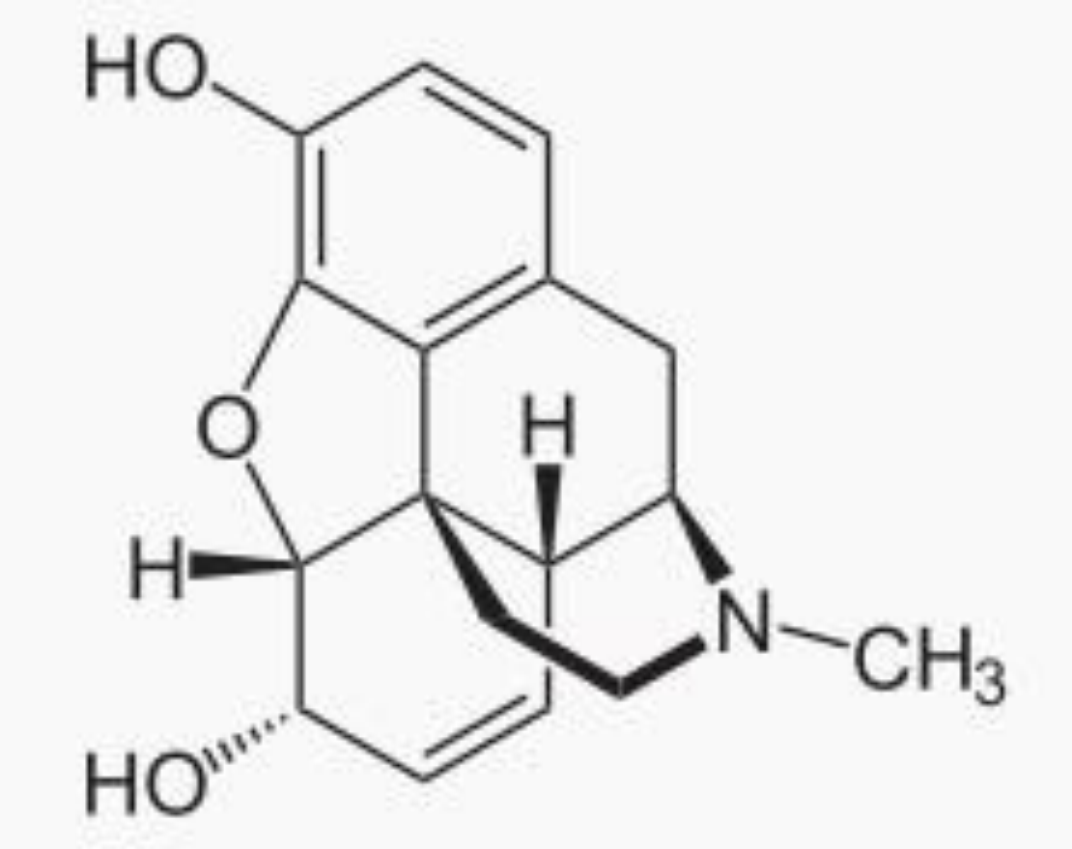
Advantages of a synthetic classification

- 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

NAME	FORMULA	UDC	A synthetic classification
Ferric chloride	$FeCl_2$	546.722’131	eFe o+2 eCl
Ferrous chloride	$FeCl_3$	546.723’131	eFe o+3 eCl
Toluene (methylbenzene)	$C_6H_5.CH_3$	547.533	a6 m1 c1 m1
Dimethylbenzene	$C_6H_4(CH_3)_2$	547.534.2	a6 m1 c1 m2
Dinitrobenzene	$C_6H_4(NO_2)_2$		a6 m1 (eN eO m2) m2

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

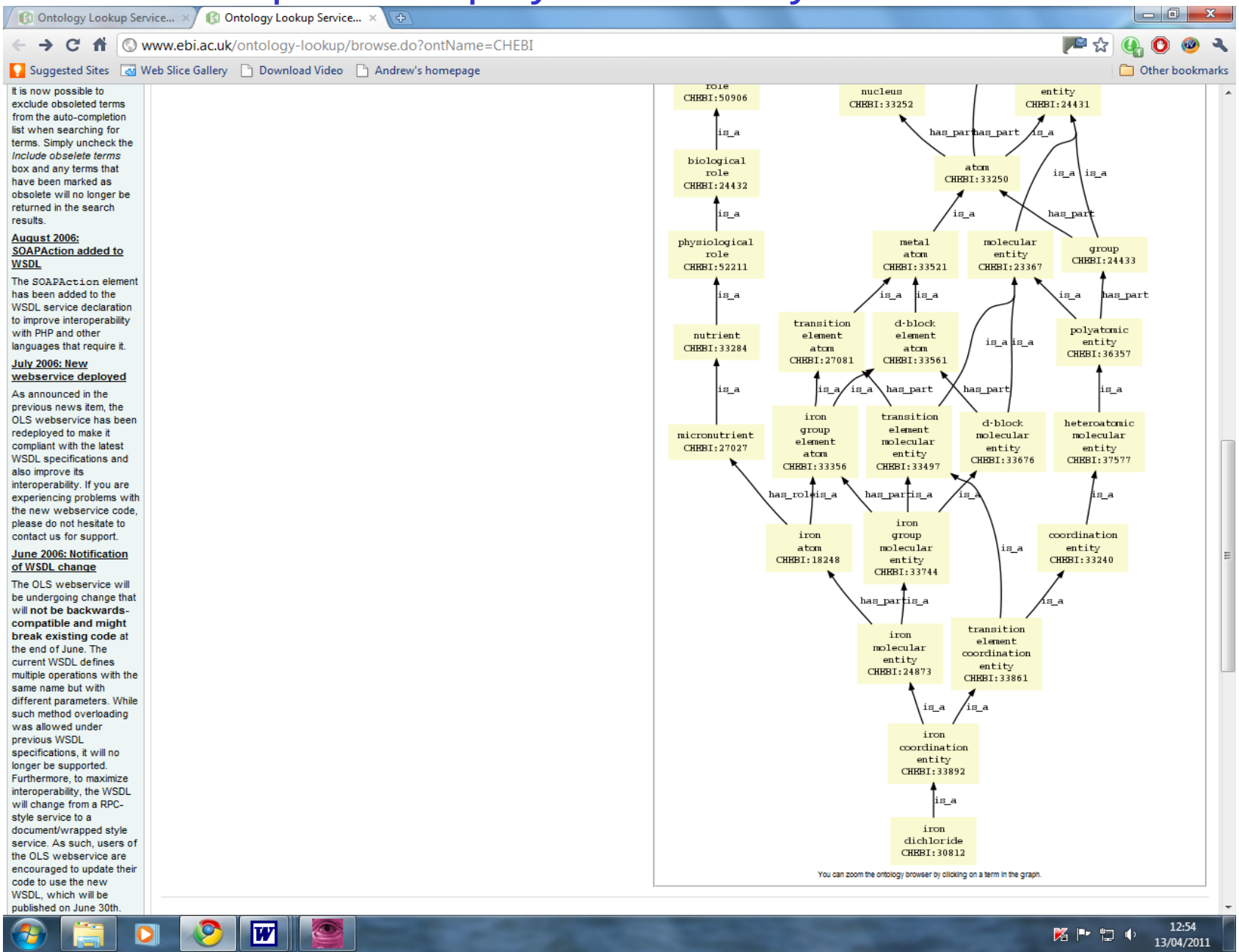
sodium chloride *is a* inorganic sodium salt

caffeine monohydrate *has part* caffeine

warfarin *has role* rodenticide

warfarin *has role* anticoagulant

Graphical display of hierarchy in ChEBI



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

chemical entity molecular entity main group molecular entity p-block molecular entity carbon group molecular entity organic molecular entity natural product carbohydrate sugar monosaccharide aldose aldohexose glucose	54 Chemistry 547 Organic chemistry 547.4 Multivalent acyclic compounds. Acyclic compounds with mixed functions 547.45 Aldehyde alcohols. Ketone alcohols 547.454 Carbohydrates 547.455 Simple sugars or monosaccharides 547.455.6 Hexoses 547.455.62 Aldohexoses with six carbon atoms 547.455.623 Glucose
Hierarchy in ChEBI	Hierarchy in UDC

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.