Chemical Knowledge Representation with Description Graphs and Logic Programming

Despoina Magka, Boris Motik and Ian Horrocks

Department of Computer Science, University of Oxford

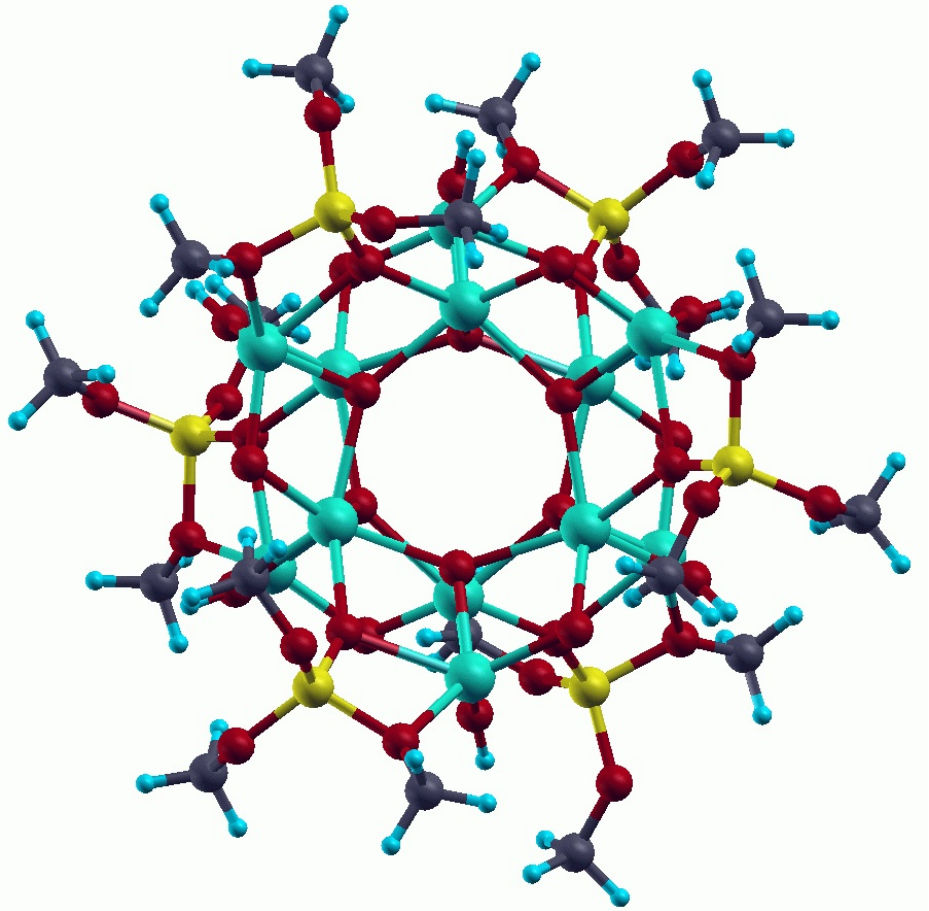
December 9, 2011

MODELLING CHEMISTRY WITH OWL

- OWL for the representation of molecular structures

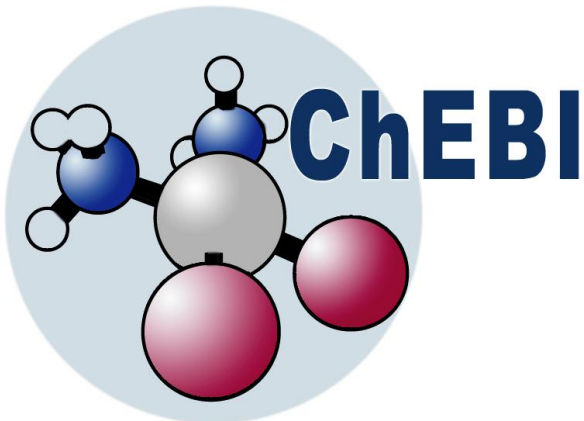
MODELLING CHEMISTRY WITH OWL

- OWL for the representation of molecular structures



Modelling chemistry with OWL

- OWL for the representation of molecular structures
- OWL ontology Chemical Entities of Biological Interest



EMBL-EBI

Modelling chemistry with OWL

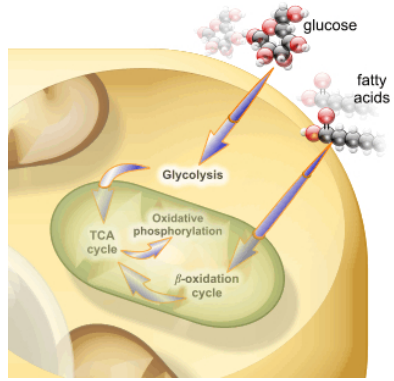
- OWL for the representation of molecular structures
- OWL ontology Chemical Entities of Biological Interest
  - High quality annotation and taxonomy of chemical compounds

Modelling chemistry with OWL

- OWL for the representation of molecular structures
- OWL ontology Chemical Entities of Biological Interest
  - High quality annotation and taxonomy of chemical compounds
  - Interoperability between researchers

Modelling chemistry with OWL

- OWL for the representation of molecular structures
- OWL ontology Chemical Entities of Biological Interest
  - High quality annotation and taxonomy of chemical compounds
  - Interoperability between researchers
  - Drug discovery and elucidation of metabolic pathways



AUTOMATE CHEMICAL CLASSIFICATION

- ChEBI is manually incremented

AUTOMATE CHEMICAL CLASSIFICATION

- ChEBI is manually incremented
- Currently $\sim$ 27,000 fully annotated entries

Automate chemical classification

- ChEBI is manually incremented
- Currently $\sim$ 27,000 fully annotated entries
- Growth rate: $\sim$ 1,500 entries per curator per year

Automate chemical classification

- ChEBI is manually incremented
- Currently $\sim$ 27,000 fully annotated entries
- Growth rate: $\sim$ 1,500 entries per curator per year
- Biologically interesting entities possibly $>$ 1,000,000

AUTOMATE CHEMICAL CLASSIFICATION

- ChEBI is manually incremented
- Currently $\sim$ 27,000 fully annotated entries
- Growth rate: $\sim$ 1,500 entries per curator per year
- Biologically interesting entities possibly $>$ 1,000,000
- Each new molecule is subsumed by several chemical classes

AUTOMATE CHEMICAL CLASSIFICATION

- ChEBI is manually incremented
- Currently $\sim$ 27,000 fully annotated entries
- Growth rate: $\sim$ 1,500 entries per curator per year
- Biologically interesting entities possibly $>$ 1,000,000
- Each new molecule is subsumed by several chemical classes
  - Is dinitrogen inorganic?

AUTOMATE CHEMICAL CLASSIFICATION

- ChEBI is manually incremented
- Currently $\sim$ 27,000 fully annotated entries
- Growth rate: $\sim$ 1,500 entries per curator per year
- Biologically interesting entities possibly $>$ 1,000,000
- Each new molecule is subsumed by several chemical classes
  - Is dinitrogen inorganic?
  - Does cyclobutane contain a four-membered ring?

AUTOMATE CHEMICAL CLASSIFICATION

- ChEBI is manually incremented
- Currently $\sim$ 27,000 fully annotated entries
- Growth rate: $\sim$ 1,500 entries per curator per year
- Biologically interesting entities possibly $>$ 1,000,000
- Each new molecule is subsumed by several chemical classes
  - Is dinitrogen inorganic?
  - Does cyclobutane contain a four-membered ring?
  - Is acetylene a hydrocarbon?

Automate chemical classification

- ChEBI is manually incremented
- Currently $\sim$ 27,000 fully annotated entries
- Growth rate: $\sim$ 1,500 entries per curator per year
- Biologically interesting entities possibly $>$ 1,000,000
- Each new molecule is subsumed by several chemical classes
  - Is dinitrogen inorganic?
  - Does cyclobutane contain a four-membered ring?
  - Is acetylene a hydrocarbon?
  - Does benzaldehyde contain a benzene ring?

Automate chemical classification

- ChEBI is manually incremented
- Currently $\sim$ 27,000 fully annotated entries
- Growth rate: $\sim$ 1,500 entries per curator per year
- Biologically interesting entities possibly $>$ 1,000,000
- Each new molecule is subsumed by several chemical classes
  - Is dinitrogen inorganic?
  - Does cyclobutane contain a four-membered ring?
  - Is acetylene a hydrocarbon?
  - Does benzaldehyde contain a benzene ring?
- Speed up curating tasks with automated reasoning tools

AUTOMATE CHEMICAL CLASSIFICATION

- ChEBI is manually incremented
- Currently $\sim$ 27,000 fully annotated entries
- Growth rate: $\sim$ 1,500 entries per curator per year
- Biologically interesting entities possibly $>$ 1,000,000
- Each new molecule is subsumed by several chemical classes
  - Is dinitrogen inorganic? $\rightsquigarrow$ Yes
  - Does cyclobutane contain a four-membered ring? $\rightsquigarrow$ Yes
  - Is acetylene a hydrocarbon? $\rightsquigarrow$ Yes
  - Does benzaldehyde contain a benzene ring? $\rightsquigarrow$ Yes
- Speed up curating tasks with automated reasoning tools

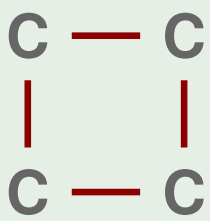
(MIS)REPRESENTING CHEMICALS WITH OWL

- Chemical compounds are highly cyclic

(MIS)REPRESENTING CHEMICALS WITH OWL

- Chemical compounds are highly cyclic

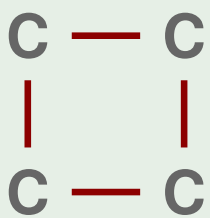
EXAMPLE



(Mis)representing chemicals with OWL

- Chemical compounds are highly cyclic
- OWL fundamentally inable to represent cycles
  - Does cyclobutane contain a four-membered ring? ×

Example

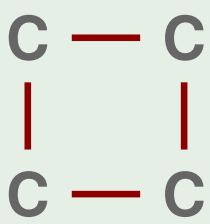


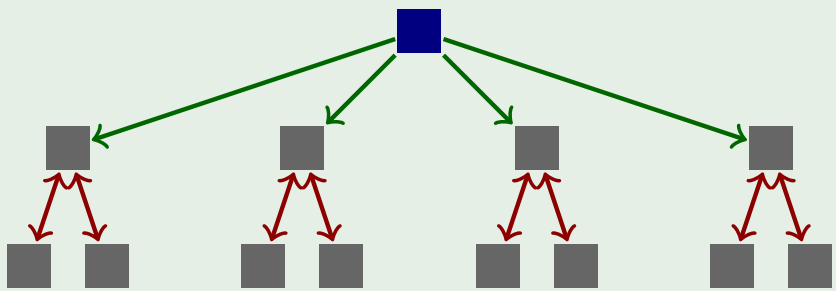
(Mis)representing chemicals with OWL

- Chemical compounds are highly cyclic
- OWL fundamentally inable to represent cycles
  - Does cyclobutane contain a four-membered ring? ×

Example

$$\text{Cyclobutane} \sqsubseteq (=4)\text{hasAtom}.(\text{Carbon} \sqcap (=2)\text{hasBond}.\text{Carbon})$$



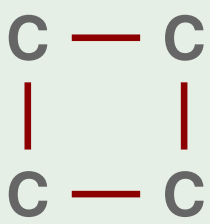


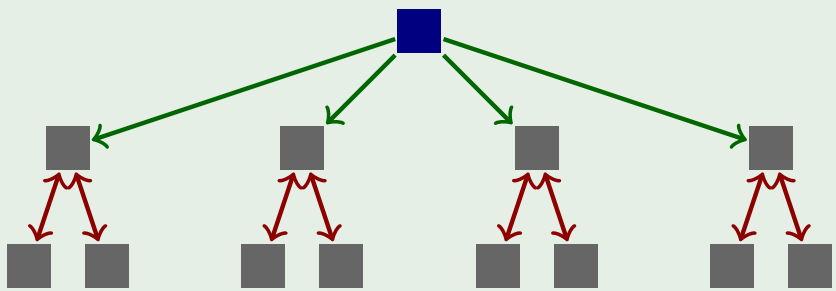
(Mis)representing chemicals with OWL

- Chemical compounds are highly cyclic
- OWL fundamentally inable to represent cycles
  - Does cyclobutane contain a four-membered ring? ×

Example

$$\text{Cyclobutane} \sqsubseteq (=4)\text{hasAtom}.(\text{Carbon} \sqcap (=2)\text{hasBond}.\text{Carbon})$$

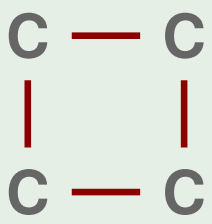


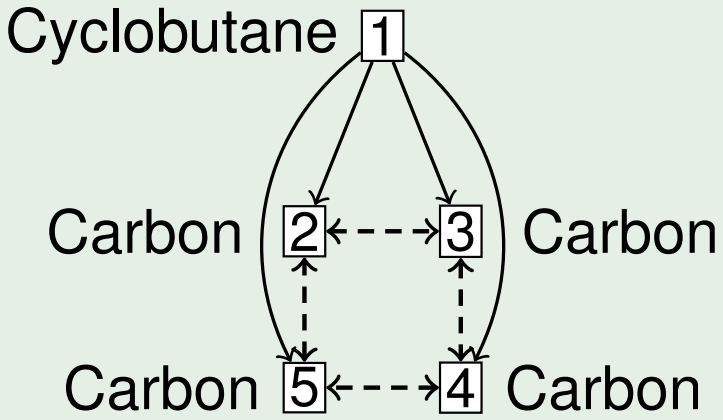


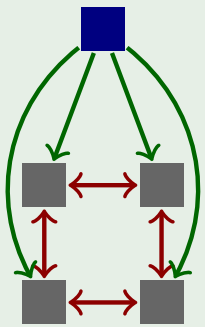
(Mis)representing chemicals with OWL

- Chemical compounds are highly cyclic
- OWL fundamentally inable to represent cycles
  - Does cyclobutane contain a four-membered ring? ×
- Limitations of OWL (partially) remedied by extension of OWL with Description Graphs and rules [Motik et al., 2009]

Example



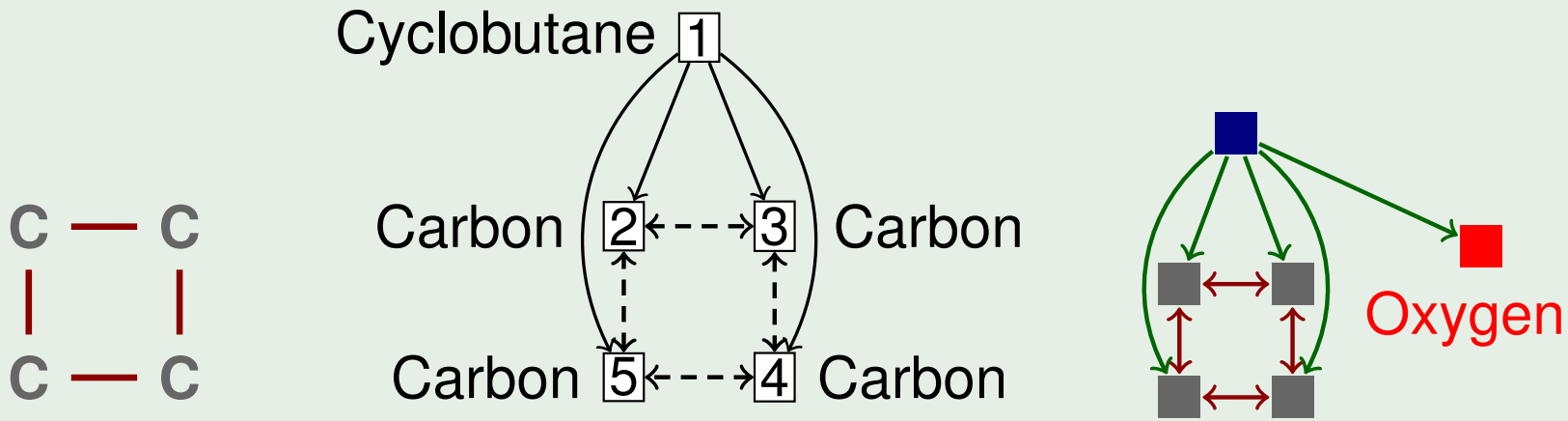




(Mis)representing chemicals with OWL

- Chemical compounds are highly cyclic
- OWL fundamentally inable to represent cycles
  - Does cyclobutane contain a four-membered ring? ×
- Limitations of OWL (partially) remedied by extension of OWL with Description Graphs and rules [Motik et al., 2009]

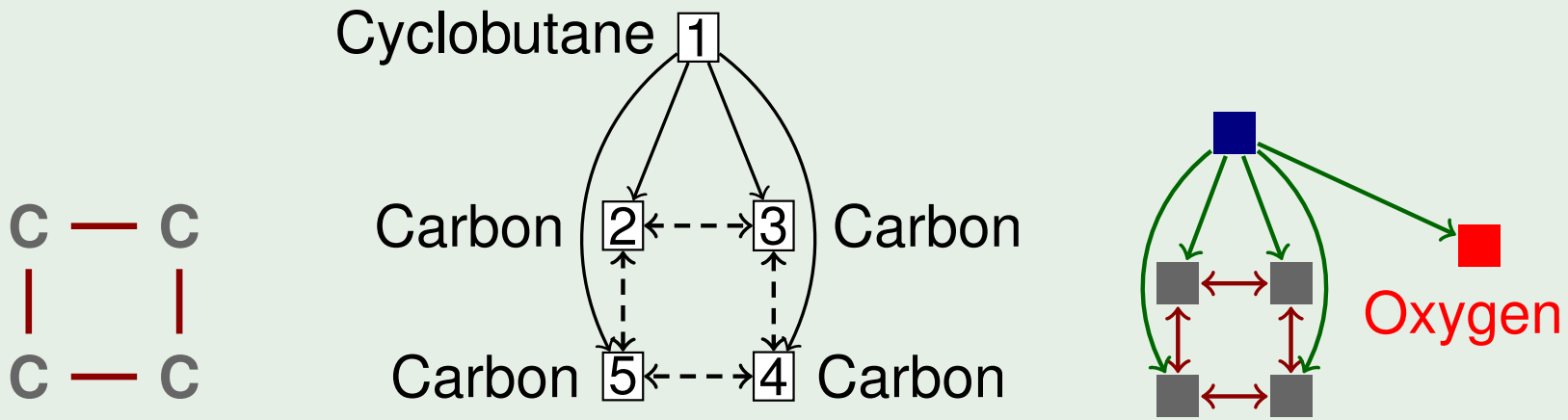
Example



(Mis)representing chemicals with OWL

- Chemical compounds are highly cyclic
- OWL fundamentally inable to represent cycles
  - Does cyclobutane contain a four-membered ring? ×
- Limitations of OWL (partially) remedied by extension of OWL with Description Graphs and rules [Motik et al., 2009]
  - Is cyclobutane a hydrocarbon? ×

Example



Description Graph Logic Programs

- Switch to logic programming

Description Graph Logic Programs

- Switch to logic programming
  - Replace classical negation with negation-as-failure

Description Graph Logic Programs

- Switch to logic programming
  - Replace classical negation with negation-as-failure
  - Translate DGs into logic programs with function symbols

Description Graph Logic Programs

- Switch to logic programming
  - Replace classical negation with negation-as-failure
  - Translate DGs into logic programs with function symbols
  - Ensure decidability via a novel acyclicity condition based on the derivations of the reasoning algorithm

Description Graph Logic Programs

- Switch to logic programming
  - Replace classical negation with negation-as-failure
  - Translate DGs into logic programs with function symbols
  - Ensure decidability via a novel acyclicity condition based on the derivations of the reasoning algorithm
- Expressive and decidable logic-based formalism for chemical knowledge representation

Description Graph Logic Programs

- Switch to logic programming
  - Replace classical negation with negation-as-failure
  - Translate DGs into logic programs with function symbols
  - Ensure decidability via a novel acyclicity condition based on the derivations of the reasoning algorithm
- Expressive and decidable logic-based formalism for chemical knowledge representation
  - Is dinitrogen inorganic?
  - Does cyclobutane contain a four-membered ring?
  - Is acetylene a hydrocarbon?
  - Does benzaldehyde contain a benzene ring?

Description Graph Logic Programs

- Switch to logic programming
  - Replace classical negation with negation-as-failure
  - Translate DGs into logic programs with function symbols
  - Ensure decidability via a novel acyclicity condition based on the derivations of the reasoning algorithm
- Expressive and decidable logic-based formalism for chemical knowledge representation
  - Is dinitrogen inorganic? ✓
  - Does cyclobutane contain a four-membered ring? ✓
  - Is acetylene a hydrocarbon? ✓
  - Does benzaldehyde contain a benzene ring? ✓

Description Graph Logic Programs

- Switch to logic programming
  - Replace classical negation with negation-as-failure
  - Translate DGs into logic programs with function symbols
  - Ensure decidability via a novel acyclicity condition based on the derivations of the reasoning algorithm
- Expressive and decidable logic-based formalism for chemical knowledge representation
  - Is dinitrogen inorganic? ✓
  - Does cyclobutane contain a four-membered ring? ✓
  - Is acetylene a hydrocarbon? ✓
  - Does benzaldehyde contain a benzene ring? ✓
- Encouraging results of a preliminary evaluation

Description Graph Logic Programs

- Switch to logic programming
  - Replace classical negation with negation-as-failure
  - Translate DGs into logic programs with function symbols
  - Ensure decidability via a novel acyclicity condition based on the derivations of the reasoning algorithm
- Expressive and decidable logic-based formalism for chemical knowledge representation
  - Is dinitrogen inorganic? ✓
  - Does cyclobutane contain a four-membered ring? ✓
  - Is acetylene a hydrocarbon? ✓
  - Does benzaldehyde contain a benzene ring? ✓
- Encouraging results of a preliminary evaluation
- Thank you. See you in the poster session!