REACTIVE CHEMICAL HAZARD ALERTING IN PHARMACEUTICAL ELECTRONIC LAB NOTEBOOKS

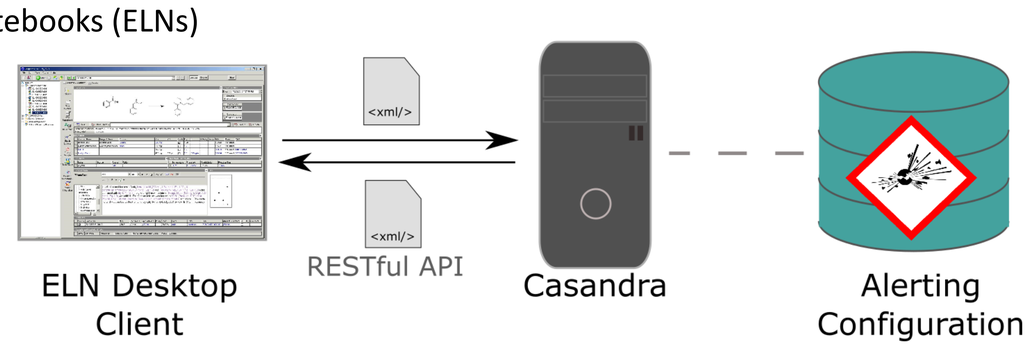
John May and Roger Sayle
NextMove Software Ltd, Cambridge, UK.

1. Introduction

Of the many chemical reactions performed by synthetic chemists in the pharmaceutical industry and academia, some are potentially more hazardous than others. Fortunately, best practices, compliance, and education helps ensure that incidents are rare, but as highlighted by the recent explosion and building evacuation at two UK universities in March 2015, constant vigilance is necessary to ensure a safe work environment.

2. Overview

Chemical safety information from MSDS/SDS datasheets, Bretherick's Handbook[1] or the internet, is readily available but the volume of such information makes it difficult for an experimentalist to identify relevant risks in a timely manner. Casandra is a server for delivering real time safety warnings of experimental hazards straight to the pharmaceutical electronic laboratory notebooks (ELNs)



In Casandra, safety alerts from an in-house or public ruleset are encoded as XML (Alerting Configuration). The server then builds an expression tree of this ruleset allowing efficient matching of potential hazards in real time. Expressions can extend beyond simple name lookup utilising the chemical structure and quantities of reactants to identify hazards.

Experimental reactions are encoded in XML by an ELN desktop client and submitted to the server for analysis. Potential hazards are identified, a report is generated and sent back to the client. The report can contain a concise description, citation, classification, actions (e.g. SOP), and a highlighted structure diagram. The ELN client handles the report and chooses whether to prompt the user for compliance.

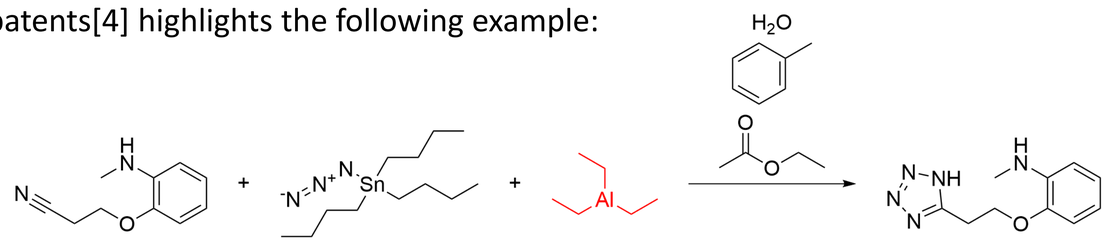
3. Identity matching

Sometimes just the presence of a compound in a reaction may be cause for caution. Examples may include: toxic, volatile, radioactive, or odorous compounds. Occurrence of a compound is identified with Casandra through identity matching.

The configuration encodes a name and optionally a structure. If no structure has been specified, Casandra can assign one using name-to-structure tools such as OPSIN[2] or NextMove Software's LeadMine.

```
<alert mesg="Pyrophoric - ignites spontaneously"/>
  <containsComponent name="triethylaluminum" smiles="CC[Al](CC)CC"/>
</alert>
```

Identity matching uses name, InChI, and canonical SMILES comparison. When no structure can be assigned name matching is still utilised. Chemical spelling correction[3] may be used for non-exact name matches. Running the above configuration in Casandra over 1.2 million reactions extracted from the US patents[4] highlights the following example:



US20120122891A1 [paragraph:499]

In this particular case the risk may be minimal since the reagent is present in toluene (rather than being powdered) and only at small amounts (9.3 ml). ELN desktop clients typically include the quantity of a compound used in an experiment. The configuration can therefore be modified to only trigger above a certain threshold.

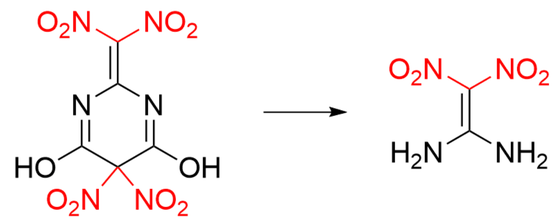
```
<alert mesg="Pyrophoric - ignites spontaneously"/>
  <containsComponent name="triethylaluminum" smiles="CC[Al](CC)CC"
                     aboveQuantity="50 ml"/>
</alert>
```

4. Substructure patterns

Capturing hazards related to specific compounds with the identity matching can be informative but is limited to known and encoded examples. Compounds that are primarily hazardous due the presence of a specific functional group can be matched with more generic patterns. In the previous section the example reaction also contains an azide which one may want to detect.

```
<alert mesg="Highly energetic group"/>
  <containsPattern name="azide" smarts="*~[ND2]~[ND2]~[ND1]"/>
  <containsPattern name="nitro" smarts="*~[ND3](~[OD1])~[OD1]"/>
</alert>
```

As with the identity matching, a quantity threshold may be specified. The condition can be tuned to only alert when the specified number of occurrences are found. For example, checking US patents for reactions where one compound contains four nitro groups finds the following:

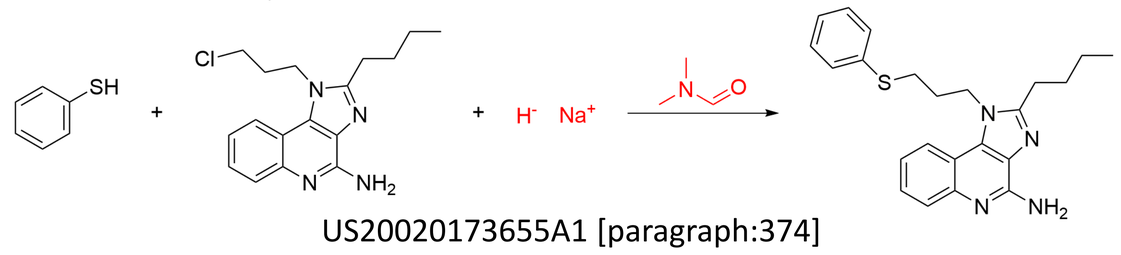


US20100081811A1 [paragraph:18]

5. Reactive incompatibles

Reactive incompatibilities representative a more subtle class of hazard where compounds may have minimal risk in isolation but are dangerous when combined.

Combinations of compounds are encoded by conjunction where all child expressions must be satisfied for the alert to trigger. An example found by Casandra in the US patents is:



US20020173655A1 [paragraph:374]

Here dimethylformamide (DMF) and sodium hydride react exothermically in a self-accelerating reaction. Care is taken when matching sodium hydride to maintain the fragment grouping.

An extension to specific incompatibles is the ability to encode hierarchical generic incompatibles[5]:

(DMF + NaH) *is child of* (amide + metal hydride)
(acetone + H_2O_2) *is child of* (ketone + peroxide)

Groupings of conditions can be nested arbitrarily and allow non-binary combinations (e.g. ammonium nitrate + formamide + H_2O).

6. Summary and future work

We have described how Casandra provides the alerting of experimental hazards. The capturing of alerts in a flexible XML format facilitates exchange and reuse. Future work includes extending conditions to check theoretical calculated property predictions, such as maximum heat of decomposition/deflagration, flashpoint, oxygen balance, radioactive half-life and vapour pressure.

7. Acknowledgements

The authors would like to thank Dana Vanderwall, Ramesh Durvasula, Dong Li and Carol McNab from Bristol-Myers Squibb for their assistance and feedback.

8. Bibliography

1. Bretherick. **Bretherick's Handbook of reactive chemical hazards**
2. Lowe DM et al. **Chemical Name to Structure: OPSIN, an Open Source Solution**. *JCIM*. 51(1). 2012
3. Sayle RS et al. **Improved Chemical Text Mining of patents with Infinite Dictionaries and Automatic Spelling Correction**. *JCIM*. 51(3). 2011.
4. Lowe DM. **Unleashing over a million reactions into the wild**. 2014. Online: nextmovesoftware.com/blog/2014/02/
5. Johnson LE and Farr JK. **CRW 2.0**. *Process Safety Progress.* 27(3). 2008