AUTOMATED CHEMICAL CLASSIFICATION – CLASSYFIRE

APPLICATIONS TO CHEMICAL HEALTH & SAFETY

Yannick Djoumbou Feunang

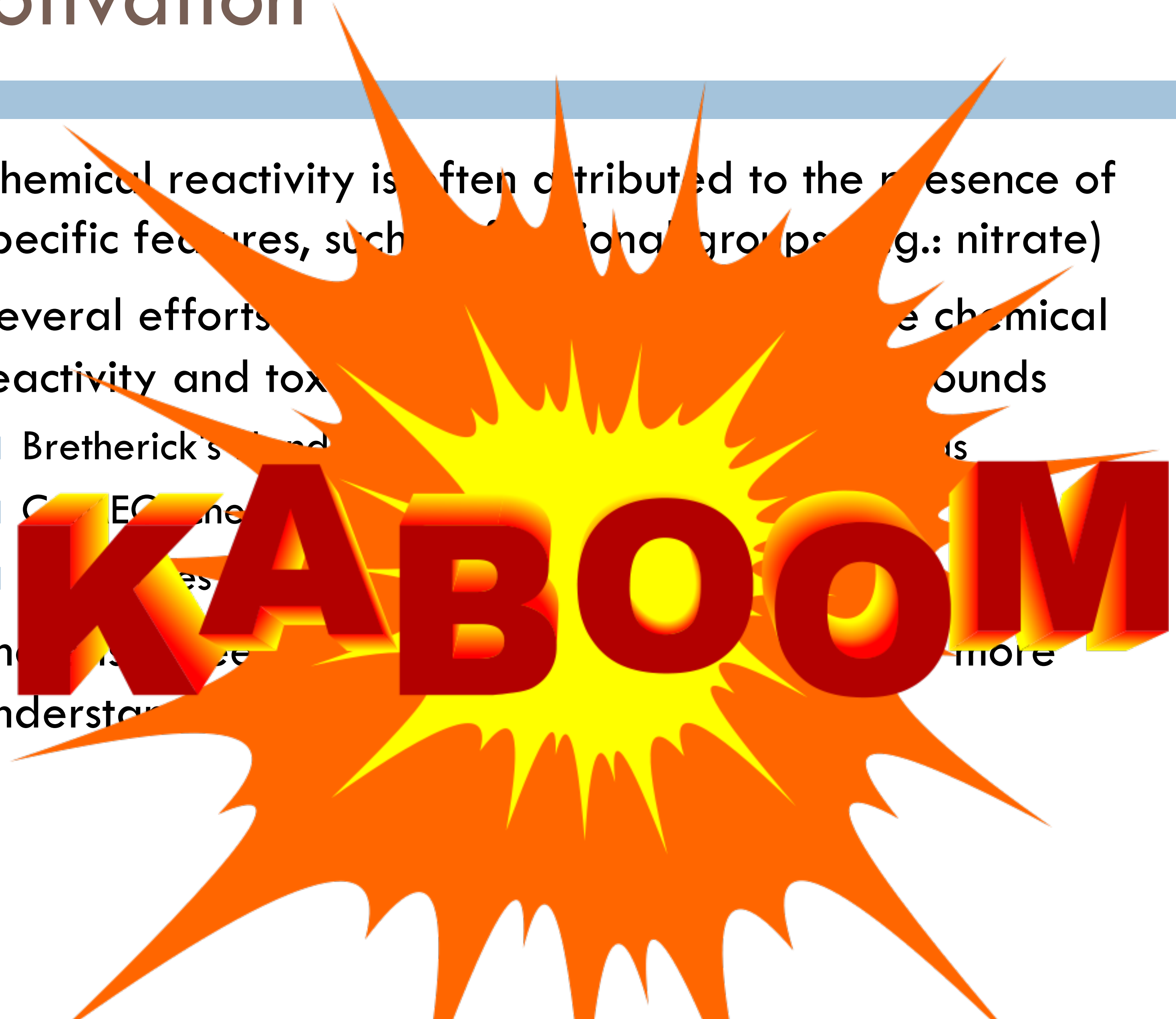
August 19, 2015

Motivation

- Chemical reactivity is often attributed to the presence of specific features, such as functional groups (e.g.: nitrate)
- Several efforts were made to study/collect the chemical reactivity and toxicity of various types of compounds
  - Bretherick's Handbook of Reactive Chemical Hazards
  - CAMEO Chemicals (NOAA)
  - Databases including HSDB, PubChem, T3DB
- There is a need to organize this data to make it more understandable, useful, and to avoid hazard

Motivation

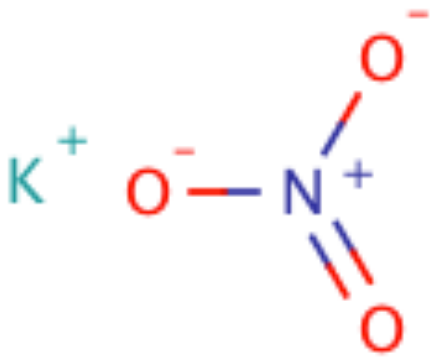
- Chemical reactivity is often attributed to the presence of specific features, such [illegible] functional groups (e.g.: nitrate)
- Several efforts [illegible] chemical reactivity and tox[illegible]ounds
  - Bretherick's [illegible]s
  - C[illegible]E[illegible]ne[illegible]
  - [illegible]es
- Th[illegible] more understa[illegible]



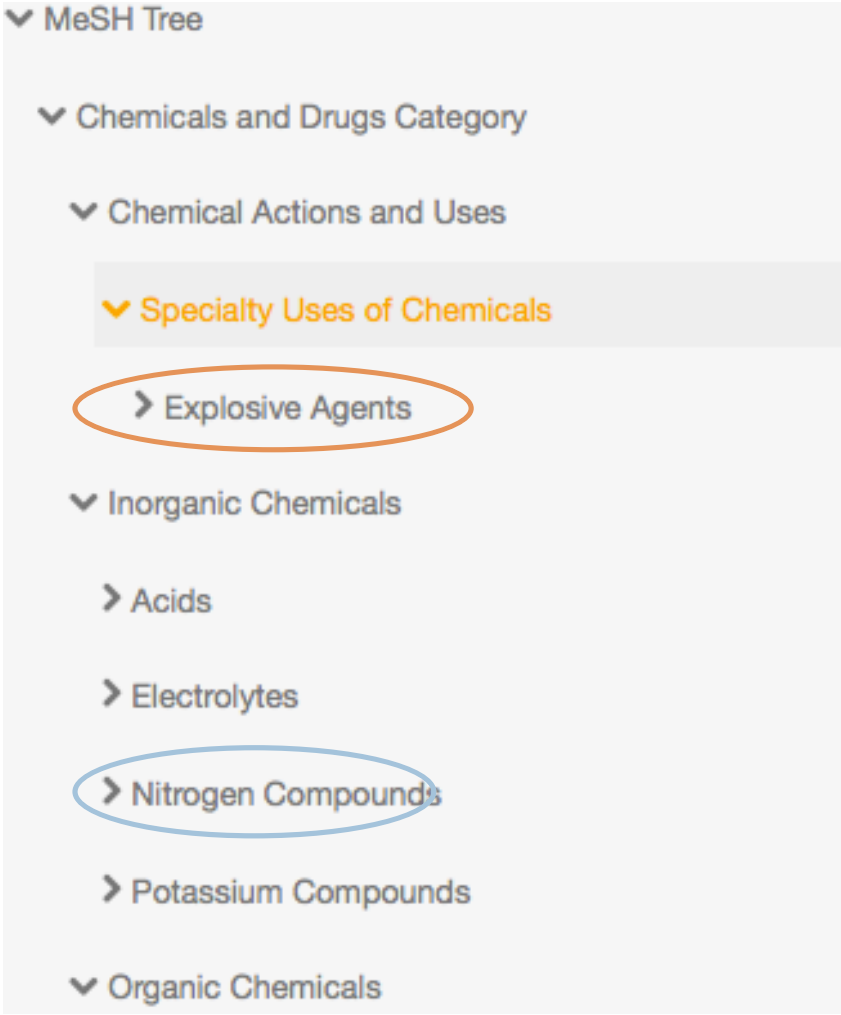
Hierarchical Chemical Classification

- Taxonomies and ontologies provide means to hierarchically organize things
- Chemicals can be classified according to their structures, reactivity profiles, etc…

Potassium Nitrate (KNO3)



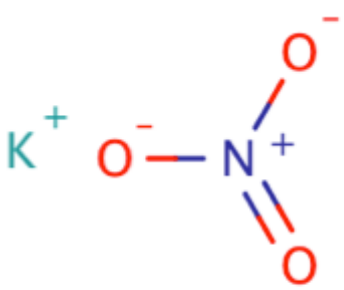
- Manual classitication is tedious, and error-prone



MeSH classification of Potassium Nitrate

Automated Classification with ClassyFire

ClassyFire, 2015 ⟶



Linnaeus, 1758 ⟶

Kingdom: Animalia
Phylum: Chordata
Class: Mammalia
Order: Artiodactyla
Family: Giraffidae
Genus: Giraffa
Species: G. camelopardalis

- Hierarchical automated data-set independent classifier
- The ontology covers 4,535 chemical classes with text-based as well as computable definitions
  - It covers inorganic and organic, metallic, natural, and synthetic compounds

Applications to Health & Safety

- Chemical reactivity can be represented in terms of Compound/ Reactivity group (RG) OR RG/RG pairs
  - E.g.: Mixing **Chlorosilanes** with **Aldehydes** may be explosive
  - ClassyFire: 7,899 Organochlorosilanes and 396,905 Aldehydes out of 41,000,000+ classified PubChem compounds.
- 51 CAMEO Chemicals Classes were Mapped to ClassyFire

CAMEO	ClassyFire
Peroxides, Organic	Organic peroxides
Acids, Carboxylic	Carboxylic acids
Carbamates	Urethanes
Organometallics	Organometallic compounds
Acids (Strong) Nonoxidizing	??

Reactivity Predictions (for each pair of reactive groups)

Esters, Sulfate Esters, Phosphate Esters, Thiophosphate Esters, and Borate Esters *mixed with* Isocyanates and Isothiocyanates

Hazard Predictions

- Reaction liberates gaseous products and may cause pressurization

Compounds containing P=O functionality catalyze decomposition of isocyanates to diimides, with evolution of CO2 gas (Six, C., F. Richter. 2003. Isocyanates, Organic. In Ullmann's Encyclopedia of Industrial Chemistry. Wiley-VCH Verlag GmhH & Co. KGaA. (Online)).

Isocyanates may decompose in the presence of organic phosphates to liberate inert CO2 gas (Smith, P. A. S., Open-Chain Nitrogen Compounds, Vol. I. New York: W. A. Benjamin, Inc., 1965, pp. 243).

Potential Gas Byproducts

- Carbon Dioxide (CO2)

Reactivity Profile

POTASSIUM NITRATE mixed with alkyl esters may explode, owing to the formation of alkyl nitrates; mixtures with phosphorus, tin (II) chloride, or other reducing agents may react explosively [Bretherick 1979. p. 108-109]. Powdered antimony mixed with potassium nitrate

The Future?

- The ClassyFire ontology is flexible and evolving
  - Adding more relevant classes for a better data representation
  - Adding more object properties to formally incorporate more information
- The ClassyFire ontology could be integrated with other ontologies (Hazards, Health effects, etc.) for a complete data pipeline
- ClassyFire also offers:
  - Rapid classification tool : ~ 500ms (4CPUs, 3.6 GB, Intel platform)
  - Web-accessible service: submit and retrieve results for free
  - Means to represent (bio-)chemical reactions and thus, infer/predict hazard information

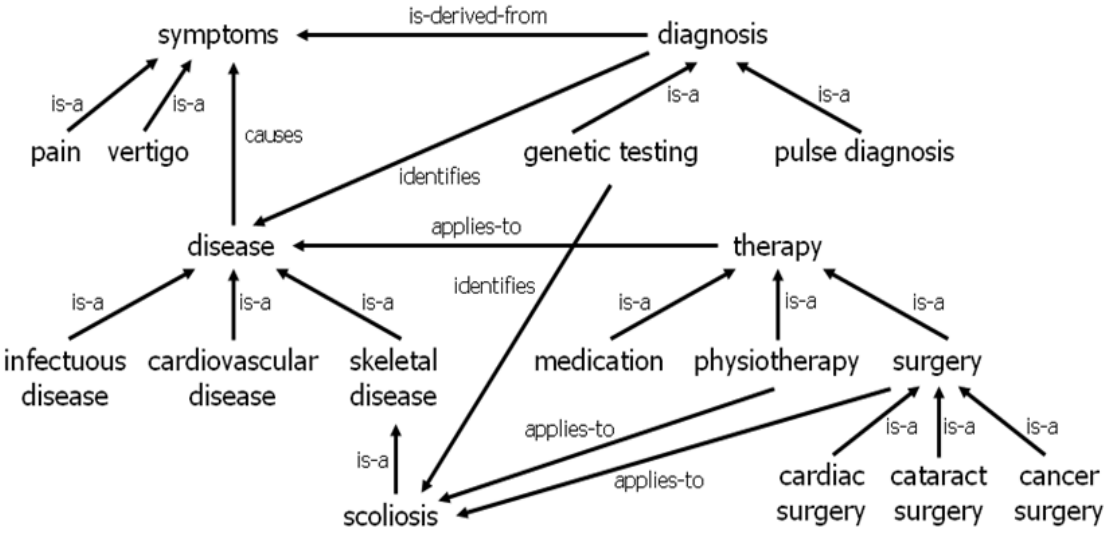
Thank You

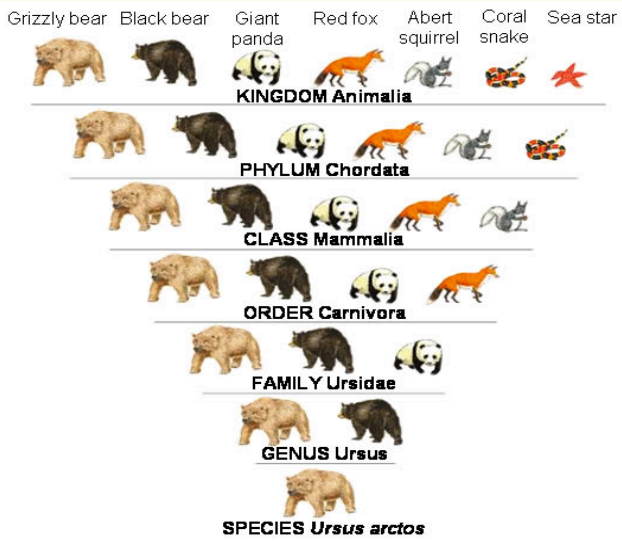
- University of Alberta
  - Wishart Lab / Greiner Lab
- NIH/PubChem
  - Evan Bolton
- EBI/ChEBI
  - Christopher Steinbeck/Jana Hastings/Gareth Owen
- LIPID MAPS, University Of California San Diego
  - Fahy Eoin
- IBM
  - Stephen K Boyer
- Leah McEwen and Ralph Stuart for this wonderful opportunity

SUPPLEMENTARY MATERIAL

What are taxonomies/ontologies?

- Taxonomies and ontologies have long been used to organize concepts and their relationships (e.g. Species, Genes, Rocks)
- They organize concepts hierarchically, allowing to:
  - Represent or/and infer relationships between concepts
  - Link entities to other concepts by deciphering relations between classes and related data from other domains





ClassyFire – The RESTful application

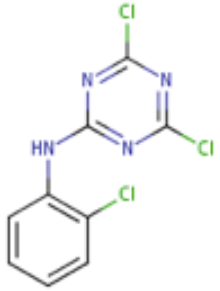
localhost:3000/entities/IMHBYKMAHXWHRP-UHFFFAOYSA-N ← Access/search by InChIKey

ClassyFire | Browse | Classify | About ClassyFire | Contact | Advanced Search

Compound Identification

SMILES

Clc1c(Nc2nc(Cl)nc(Cl)n2)cccc1



InChIKey

InChIKey=IMHBYKMAHXWHRP-UHFFFAOYSA-N

Mass

275.52

Taxonomic Classification

Taxonomy Tree

Kingdom: Organic compounds
Superclass: Organoheterocyclic compounds
Class: Triazines
Subclass: Aminotriazines

Levels of classification. They deeper the level of a class, the higher the similarity of its compounds.

Kingdom

Organic compounds

Superclass

Organoheterocyclic compounds

Class

Triazines

ClassyFire – The RESTful application

Some Applications Of Ontological Structure-based Classification Of Chemicals

- **Representation of (bio-)chemical transformations**

Transformation of 1,3-Diketones to Ketones and Carboxylates

- **Toxicology**
  - Aflatoxins bind and mutate human DNA
- **Hazard classification**
  - Oxidizers (e.g.: Organic peroxides)
  - Corrosive substances (e.g.: Chlorosilanes)

- **Database Search**
  - Compounds can be described based on their structural features.
  - Text-based search (search by Categories) – can be more efficient than structure-based search (e.g. searching for Alkaloids)