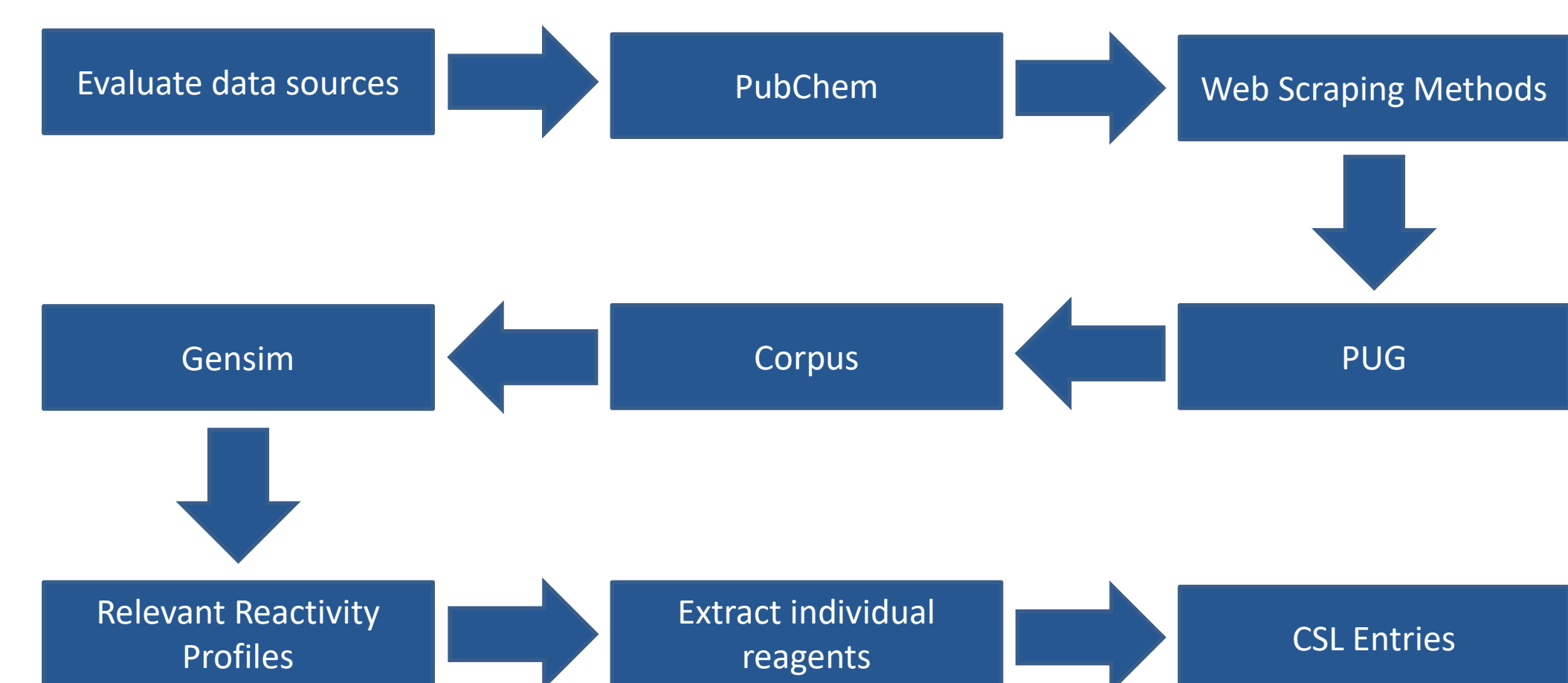


INTRODUCTION

- Pistoia Alliance and CAS: Not-for-profit organizations committed on obtaining crowd sourced data to promote safety in laboratories
- Pistoia's CSL: An open access platform developed by CAS that provides hazardous reaction information to alert scientists of dangerous reactions
- Goal: Expand the CSL by Identifying new hazardous 2+ reagent reactions
- Method: Develop model to find most reactive chemicals to enter into CSL database

PROJECT WORKFLOW



EXAMPLE SUBMISSION / ENTRIES TO CSL

Incident Report Form

Reagent/Reactants Involved in Incident Report*

Reagent/Reactant 1 - Name

Reagent/Reactant 1 - CAS Registry Number

Warning Message*

Please enter a brief description of what happened and what could have been done differently.

Reagent(s)

BROMINE

CAS RN: 7726-95-6
 MFCD: MFCD00010896
 SMILES: BrBr
 InChi: InChi=1S/Br2/c1-2
 InChi Key: GDTBXPJZTBHREO-UHFFFAOYSA-N

SULFURIC ACID

CAS RN: 7664-93-9
 MFCD: MFCD00064589
 SMILES: OS(=O)(=O)O
 InChi: InChi=1S/H2O4S/c1-5(2,3)4/h(H2,1,2,3,4)
 InChi Key: QAOWNCQODCNURD-UHFFFAOYSA-N

SODIUM BROMIDE

CAS RN: 59217-63-9, 7647-15-6
 MFCD: MFCD00003475
 SMILES: [Na+].[Br-]
 InChi: InChi=1S/BrH.Na/h1H;/q:+1/p-1
 InChi Key: JHJLBTNAGRQEKs-UHFFFAOYSA-M

Sample Entry

Date Created: 11/5/17

Modified Date: 5/24/18

Warning Message: Warning: Incorrect concentrations of sodium bromide with concentrated sulfuric acid can cause bromine gas to be released

GHS Category: Corrosive, Toxic

Functional Group:

Reaction Class: CSL00040 Bromination BROMINE

Source: User-Reported

RESEARCH METHODOLOGY 1: GENSIM

- Created a corpus with reactivity profiles of a list of compounds from PubChem
- Used Gensim to create a frequency list of select key words including:
 - Explosive, flammable, combustible, accidental, hazard
- Used to determine which compounds are the most dangerous by comparing how often these key words are shown in profiles



Why Reactivity Profiles?

- Existence of the structured Reactivity Profile indicates of a hazardous reaction
- Contains all necessary information to file incident report



List of reactivity profiles



Filter for key hazard words

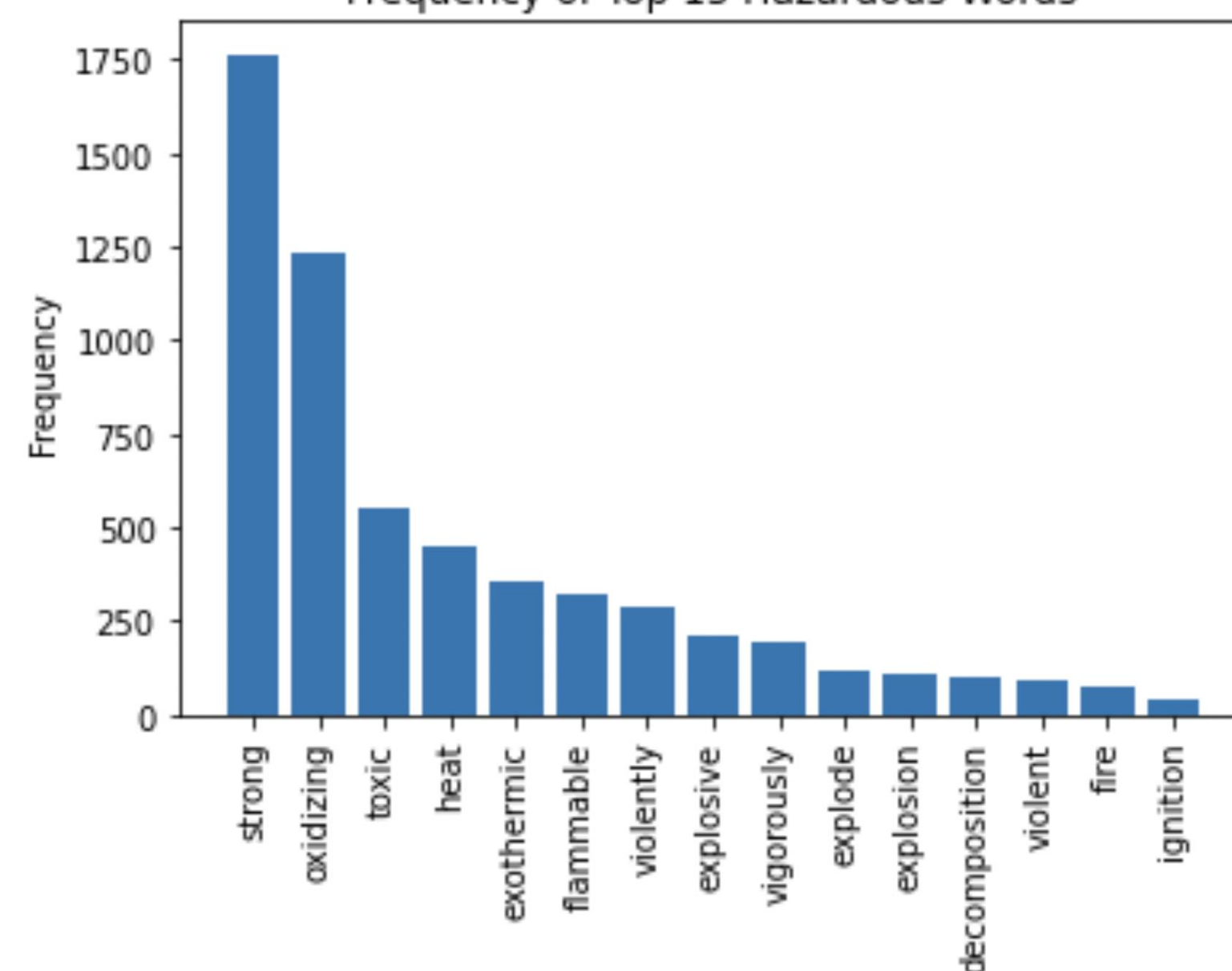


Compare back to reactivity profiles



Sort chemicals by highest frequency = most dangerous

Frequency of Top 15 Hazardous Words



9.2 Reactivity Profile

Mixing acetic acid in equal molar portions with any of the following substances in a closed container caused the temperature and pressure to increase: **2-Aminoethanol, chlorosulfonic acid, ethylene diamine, ethyleneimine** [NFPA 1991]. Acetic acid or **acetic anhydride** can **explode** with nitric acid if not kept cold. Potassium hydroxide residue in a catalyst pot reacted **violently** when acetic acid was added [MCA Case History 920. 1963]. During the production of **terephthalic acid**, n-xylene is oxidized in the presence of acetic acid. During these processes, **detonating** mixtures may be produced. Addition of a small amount of **water** may largely eliminate the risk of **explosion** [NFPA 491M.1991.p. 7]. **Acetaldehyde** was put in drums previously pickled with acetic acid. The acid caused the **acetaldehyde** to polymerize and the drums got hot and vented [MCA Case History 1764. 1971]. A mixture of **ammonium nitrate** and acetic acid **ignites** when warmed, especially if concentrated [Von Schwartz 1918. p. 322]. Several laboratory **explosions** have been reported using acetic acid and **phosphorus trichloride** to form **acetyl chloride**. Poor heat control probably caused the formation of **phosphine** [J. Am. Chem. Soc. 60:488. 1938]. Acetic acid forms **explosive** mixtures with **p-xylene** and air [Shraer, B.I. 1970. Khim. Prom. 46(10):747-750].

METHODOLOGY 2: REACTIVITY PROFILE EXTRACTION

- Web-scraped PubChem using the PUG-VIEW API to request information such as the reactivity profile and IUPAC name for a particular compound
- Split reactivity profile into sentences and extracted compounds, shown as embedded links, to form reactant groups along with the specific compound
- Filtered out sentences not containing hazard words to create warning messages for the reactant group

Reactant 1: hydrogen
 Reactant 2: 1-aminopropan-2-ol
 Warning Message: Flammable gaseous hydrogen is generated by amines in combination with strong reducing agents, such as hydrides.

Reactant 1: hydrazine hydrate
 Reactant 2: 1-chloro-2,4-dinitrobenzene
 Warning Message: The mixture of this compound with hydrazine hydrate caused a violent reaction.

Reactant 1: ammonia
 Reactant 2: 1,2-dichloroethane
 Warning Message: Liquid ammonia and ethylene dichloride can cause an explosion when mixed, nfp4 491m, 1991.

- Tf-idf model to sort reactivity profiles by highest frequency of hazard words
 - Compares individual reactivity profile to pre-determined list of words
 - Determines most dangerous compounds and reactions to add to CSL first

Chemical Name	Similarity Score
Acetyl Chloride	0.62683517
Isopropylamine	0.6027288
Bromoform	0.55904955
Ethyl Chloride	0.5437039

CONCLUSION

- Successfully built up models to extract reactivity information from PubChem
- Sorted through the corpus using dangerous word frequencies
- Formed an organized list of 2+ reagent reactions with warning messages that can be manually inputted into the CSL
- Next steps:
 - Running key word searches on different data sources
 - Automation of incident report filing
 - Include reactions of 3+ reagents

ACKNOWLEDGEMENTS & REFERENCES

Special thanks to:

- Pistoia Alliance Advisors: Gabrielle Whittick, Qiong Yuan, Mark Manfredi, Grace Baysinger, Thomas Vickery, Ralph Stuart, Jian (Jeff) Zhang
- Data Mine Mentors: Rachel Pfeifer, Ryan Altman, Nick Rosenorn, Nick Lenfestey, Mark Ward, Maggie Betz, and Data Mine Staff

References:

[PubChem](#) [Gensim](#) [PUG](#) [Comeo Chemicals](#)