

RS02 Designing chemical feature analysis for Chemists

Background

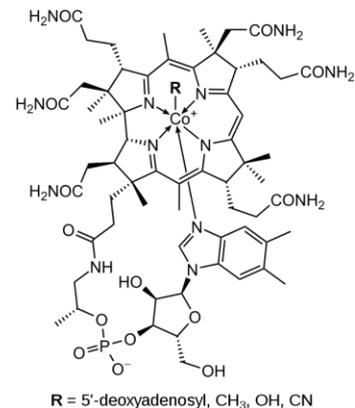
Scientists use recommender systems to quickly find the most relevant scientific literature in their field. However, in the chemistry domain, chemists are not only interested in textual relevance, but also want to find and make sense of chemical formulas contained in the literature. For example, within a publication: $C_9H_8O_4$, $C_9H_8O_4$, acetylsalicylic acid, and ASA all refer to the same compound: Aspirin! How can we design an interface that highlights such information and lets chemists quickly get an overview of all chemical compound names, chemical formulas, and drug names discussed in a publication?

Goal

- Improve the ease of finding, comparing and understanding chemical named entities (e.g. methanol, acetone, hexane, H_2O , $MG(OH)_3$, CH_2Cl_2) in scientific literature.

Tasks

- Research and compare existing toolkits for the automated extraction of chemical information from scientific literature, e.g., ChemDataExtractor (python)
- Build a prototype that visualizes the chemical information present in scientific papers to help chemist more quickly find, compare, and make sense of the chemical information present.



Corinna Breitinger
breitinger@uni-wuppertal.de



André Greiner-Petter
greinerpetter@gipplab.org

