1. Data curation

Manually curated Chemistry olympiads University exams University exercise sheets

Semi-programatically curated

Daily allowed intakes

GHS pictograms

Hazard statements

Number of NMR peaks

Electron counts

IUPAC-SMILES questions

Oxidation states

Point groups

2. Semantic annotation

Reactions -[START_RXNSMILES] [END_RXNSMILES]



[START_SMILES] [END_SMILES]



Units -

 $pu\{m^{-3}\}$



Equations $a^2 + b^2 = c^2$



Manual inspection -

Clarity and phrasing

Error analysis

-Automatic checks

Invariance to shuffling

Spelling

Schema



