chem-bla-ics

Mining the KEGG pathway database with self-organizing maps

Egon Willighagen

Published April 4, 2006

Citation

Willighagen, E. (2006). Mining the KEGG pathway database with self-organizing maps. In *chem-bla-ics*. chem-bla-ics. https://doi.org/10.59350/tzwfe-ww931

Keywords

Kegg, Chemometrics

chem-bla-ics

The Self-organizing map (SOM) is a popular (again) and intuitive non-linear mapping method: it transforms a multidimensional space into two dimensions (normally: they are so easy to visualize). Latino and Aires-de-Sousa published a paper that uses this method to analyze the whole KEGG pathway database: *Genome-Scale Classification of Metabolic Reactions: A Chemoinformatics Approach* (DOI: anie.200503833).

The method is based on earlier work by Zhang and Aires-de-Sousa: *Structure-Based Classification of Chemical Reactions without Assignment of Reaction Centers* (DOI: 10.1021/ci0502707). A non-trivial feature of the suggested method is the use of two SOMs. The first maps the reaction onto a fixed-length vector (coined MOLMAP), which is used as input vector for the second map. This later map is used to cluster the KEGG reactions on a purely chemical basis. The resemblence with the EC numbering system is striking.