INISOR : Integrated Information System on Organic Reactions

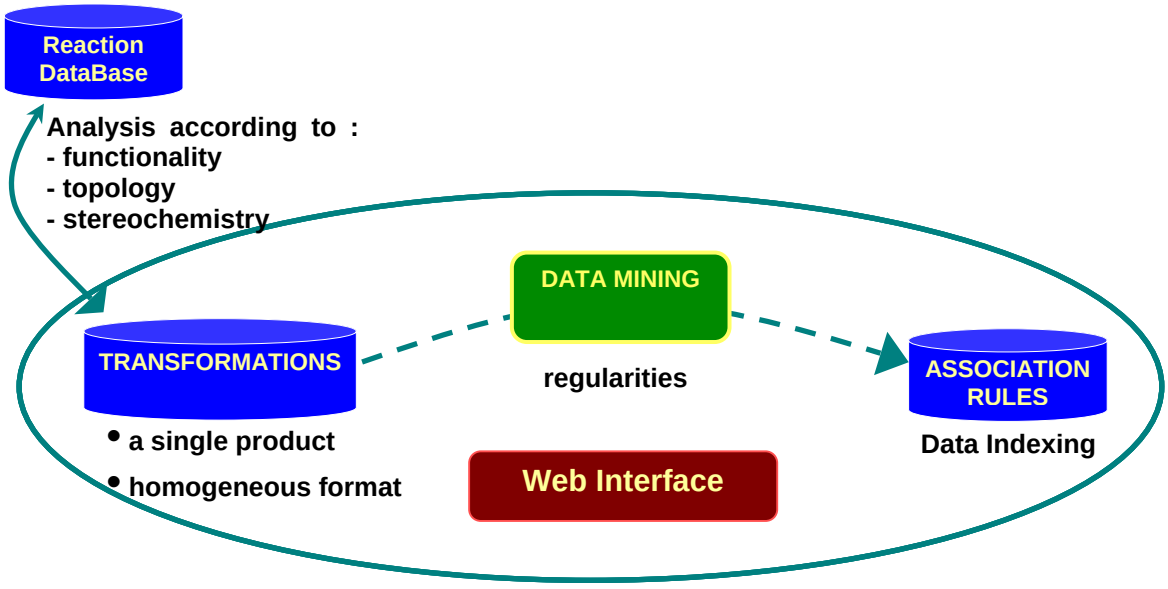
Julien Burton, Sophie Cauquil, Gilles Niel, Philippe Jauffret[1]*, Claude Laurenço*
UMR 5076 CNRS-ENSCM 8, rue de l'Ecole Normale 34296 Montpellier CEDEX 5

We describe herein the prototype of an integrated information system gathering a reaction database (RDB) and a metadata database, accessible through a Web interface. Within the domain of functional transformations in organic synthesis, these metadata may are useful for indexing reactions.

INISOR was designed because of difficulties met when an organic chemist searches for relevant information in commercial RDBs. These databases show some shortcomings, *e.g.* use of loose reaction models, lack of concept verbalization, and lack of organized knowledge compared to the huge amount of crude facts they contain. We aim at improving these information systems by first identifying required knowledge, either intellectually or through a knowledge extraction process from RDBs, and in a second time by integrating this knowledge with the initial reaction database. The purposes of this approach are i) the construction of a system involving several knowledge levels and several points of view about the reactions, ii) the help in problem solving by means of new query modes, based on synthetic objectives that are reached by these reactions.

The first step of the knowledge extraction process consists in analyzing the reactions from a particular RDB according to functional, topological and stereochemical points of view. INISOR deals currently only with the functional point of view. This analysis relies on a modelling of the functionality and on a formal definition of both chemical functions and functional interchanges.[2] During the second step, resulting structural data are transformed into textual data that are analyzed by data mining algorithms. Data mining produces association rules, which give new knowledge (metadata) about the initial database content *e.g.* how frequent functions are formed or destroyed, interchange distribution, etc. Such information is not usually associated with reaction data.[3]

This whole information is managed through a relational database. INISOR enables through its Web interface several different query modes: i) the use of metadata allows to choose the most relevant transformations to reach a given synthetic goal *e.g.* "which other function does react with an alkene to form an alcohol?", ii) a straightforward access to factual data *e.g.* "which known transformations present in the database allow the alcohol function to be formed"?



[1] philippe.jauffret@enscm.fr

[2] RESYN_Mining project, developed in collaboration with LIRMM (Laboratoire d'Informatique, Robotique et Microélectronique de Montpellier) and A. Napoli's research group of LORIA (Laboratoire Lorrain de Recherche en Informatique et ses Applications UMR 7503 CNRS/INRIA/Universités Nancy 1&2)

[3] An experiment on knowledge discovery in chemical databases. Berasaluce S., Laurenço C., Napoli A. and Niel G. *8th European Conference on Principles and Practice of Knowledge Discovery in Databases – PKDD*, 2004. (Pisa, Italia). Springer Verlag, 2004. *Lecture Notes in Artificial Intelligence*. vol 3202. pp.39-51