R. Nugmanov [1]
B. Sattarov [1]
A. Fatykhova [1]
A. Kutlushina [1]
F. Bekmuratova [1]
Z. Khayrullina [1]
E. Moiseeva [1]
R. Gimatev [1]
T. Madzhidov [1]
A. Varnek [1,2]

CGR-DB. INTERACTIVE REACTION DATABASE AND CGR-BASED SEARCH ENGINE.

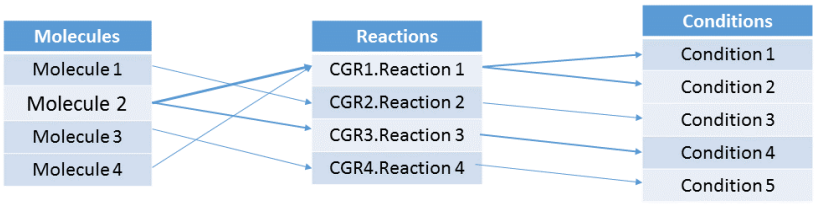
[1] A.M. Butlerov Insitute of Chemistry, Kazan Federal University, Russia
[2] University of Strasbourg, Strasbourg, France

nougmanoff@hotmail.com

At present, a large volume of data on chemical reactions has accumulated among various research teams around the world. Therefore, it is reasonable to create an online database of chemical reactions that will not only provide free access to information and search, but also provide researchers with a convenient interface for adding and modifying data.

Technology of Condensed Graph of Reaction (CGR) effectively solves complicated problems of chemical reactions storage and search. CGR usage allowed us to create a publicly accessible interactive database of chemical reactions, where substructure and similarity searches are carried out by the same approach for molecules and reactions.

Unification of the representations of molecules and reactions stored in the database allowed us to develop new effective tools for accelerating search possibilities and indexing. The problem of creating an informative and convenient reaction's similarity measure is still not completely solved, but thanks to CGR it became possible to calculate the Tanimoto index for reactions. This feature made it possible to rank reactions according to similarity. CGR-based substructure and structure searches was implemented as well.



Deduplication is a crucial principle of relational databases. In the case of chemical information storage, it means that, since the same molecules are found in various reactions, it is reasonable to put them in a separate table and bind them through many-to-many relationship to the reaction table. First, in addition to reducing the size of the stored database, this approach allows one to modify the individual structure only once, without need to rewrite each entry in different reactions.This approach greatly simplifies re-standardization. And secondly,most popular among organic chemists "reaction-by-molecule" search, now, does not require scanning all reactions or applying complex indices to speed up the search.

A separate task was to create a user-friendly mechanism for adding reactions to the database. Our implementation of "drawing-to-database" approach allows user to modify reactions directly from the database's interface, and standardization of the new reactions will be done automatically at the server side.

The research was supported by Russian Scientific Foundation, grant 14-43-00024.