From Reactivity Charts to an Industrial-Grade Engine: A Decade-Long Journey to Reaction Condition Recommendation in Reaxys

Timur Madzhidov

Elsevier Chemistry, Nielsen House, John Smith Drive, Oxford OX4 2WB, United Kingdom

Reaxys is the most comprehensive chemistry information systems, integrating manually and programmatically gathered data from millions of journal articles and patents into a unified, searchable knowledge base of substances, reactions, properties and conditions. Among the questions a synthetic chemist asks most often, “which conditions should I use?” remains one of the hardest to answer, yet one of the most valuable. This presentation reviews a decade-long effort, begun in 2016 within a long-standing Reaxys R&D collaboration, to turn that question into a reliable, scalable capability for every Reaxys user.

The range of approaches evaluated along the way. It begun with knowledge-driven systems that reconstructed classical reactivity guidance from large-scale data[1], which could inform a chemist’s intuition but not offer concrete recommendations. A broad spectrum of predictive methods explored over several years, spanning ranking and similarity-based models[2] and modern generative architectures[3]. A recurring observation is highlighted: methods that excel on small, focused datasets do not necessarily transfer to the full diversity of chemistry, and apparent sophistication does not guarantee real-world utility.

A central principle shaped the design: rather than beginning from a model, the work begins from the chemist. The system follows the user journey - mirroring how real chemists select conditions and the questions they want answered - informed by a careful study of user feedback gathered across the globe. In this sense the approach is human-inspired rather than AI-driven: fast enough for interactive, industrial-scale use while remaining transparent and trustworthy. Careful attention to the underlying data - including atom-to-atom mapping refined against a purpose-built reference set, and a structured ontology for representing reagents, catalysts and solvents - was essential to reaching production quality.

Building on these findings, we are developing the Reaxys Conditions Navigator: a capability that surfaces relevant precedents, summarises the conditions chemists have actually used, and provides the information chemists need to choose conditions - all backed by data.

Keywords: reaction condition prediction; similarity search; explainable AI; reaction data curation; condition ontology; Reaxys

References

(1) Lin, A. I.; Madzhidov, T. I.; Klimchuk, O.; Nugmanov, R. I.; Antipin, I. S.; Varnek, A. Automatized Assessment of Protective Group Reactivity: A Step Toward Big Reaction Data Analysis. *J. Chem. Inf. Model.* **2016**, *56* (10), 2140–2148. DOI: 10.1021/acs.jcim.6b00319.

(2) Afonina, V. A.; Mazitov, D. A.; Nurmukhametova, A.; Shevelev, M. D.; Khasanova, D. A.; Nugmanov, R. I.; Burilov, V. A.; Madzhidov, T. I.; Varnek, A. Prediction of Optimal Conditions of Hydrogenation Reaction Using the Likelihood Ranking Approach. *Int. J. Mol. Sci.* **2022**, *23* (1), 248. DOI: 10.3390/ijms23010248.

(3) Mazitov, D. A.; Madzhidov, T. I.; et al. Conditional Variational AutoEncoder to Predict Suitable Conditions for Hydrogenation Reactions. *Molecules* **2026**, *31* (1), 75. DOI: 10.3390/molecules31010075.