Transformer-based tool recommendation system in Galaxy

Anup Kumar[1,*], Björn Grüning[1], Rolf Backofen[1,2]

[1] Bioinformatics Group, Department of Computer Science, University of Freiburg, Georges-Koehler-Allee 106, 79110 Freiburg, Germany

[2] Signalling Research Centres BIOSS and CIBSS, University of Freiburg, Schaenzlestr. 18, 79104 Freiburg, Germany

Bioinformatics Group, Department of Computer Science, University of Freiburg, Georges-Koehler-Allee 106, 79110 Freiburg, Germany
* kumara@informatik.uni-freiburg.de
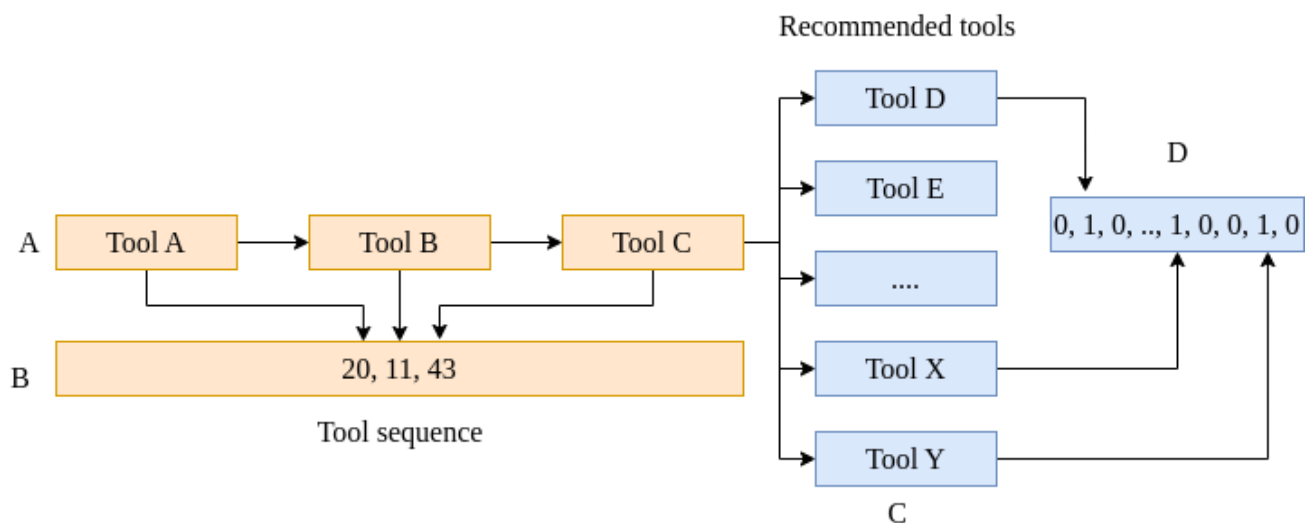
The figure shows a sample workflow with 3 tools (Tool A, Tool B and Tool C). This workflow is represented as a sequence of integers shown in step B. The set of recommended tools is shown in step C which is represented as a one-hot encoded vector shown in step D. The position corresponding to the index of a tool is set to 1. Other positions remain 0. The corresponding pairs of B and D are used to create training and test datasets. The training dataset is used for training Transformer and RNN models and the test dataset is used for evaluating the models.