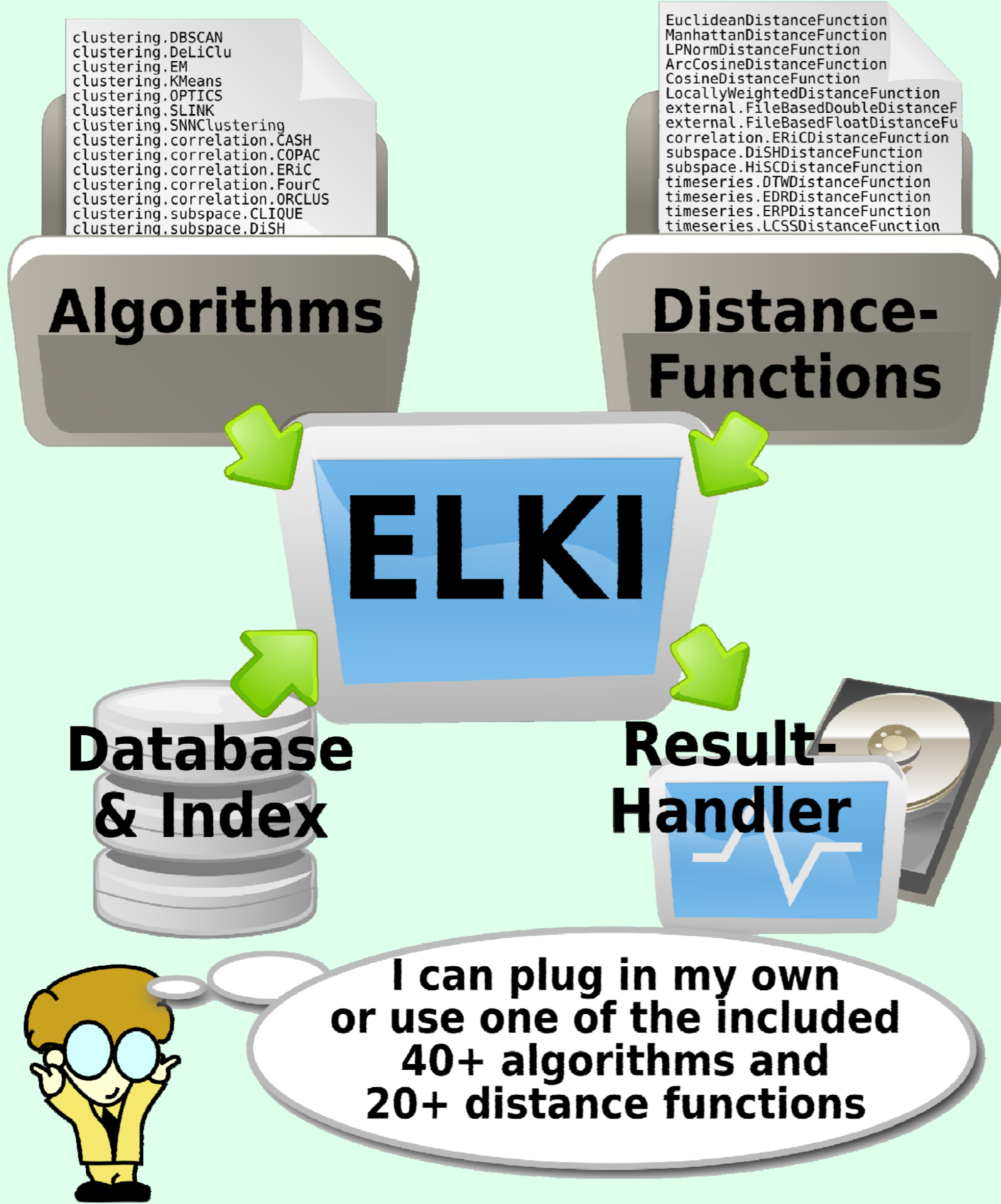


## Environment for DeveLoping KDD-Applications Supported by Index-Structures

### Motivation:

In high dimensional data, and especially in time series data, the choice of a distance measure suitable and meaningful w. r. t. the data in question is essential. In many implementations of algorithms, either provided by authors or implemented in general frameworks, the Euclidean distance is invariably used as a standard distance measure.

In the software system ELKI, we facilitate the use of a wide range of different algorithms along with a wide choice of distance measures.

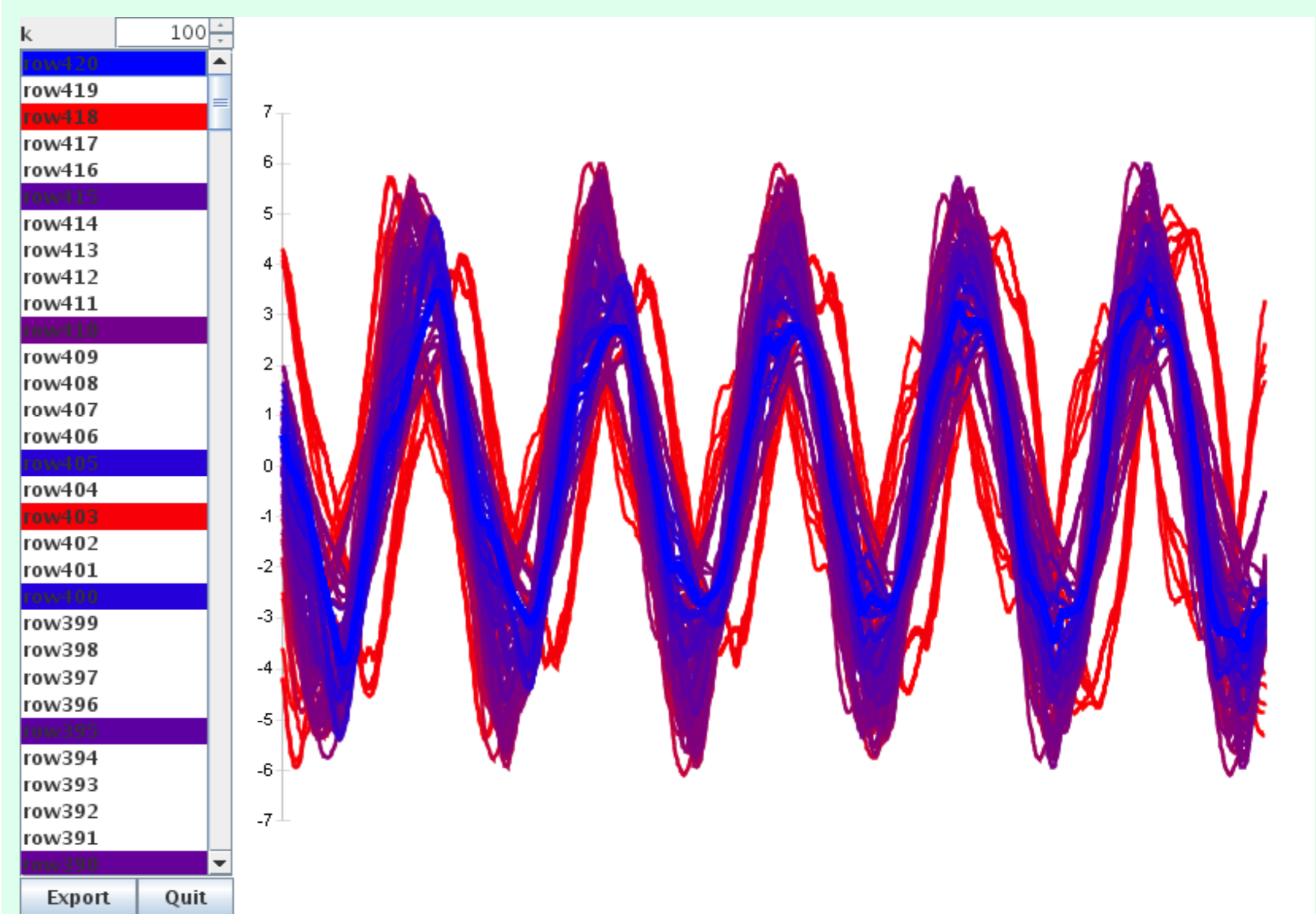


### Flexible Framework:

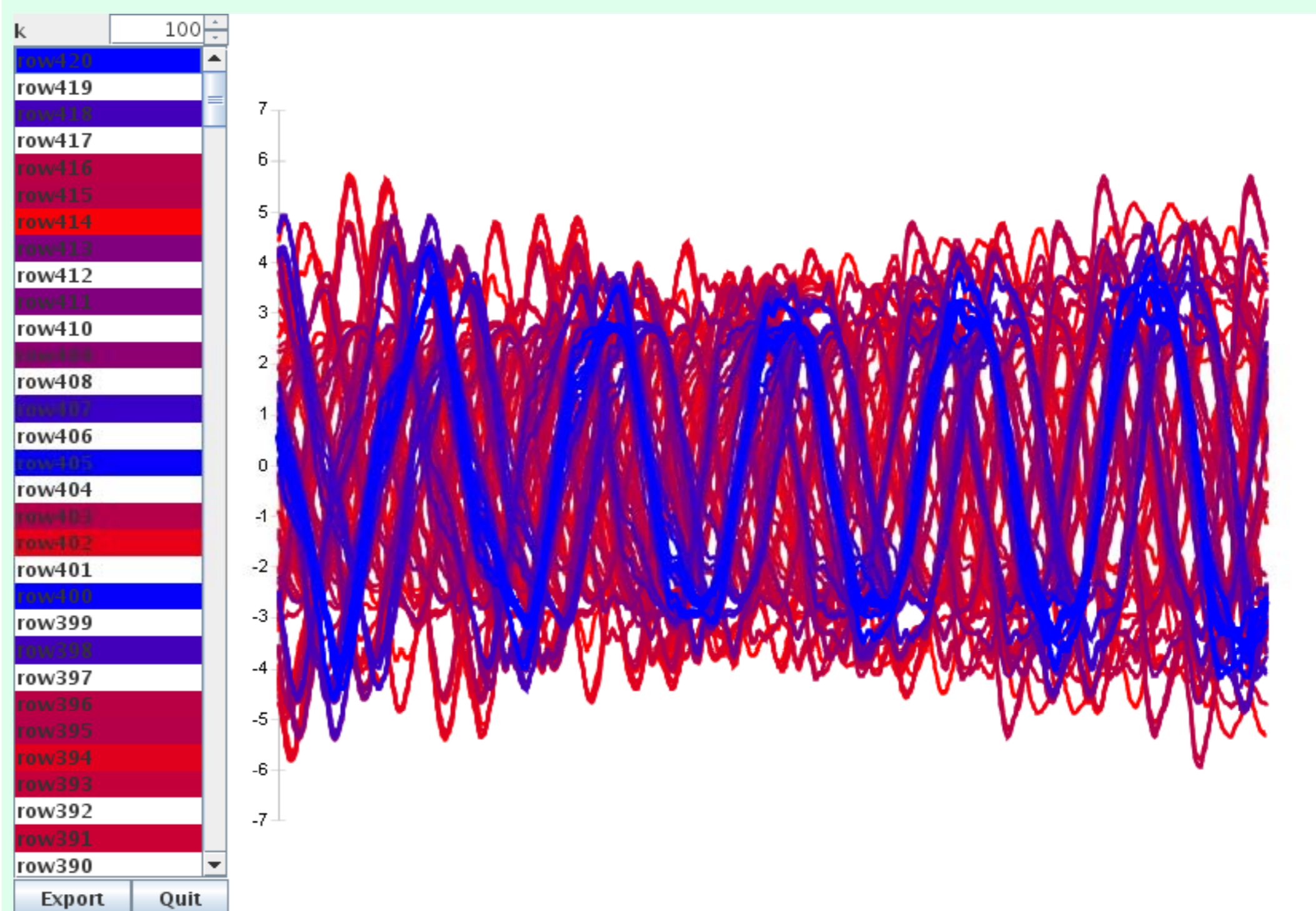
As a framework, ELKI is flexible in a sense, that it allows to read arbitrary data types (provided there is a suitable parser for your data file or adapter for your database), and that it supports the use of any distance or similarity measure appropriate for the given data type. Usually, an algorithm needs to be provided with a distance function of some sort. Thus, distance functions connect arbitrary data types to arbitrary algorithms.

### Performance of Distance Functions:

In ELKI 0.2, the performance of different distance measures can be directly assessed and visualized to enable the researcher to get a feeling for the meaning, benefits and drawbacks of a specific distance measure. In a data set of time series, a specific time series can be picked and a  $k$ -NN query can be performed for this time series within the data set for any  $k$  and any distance function. The result of the query is e.g. visualized by assigning colors of degrading similarity to the time series in the query result according to the decreasing similarity w.r.t. the given distance measure.



k-NN query using  
Euclidean Distance



k-NN query using  
LCSS Distance

### Availability:

This work is continued – find the source-code and binaries, documentation, and bug-reports via:

<http://www.dbs.ifi.lmu.de/research/KDD/ELKI/>