

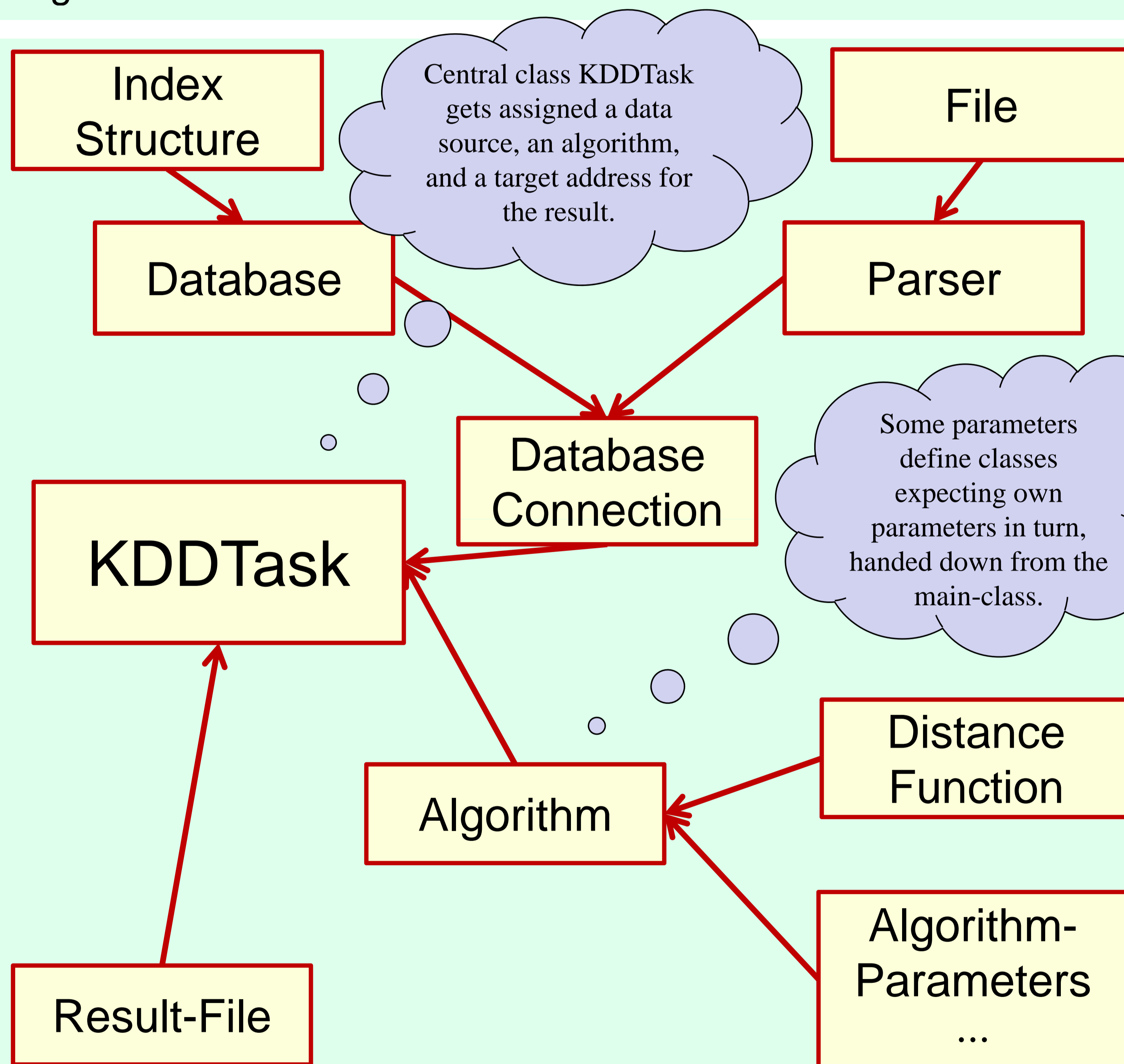
Environment for DeveLoping KDD-Applications Supported by Index-Structures

Motivation:

Data mining research leads to many algorithms for similar tasks. A fair and useful comparison of these algorithms is difficult due to several reasons:

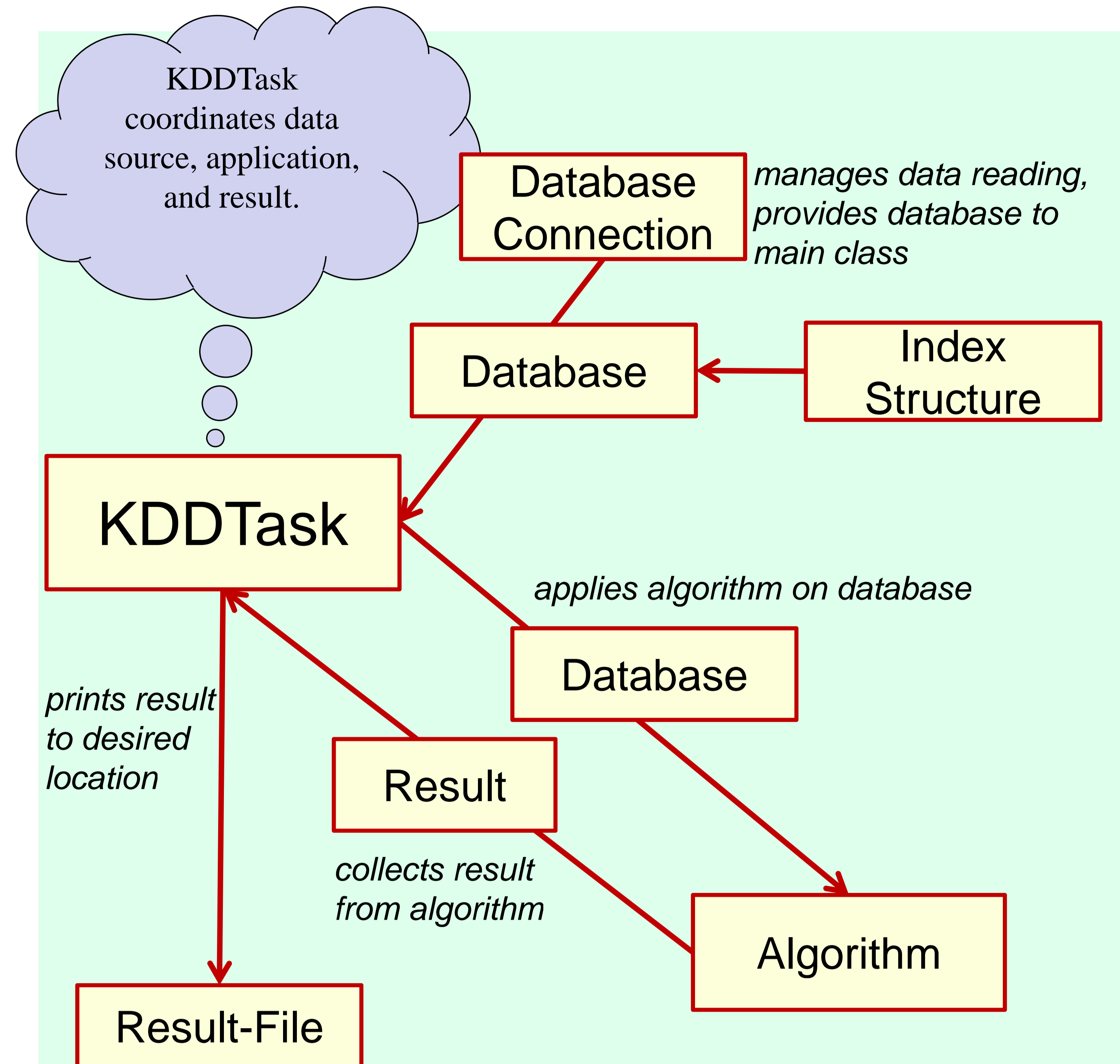
- Implementations of comparison partners are not at hand.
- If implementations of different authors are provided, an evaluation in terms of efficiency is biased to evaluate the efforts of different authors in efficient programming instead of evaluating algorithmic merits.

On the other hand, efficient data management tools like index-structures can show considerable impact on data mining tasks and are therefore useful for a broad variety of algorithms.



Idea:

In ELKI, data mining algorithms and data management tasks are separated and allow for an independent evaluation. This separation makes ELKI unique among data mining frameworks like Weka or YALE and frameworks for index structures like GiST. At the same time, ELKI is open to arbitrary data types, distance or similarity measures, or file formats. The fundamental approach is the independence of file parsers or database connections, data types, distances, distance functions, and data mining algorithms. Helper classes, e.g. for algebraic or analytic computations are available for all algorithms on equal terms.



USAGE:

```

java de.lmu.ifi.dbs.elki.KDDTask
-algorithm : <class> Classname of an algorithm (implementing
de.lmu.ifi.dbs.elki.algorithm.Algorithm -- available classes:
-->de.lmu.ifi.dbs.elki.algorithm.APRIORI
-->de.lmu.ifi.dbs.elki.algorithm.clustering.DBSCAN
-->de.lmu.ifi.dbs.elki.algorithm.clustering.DeLiClu
-->de.lmu.ifi.dbs.elki.algorithm.clustering.EM
-->de.lmu.ifi.dbs.elki.algorithm.clustering.KMeans
-->de.lmu.ifi.dbs.elki.algorithm.clustering.OPTICS
-->de.lmu.ifi.dbs.elki.algorithm.clustering.SLINK
-->de.lmu.ifi.dbs.elki.algorithm.clustering.SNNClustering
-->de.lmu.ifi.dbs.elki.algorithm.clustering.biclustering.ChengAndChurch
-->de.lmu.ifi.dbs.elki.algorithm.clustering.biclustering.FLOC
-->de.lmu.ifi.dbs.elki.algorithm.clustering.biclustering.MaPle
-->de.lmu.ifi.dbs.elki.algorithm.clustering.biclustering.PClustering
-->de.lmu.ifi.dbs.elki.algorithm.clustering.correlation.CASH
-->de.lmu.ifi.dbs.elki.algorithm.clustering.correlation.COPAC
-->de.lmu.ifi.dbs.elki.algorithm.clustering.correlation.ERIC
-->de.lmu.ifi.dbs.elki.algorithm.clustering.correlation.FourC
-->de.lmu.ifi.dbs.elki.algorithm.clustering.correlation.ORCLUS
-->de.lmu.ifi.dbs.elki.algorithm.clustering.subspace.CLIQUE
-->de.lmu.ifi.dbs.elki.algorithm.clustering.subspace.DiSH
-->de.lmu.ifi.dbs.elki.algorithm.clustering.subspace.PreDeCon
-->de.lmu.ifi.dbs.elki.algorithm.clustering.subspace.PROCLUS
-->de.lmu.ifi.dbs.elki.algorithm.clustering.subspace.SUBCLU
). Either full name to identify classpath or only classname, if its package is
de.lmu.ifi.dbs.elki.algorithm.
-dbc : <class> Classname of a database connection (implementing
de.lmu.ifi.dbs.elki.database.connection.DatabaseConnection --
available classes:
-->de.lmu.ifi.dbs.elki.database.connection.FileBasedDatabaseConnection
-->de.lmu.ifi.dbs.elki.database.connection.InputStreamDatabaseConnection
-->de.lmu.ifi.dbs.elki.database.connection.MultipleFileBasedDatabaseConnection
). Either full name to identify classpath or only classname, if its package is
de.lmu.ifi.dbs.elki.database.connection. Default:
de.lmu.ifi.dbs.elki.database.connection.FileBasedDatabaseConnection.
-description : <class> Name of a class to obtain a description - for classes
that implement
de.lmu.ifi.dbs.elki.utilities.optionhandling.Parameterizable)
-- no further processing will be performed.
-h : Flag to obtain help-message, either for the main-routine or
for any specified algorithm. Causes immediate stop of the
program.
-help : Flag to obtain help-message, either for the main-routine or
for any specified algorithm. Causes immediate stop of the
program.
-norm : <class> Classname of a normalization in order to use a
database with normalized values (implementing
de.lmu.ifi.dbs.elki.normalization.Normalization -- available
classes:
...
    
```

For supported algorithms, a short description is provided on request:

```

java -jar elki.jar -description de.lmu.ifi.dbs.elki.algorithm.clustering.correlation.ERIC
ERIC: Exploring Relationships among Correlation Clusters
Performs the DBSCAN algorithm on the data using a special distance function
taking into account correlations among attributes and builds a hierarchy
that allows multiple inheritance from the correlation clustering result.

Reference: E. Achtert, C. Böhm, H.-P. Kriegel, P. Kröger, and A. Zimek: On
Exploring Complex Relationships of Correlation Clusters. In Proc. 19th
International Conference on Scientific and Statistical Database Management
(SSDBM 2007), Banff, Canada, 2007
    
```

This work is continued – find the source-code and binaries, documentation, and bug-reports via:
<http://www.dbs.ifi.lmu.de/research/KDD/ELKI/>