Goal

Build association rule from binary dataset.
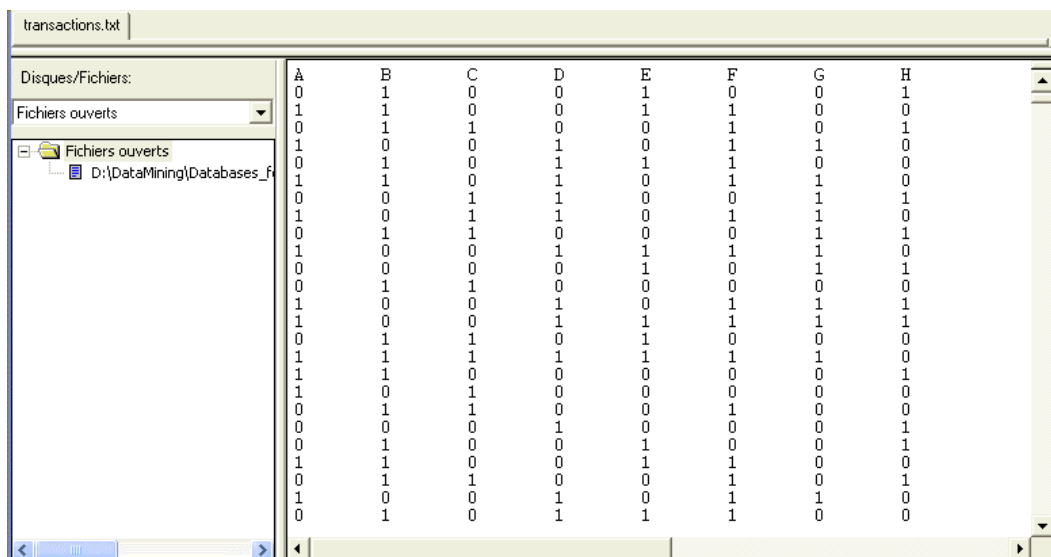
Dataset

Association rules can be built from attribute-value dataset, which is re-coded as binary table. In certain cases, we have a transaction dataset, which is already a binary table. It is not necessary to re-code this one. How to handle this kind of dataset?

TANAGRA can handle only attribute-value dataset: the absence of one item in a transaction is coded as 0, other values are seeing as a presence (1 value if the file is correctly built).

# Build association rule from a binary dataset

Dataset

This file comes from the web[1]; they represent the presence of 8 items in 10000 transactions. You must build the dataset as a tabular file where "0" is an absence of an item for one transaction, "1" the presence. Use a text file format and tab separator.

You can import the dataset into TANAGRA, all attributes are seeing as continuous.

[1] http://www2.cs.uregina.ca/~dbd/cs831/notes/itemsets/itemset_prog1.html
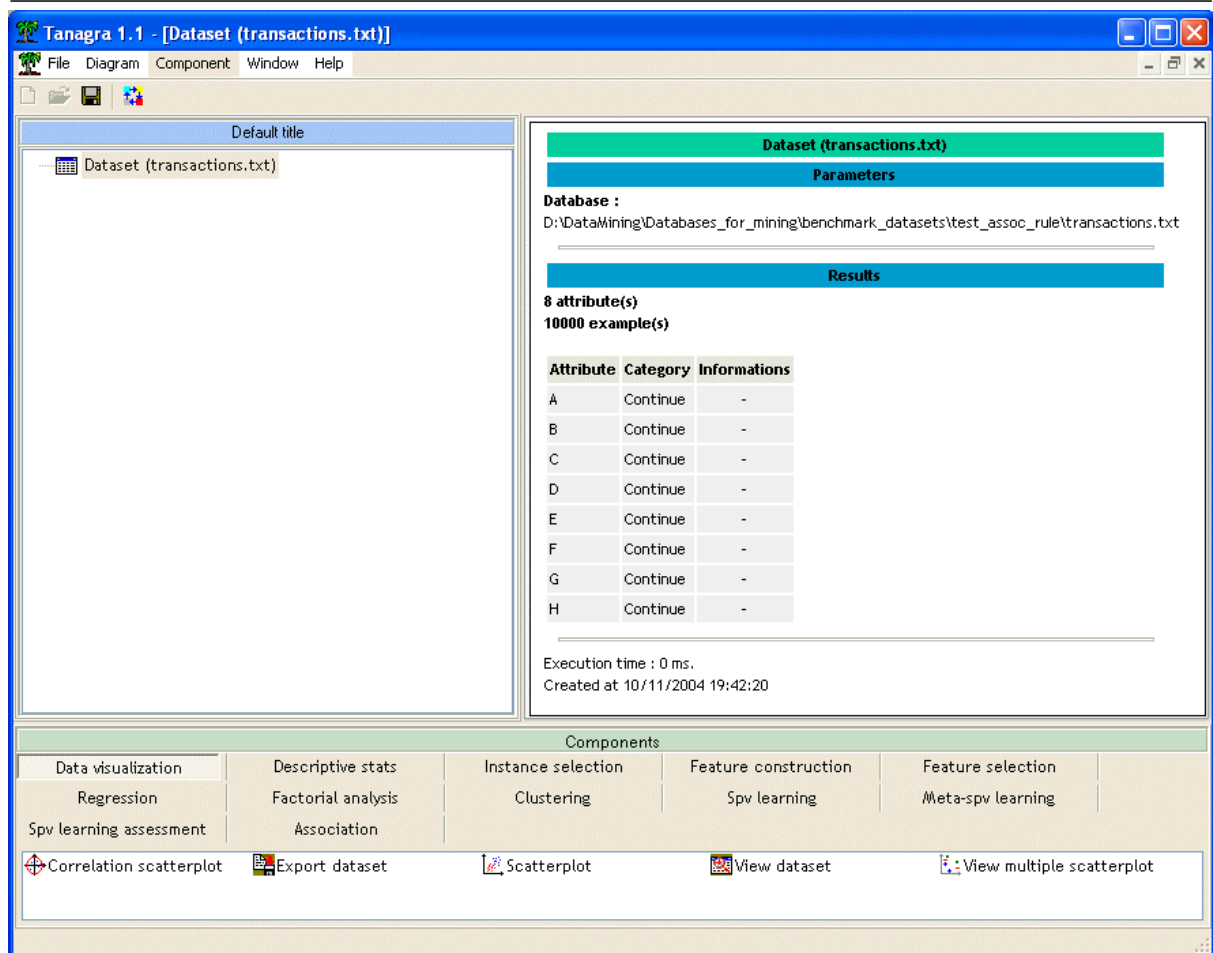
Define attributes status

The next step is to define attributes status. TANAGRA accepts two kind of selection:

1. All attributes are discrete. In this case, it re-code the dataset, each attribute-value pair becomes a binary column.

2. All attributes are continuous. In this case, each attribute becomes a binary column, "0" is seeing as "false" (absence), and the others ("1" if the file is correctly built) as "true" (presence).
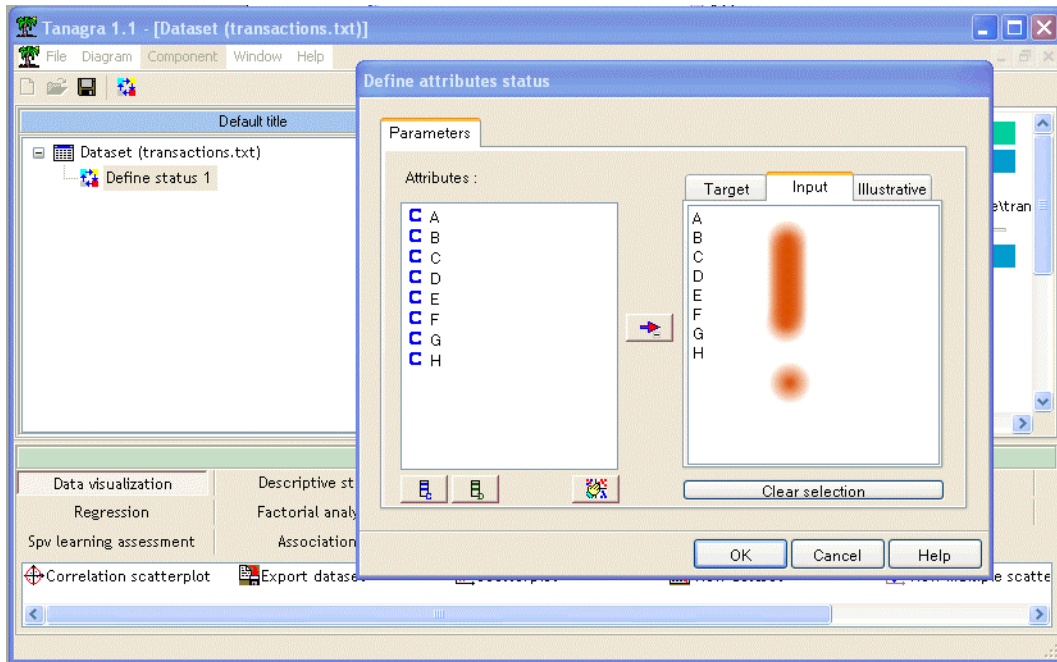
Analysis

You can launch the A PRIORI component and run analysis. Only "positive" rules are extracted.