Viscovery® SOMine 7 – Data Sheet

Extension Module | Enterprise Data

Connectivity with enterprise environments and support of voluminous data

Main Functions and Features

General characteristics

- **Enterprise Data** is an extension module of the Viscovery SOMine Suite. It requires the **Visual Clusters** core module and can optionally be combined with other modules of the suite.
- This module provides features to connect with enterprise data sources, to handle very high-dimensional data and to support the work process in complex analysis projects.
- The module enables processing of data sets with an unlimited number of records and variables in the complete suite.

Database connectivity

- ODBC/OLEDB interface for accessing all common database systems, such as Oracle and SQL Server
- Import data in the "Import Data" step from database tables and views
- Export of classification and scoring results and of Viscovery data marts to a database table

Additional data interfaces

- Import and export of SPSS files (*.sav) and of Viscovery XML files (*.xml)
- Import data from comma-separated text files (*.csv)
- Import of original SOM_PAK model files (*.cod)

Export of Viscovery data marts

- Export of data marts created in the **Preprocess** workflow to flat files or database tables
- Optional use of defined replacements and transformations for exported attributes
- Optional replacement of missing values by self-organizing map (SOM) node values for exported attributes
- Optional export of nominal variables as dichotomous values

Join functionality

- Joining of data files and/or database tables directly in the "Import Data" step
- Join of data from multiple data sources on arbitrary columns ("left outer join")
- Highlighting of possible join attributes and name conflicts

Handling of voluminous and high-dimensional data

- Random sampling of large data sets to smaller sizes in Viscovery data marts
- Application of Benjamini-Hochberg multiple-testing correction to adjust statistical confidence measures if a large number of attributes are used to characterize clusters or to compare groups

Preprocessing protocol

- Import and export of preprocessing settings of the **Preprocess** workflow from/to a spreadsheet document
- Compact overview of all parameters ("preprocessing protocol"), which can be edited outside the Viscovery system
- Organization into a table with columns of predefined content, which allows specification of variable definitions, setting allowed variable ranges, defining the treatment of outliers and irregularities, and addition of descriptions, supporting a clear specification of preprocessing options