

<b>Projekt LISp-Miner</b>		Číslo dokumentu: 119	Kategorie: D
Odpovědná osoba: Milan Šimůnek	Email: <a href="mailto:simunek@vse.cz">simunek@vse.cz</a>	Poslední změna: 4. 6. 2011	Úroveň: I
Poznámka:			

# Export pro SEWEBAR

## popis implementace

### Abstrakt

Tento dokument obsahuje popisuje modul *LM SwblImporter* pro import dat ve formátu PMML z prostředí SEWEBAR.

## Obsah

<b>Abstrakt.....</b>	<b>1</b>
<b>Úvod .....</b>	<b>2</b>
<b>Popis implementace importu .....</b>	<b>2</b>
<b>Chybová hlášení.....</b>	<b>2</b>
Obecná .....	2
Matrix .....	2
DataDictionary .....	2
TransformationDictionary .....	3
4ftTask .....	3
4ftZadaniCedentu .....	3
4ftZadaniDilcihoCedentu .....	3
4ftZadaniLiteralu .....	3
<b>Úprava textových hodnot ze schématu PMML .....</b>	<b>4</b>
<b>LMSwblImporter .....</b>	<b>4</b>

## Úvod

– zde je třeba doplnit text

## Popis implementace importu

– zde je třeba doplnit text

Základní importované sekce:

- DataDictionary
- TransformationDictionary
- 4ftTask – guha:AssociationModel

Importér se pokouší načíst všechny sekce, které se v předaném PMML souboru vyskytují. Libovolná ze sekcí může v PMML souboru chybět. Potom však musí být všechny datové struktury odkazované z jiné sekce již vytvořené v metabázi.

Příklady:

- import pouze 4ftTask (když už je v metabázi nadefinován DataDictionary a TransformationDictionary)
- import pouze TransformationDictionary
- import DataDictionary (zejména nastavení primárního klíče) + TransformationDictionary
- import úplně všeho

– zde je třeba doplnit text

## Chybová hlášení

Má-li být LM SwbImporter spuštěn na pozadí, bez možnosti interakce s uživatelem, je třeba zadat parametry /Quiet a /NoProgress. V takovém případě jsou veškerá chybová hlášení vypisována pouze do log-souboru (defaultně \_AppLog.dat ve stejném adresáři jako LMSwbExporer.ext, je však možné změnit parametrem /AppLog).

Pro ladící účely není vhodné tichý režim používat a je lepší nechat zobrazovat jak chybová hlášení, tak i průběh generování.

Následuje přehled chybových hlášení a jejich identifikací (řazeno podle typu importovaných dat).

### Obecná

- (General.001) Error reading the import file <file\_name>
- (General.002) Error reading the alias file <file\_name>

### Matrix

- (Matrix.001) SwbImport.LookupMatrix - XML Header tag not found
- (Matrix.002) SwbImport.LookupMatrix - XML tag 'Header' extension 'dataset' not found
- (Matrix.003) SwbImport.LookupMatrix - data matrix <matrix\_name> not found in the metabase
- (Matrix.004) SwbImport.LookupMatrix - data matrix <matrix\_name> cannot be initialised
- (Matrix.005) SwbImport.LookupMatrix - data matrix <matrix\_name> cannot be used
- (Matrix.006) SwbImport.LookupMatrix - no primary key defined for the data matrix <matrix\_name>
- (Matrix.007) SwbImport.LookupMatrix - duplicities found in the primary key for data matrix <matrix\_name>

### DataDictionary

- (DataDictionary.001) SwbImport.DataDictionary - no DataDictionary found

- (DataDictionary.002) SwbImport.DataDictionary - syntax error reading PrimaryKeyPosition
- (DataDictionary.003) SwbImport.DataDictionary - matrix attribute '%s' not found in the matrix <matrix\_name>
- (DataDictionary.004) SwbImport.DataDictionary - derived attribute '%s' could not be part of the primary key in matrix <matrix\_name>
- (DataDictionary.005) SwbImport.DataDictionary - column <column\_name> not found in the table <table\_name>

### TransformationDictionary

- (TransformationDictionary.001) SwbImport.TransformationDictionary - no DataDictionary found
- (TransformationDictionary.002) SwbImport.TransformationDictionary - no TransformationDictionary found
- (TransformationDictionary.003) SwbImport.Attribute - no attribute name found
- (TransformationDictionary.004) SwbImport.Attribute - attribute definition is empty
- (TransformationDictionary.005) SwbImport.Attribute - database column name not found
- (TransformationDictionary.006) SwbImport.Attribute - datafield name not found
- (TransformationDictionary.007) SwbImport.Attribute - column <column\_name> not found in the data matrix <matrix\_name>
- (TransformationDictionary.008) SwbImport.Category.Enum - field and/or column name not found
- (TransformationDictionary.009) SwbImport.Category.Enum - error parsing category value
- (TransformationDictionary.010) SwbImport.Category.Enum - category <category\_name> already exists
- (TransformationDictionary.011) SwbImport.Category.Interval - interval definition not found
- (TransformationDictionary.012) SwbImport.Category.Interval - error parsing interval boundaries values
- (TransformationDictionary.013) SwbImport.Category.Interval - category <category\_name> already exists

### 4ftTask

- (4ftTask.001) SwbImport.4ftTask - no 4ftTask found
- (4ftTask.002) SwbImport.4ftTask - <TaskSetting> tag not found
- (4ftTask.003) SwbImport.4ftTask - task with name <task\_name> already exists
- (4ftTask.004) SwbImport.4ftTask.FTQuantifier - unknown quantifier name
- (4ftTask.005) SwbImport.4ftTask.FTQuantifier - could not read the quantifier threshold value
- (4ftTask.006) SwbImport.4ftTask.FTQuantifier - unknown quantifiers value compare type
- (4ftTask.007) SwbImport.4ftTask.FTQuantifier - could not read the quantifier order value

### 4ftZadaniCedentu

- (FTCedentBagD.001) SwbImport.FTCedentBagD - <DBASettings> tag not found
- (FTCedentBagD.002) SwbImport.FTCedentBagD - <BBASettings> tag not found
- (FTCedentBagD.003) SwbImport.FTCedentBagD - <DBASetting> tag <cedentbag\_id> not found

### 4ftZadaniDilcihoCedentu

- (FTCedentD.001) SwbImport.FTCedentD - <DBASetting> tag <cedent\_id> not found

### 4ftZadaniLiteralu

- (FTLiteralD.001) SwbImport.FTLiteralD - <DBASetting> tag <literal\_id> not found
- (FTLiteralD.002) SwbImport.FTLiteralD - <BBASetting> tag <literal\_id> not found
- (FTLiteralD.003) SwbImport.FTLiteralD - attribute <attribute\_name> not found
- (FTLiteralD.004) SwbImport.FTLiteralD - <Coefficient> tag not found

- (FTLiteralD.005) SwbImport.FTLiteralD - unknown coefficient type
- (FTLiteralD.006) SwbImport.FTLiteralD - no category defined for coefficient type of 'One category'
- (FTLiteralD.006) SwbImport.FTLiteralD - category <category\_name> not found in <attribute\_name> for coefficient type of 'One category'

## Úprava textových hodnot ze schématu PMML

Textové hodnoty v číselnících (názvy kvantifikátorů, typů koeficientů, datové typy sloupců, kardinalita) používané v systému LISp-Miner se v některých případech liší od těch v používaném PMML schématu.

Při importu PMML je proto nutné zajistit jejich nahrazení (v opačném směru než při exportu do PMML). Mapovací tabulka je definována v souboru `LM.PMML.Alias.txt`, stejně jako pro `LM SwbExporter`. Nahrazování se však provádí v opačném směru (hodnoty v druhém sloupci jsou nahrazeny hodnotami v prvním sloupci). To znamená, že pro import i export je možné použít stejný mapovací soubor.

## LMSwbImporter

V systému LM přibyl nový modul `LM SwbImporter`, který je možné spouštět samostatně. Stačí předat název ODBC datového zdroje a název importovaného PMML souboru. Datové struktury se poté naimportují do metabáze.

Syntaxe:

<code>LMSwbImporter.exe</code>	name of the executable file
<code>/DSN:&lt;data-source-name&gt;</code>	data source name of the metabase (if the data source name contains spaces, the whole /DSN parameter has to be enclosed in quotations mark, e.g. "/DSN:LM Barbora MB")
<code>/Input:&lt;pmml_file&gt;</code>	path and name to the input PMML file
<code>/Alias:&lt;alias_file&gt;</code>	(O) path and name to the alias file
<code>/NoCheckPrimaryKeyUnique</code>	(O) skip test of unique values in the database table primary key
<code>/Quiet</code>	(O) for run in the background – no information dialog boxes shown and no user's interaction is expected (errors reported to the <code>_AppLog.dat</code> only)
<code>/NoProgress</code>	(O) for run in the background -- no progress dialog is displayed
<code>/AppLog:&lt;log_file&gt;</code>	(O) alternative path and file name for logging

(O) znamená „nepovinný parametr“.