ifis
Institut für Informationssysteme
Technische Universität Braunschweig

Exercises for DW & DM

Sheet 3

No need to handout solutions for this homework. The following exercises are meant to help you familiarize with tools for simple data cleaning as well as OLAP operations.

Exercise 1

1. Install Eobjects Data Cleaner (http://datacleaner.eobjects.org/downloads). Perform the following tasks, by using the sample database provided with the software. As a data source, select the Data Cleaner example database named orderdb. For analysis purposes, choose the CUSTOMERS table, and as data subset the CONTACTLASTNAME and CONTACTFIRSTNAME attributes.

   a. Add a string pattern match filter and configure you own string pattern. Compose a regular expression which validates only strings which contain letters only (no spaces or other characters than letters), start with only one capital letter, and continue with at least one, up to 20 small letters. See examples in Annex 1.

   b. Select the regular expression from 2.a to identify the invalid entries for the CONTACTLASTNAME and CONTACTFIRSTNAME attributes. Write down the lastname and firstname of the clients which did not pass the validation.

   c. Give three examples (of different patterns) of strings which pass the validation of the following regular expression, and one that doesn't:

   $$(\backslash+\backslash d\{1,2\})?((\backslash(\backslash d\{1,4\}\backslash))|(\backslash d)\{3,5\}[-/]?)((\backslash d)\{1,5\})$$

Exercise 2

Install Mondrian a MOLAP solution from Pentaho. You can download the code from http://sourceforge.net/projects/pentaho/files/Business%20Intelligence%20Server/4.5.0-stable/ After you have decompressed it, start the OLAP server (which comprises also a http server - Tomcat) by running the ..\mondrian\biserver-ce\start-pentaho.bat command file. After the server is up, open a browser window and type http://localhost:8080/pentaho. Check that no other service is running on the 8080 port. If so, then search for the server URL in the console window. At login go for the evaluation login, and choose to login with the Business User. If you have gotten so far, you are now able to perform some operations on the example data. Choose the Analysis View from the File\New menu. Use the default set-

tings for the data source. Just play around/experiment. You can now test typical OLAP operations like drill-down, drill-across, pivot, slice, dice, etc. Pressing on the MDX toolbar icon opens a console where you can test some MDX queries.

Annex 1

String	Evaluation
A	Bad
Aa	Good
AA	Bad
aa	Bad
A1	Very bad
Thomas	Good
Thomas Mann	Bad