

## Exercises for DW & DM

### Sheet 8 (until 07.01.2011)

You may hand in your solutions into the mailbox at the IFIS floor (Mühlenpfordtstraße 23, 2nd floor). For the ITIS students only, please send your solutions to [silviu@ifis.cs.tu-bs.de](mailto:silviu@ifis.cs.tu-bs.de). The deadline is Friday, after the next lecture (date is also mentioned above). You may answer in either German or English. **You are encouraged to work in teams of 2 students** (not more than 2), and send your solution as a team. Please mention the **name of both students** together with the corresponding **inmatriculation numbers**.

#### Exercise 1 (9P)

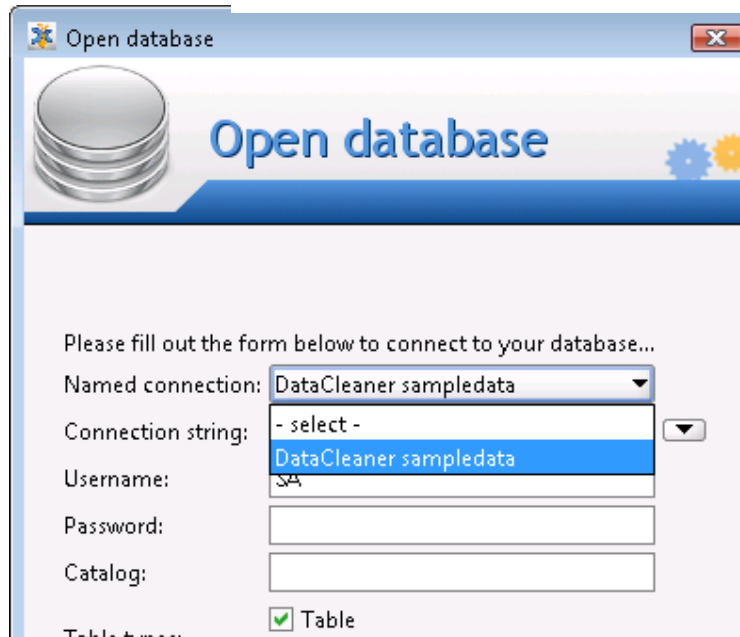
1. Install Eobjects Data Cleaner (<http://datacleaner.eobjects.org/downloads>). Perform the following tasks, by using the sample database provided with the software (by choosing it from the drop down menu as observed in the Annex1)
  - a. Compose a regular expression which validates only strings which 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 2. (3P)
  - b. Use the regular expression from 2.a, and create a validation task, add as validation rule a “regex validation”, choose as data selections the CUSTOMER table, and as data subset the CONTACTLASTNAME and CONTACTFIRSTNAME attributes. Write the lastname and firstname of the clients which did not pass the validation. (If there are too many you did something wrong!!!) (3P)
  - c. Give three examples (of different patterns) of strings which pass the validation of the following regular expression, and one that doesn't:

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

#### Exercise 2 (15P)

Simulate the functionality of the Multiple Minimum Supports mining algorithm on the transactions provided in Annex 3, presenting each of the 2 steps, as well as the initialization,  $k=2$  and generalization phases for step 1. Minimum support values are also provided in the Annex 3.  $\phi = 20\%$  and  $\text{minconf} = 60\%$ . (15P)

## Annex 1



Open database

Please fill out the form below to connect to your database...

Named connection: DataCleaner sampledata

Connection string: - select -

Username: sa

Password:

Catalog:

Table

## Annex 2

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

### Annex 3

		Transactions
		1, 4, 6
		1
		1, 5, 6
		1, 6
		4, 6
		1, 2, 3, 5
		1, 2, 3, 5
		6
		1
		1, 6
Item	MIS %	
1	70	
2	17	
3	15	
4	30	
5	30	
6	35	