Business Intelligence Lab

Written test 22/7/2015

Deliver solutions within 4 h

Notice: use your own SQL Server credentials (the lbi account is disabled)

	$B = v_B$	$B \neq v_B$	
$A = v_A$	a	b	n_1
$A \neq v_A$	c	d	n_2
	m_1	m_2	n

$$a = |\{\sigma \in \mathcal{R} \mid \sigma[A] = v_A, \sigma[B] = v_B\}| \quad b = |\{\sigma \in \mathcal{R} \mid \sigma[A] = v_A, \sigma[B] \neq v_B\}|$$
$$c = |\{\sigma \in \mathcal{R} \mid \sigma[A] \neq v_A, \sigma[B] = v_B\}| \quad d = |\{\sigma \in \mathcal{R} \mid \sigma[A] \neq v_A, \sigma[B] \neq v_B\}|$$
$$n_1 = a + b \quad n_2 = c + d \quad m_1 = a + c \quad m_2 = b + d \quad n = n_1 + n_2 = m_1 + m_2$$

Figure 1: Contingency table.

Exercise 1 (8 pts). Let $\mathcal{R}$ be a relation, including two attributes A and B. A contingency table for values v_A and v_B is shown in Fig. 1. Develop a Java program CTable.java using JDBC that given parameters $\mathcal{R}$, A and B, outputs to a CSV file a row v_A, v_B, a, b, c, d for every contingency table at the variation of v_A and v_B.

Test the program on the table `census` of the `lbi` database. Only `SELECT` without `WHERE/ORDER BY/GROUP BY` statements from a single table is allowed.

What to deliver: CTable.java, myJDBCdef.props (with only the parameters needed for a test of the program).

Exercise 2 (8 pts). Develop a SSIS package reading table `census` on the `lbi` database, solving Exercise 1 for A being `sex` and B being `class`.

What to deliver: SSDT solution.

Exercise 3 (8 pts). Write a single MDX query that solves Exercise 1 on the `ruggieri_foodmart` for for A being `product family` and B being `customer gender`.

What to deliver: MDX query and a brief comment about it, PowerPoint file with the screenshot of the query result.

Exercise 4 (8 pts). Generalize contingency tables to subsets of $\mathcal{R}$ such that $C = v_C$ for a third attribute C. Solve the generalized problem of Exercise 1 using association rules as building blocks. Give pseudo-code of the solution using Weka API calls.

What to deliver: PowerPoint file with answer, pseudo-code as a Java program.

How to deliver: send an e-mail with a single <your surname>.zip file attached to ruggieri@di.unipi.it, including your name, surname, student ID, and computer IP address (http://www.whatismyip.com).

Results and oral exam. Results will be emailed to the students shortly, including the date and time for those who are admitted to the oral exam.