Business Intelligence Lab

Written exam 18/7/2013

Deliver exercises within 4 h from start time

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

Exercise 1 (8 pts). Consider a table with a binary column C with domain $\{c_1, c_2\}$, having p_1 percent of tuples with value $C = c_1$. A table with (at least) column A, B, and the binary column C above is called t-close if for every possible values a of A and b of B, either there is no tuple such that A=a and B=b or in the set of such tuples the proportion p with $C = c_1$ is such that $p \in [p_1 - t, p_1 + t]$. Write a Java program TClose.java using JDBC that output 'yes' if a table is t-close, and 'no' otherwise. Table name, t, and columns A, B, and C are passed as parameters. The usage of any SQL query is permitted.

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

Exercise 2 (8 pts). Develop a SSIS package reading `sales_fact_1998` from the *foodmart* database, with the purpose of writing on a text file all triples (store_id, product_category, indicator) that are not 30-close. The column indicator has value "1" if the total sales of product_category at store_id is above $ 500, and "0" otherwise. The usage of SQL queries to perform computation at server side is not permitted. All the work must be done by the SSIS package.

What to deliver: BIDS/SSDT solution.

Exercise 3 (6 pts). The Jaccard index of the sales in March 1998 of a product category A is m/n where m is the number of baskets of March 1998 including A, and n is the number of baskets including A or sold in March 1998. Write a MDX query to answer the following question on the Sales cube of the `ruggieri_foodmart` OLAP database:

- Jaccard index in March 1998 of all product categories.

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

Exercise 4 (10 pts). The Jaccard index of two items A and B is:

$$J_{A,B} = \frac{\#\{t | A \in t, B \in t\}}{\#\{t | A \in t \text{ or } B \in t\}}$$

Using association rules in Weka, find (at least 3) pairs A, B of items with $J_{A,B} \geq 0.8$ on the dataset provided by the teacher. Justify your approach.

What to deliver: a PowerPoint file with screenshots of Weka explorer, description of the steps of the analysis.

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

Results and oral exam. Results will be published on-line by tomorrow morning. Oral exams will start tomorrow afternoon at teacher office.