Business Intelligence Lab

Written exam 27/6/2013

Deliver exercises within 3 h from start time

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

Exercise 1 (8 pts). A table with (at least) column A, B, and C is called k-anonymous if for every possible value a of A, b of B and c of C either there is no tuple such that A=a and B=b and C=c, or there are at least k tuples such that A=a and B=b and C=c. Write a Java program KAnonimity.java using JDBC that output 'yes' if a table is k-anonymous, and 'no' otherwise. Table name, k, and columns A, B, and C are passed as parameters. The usage of any SQL query is permitted.

What to deliver: KAnonimity.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, the_month) that are not 100-anonymous. 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). Write a MDX query to answer the following question on the Sales cube of the `ruggieri_foodmart` OLAP database:

- Total sales for each store and product category and month that are 100-anonymous triples.

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). Using association rules in Weka, find (at least 3) triples "A=a and B=b and C=c" that make a dataset non 5-anonymous. Then make some pre-processing filtering to make the triples 5-anonymous (removing rows is not permitted). Apply the approach on the dataset provided by the teacher.

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.