Bases de données: Exam

Pierre Senellart (pierre.senellart@ens.fr)

May 30, 2018

The sole authorized documents are 2 A4 two-sided sheets (4 pages), with the content of your choice, readable with the unaided eye. This exam lasts two hours and comprises a single problem of 10 questions, for a total of 20 points.

An oenology society wishes to gather all data about its wines, its members, and their activity in a database.

Every wine is characterized by a *label* (the name of the wine, e.g., Château Margaux Grand Vin), a *vineyard* (e.g., Château Margaux), a *region* (e.g., Bordeaux), a *vintage* (the year of production of the wine), a *type* (red, rosé, white), and a *price* (in euro cents) corresponding to the amount billed by the society for every tasting of this wine. Two wines of different vineyards may have the same label. Within a vineyard, the label uniquely identifies a wine type, but whose vintage may vary. We will assume that the price is fixed for a given vineyard, label, and vintage.

Every society member has a last name, a first name, and can specify a favorite wine type (red, rosé, white), for which he or she obtains a 10% discount for every tasting. Each society member, except its creator, must be invited by an existing member, called his or her "godfather".

The society also records the date every member is tasting every wine, so as to bill these tastings to its members. At a given date, a taster can taste several different wines, and several times the same wine.

1. (2 points) Propose an entity–relation schema for this information system.
2. (2 points) Deduce a relational schema and express it with SQL `CREATE TABLE` statements; include in these statements primary keys (`PRIMARY KEY`), candidate keys (`UNIQUE`), foreign keys (`REFERENCES`), `CHECK` constraints, and specify which attributes have to be `NOT NULL`.
3. (1 point) What logical constraints of the information system are not expressed by these SQL constraints? Express them as first-order formulas if possible, and if possible as TGDs or EGDs.
4. (3 points) Give a query computing the list of members who are godfathers of a member having at least one wine tasting in the month of May 2018:
   a) in the relational algebra;
   b) in the relational calculus;
   c) in SQL.
5. (2 points) Propose two query plans for this query, and discuss of the most efficient plan, in terms of the sizes of the different relations.
6. (1 point) Give an upper bound on the asymptotic complexity of this query, in data complexity, as precise as possible.

7. (2 points) Every month, members receive a discount of 1 euro on their bill for every member they are a godfather of and who have realized at least one tasting during the month. The final bill cannot become negative.

   Write in SQL a query computing the amount due for the month of May 2018 by every member, taking into account their tastings during that month, and possible discounts due to favorite wine types and godfathering.

8. (4 points) One would like to check there is no cycle in the godfather relation between members, i.e., that there is no two distinct members x_0 and x_n such that there exists a sequence of members $x_1, \ldots, x_{n-1}$ with x_0 godfather of x_1, x_1 godfather of x_2, $\ldots$, x_{n-1} godfather of x_n et x_n godfather of x_0. Write a Boolean query testing the non-existence of such a cycle:

   a) in Datalog with negation;

   b) in fixpoint first-order logic (inflationary or non-inflationary, up to you);

   c) in second-order logic;

   d) in SQL.

9. (1.5 point) Data input (adding a new member, inputting a wine tasting, issuing a bill) is made by the society heads through a Web application. Discuss potential race condition issues that may arise in such an application, and how to solve them.

10. (1.5 point) All members have access to a Web form allowing them to search a wine by label or vineyard. They have, on the other hand, no access to the list of members. Assuming the site is sensitive to SQL injection attacks, present a potential attack that would allow to find the list of members by using the Web form for searching wines.

This exam is a reference to a schema traditionnally used for database teaching in France, due to Georges Gardarin. Excessive drinking may damage your health.