CS–A1150 Databases, 2020

Exercise Round 1 (24.2.–28.2.2020), deadline for submission 29.2.2020 in the morning.

The problems may be solved in the groups of 1–3 students. Submit your solutions to A+-system (the link is given in course MyCourses page).

Problems 1–13: Submit the solution to each of these problems separately, but each group submits only one common solution.

Problems 14–16: Submit the solutions to these problems as one pdf-file in A+-system (each group submits only one common solution). Although the problems are given in English, you may write your solutions in Finnish, English or Swedish.

1. (1 p.) Consider the following relations:

   ```
   Students(ID, name, program, year)
   Courses(code, name, credits)
   Grades(studentID, courseCode, date, grade)
   ```

   The `Students` relation gives the student ID, the name, the study program and the starting year of a student. The `Courses` relation gives the course code, the name and the number of credits of a course. The `Grades` relation gives the ID of the student who has completed the course, the code of the completed course and the date and the grade. We assume that each student can complete a certain course only once and only information on accepted grades are stored in relation `Grades`.

   Write an expression of relational algebra to answer the following query: find the code and name of all courses whose number of credits is 4.

2. (1 p.) Using the database schema in Problem 1, write an expression of relational algebra to answer the following query: Find the student ID and name of all students who have started studies in Year 2019 and whose program is TIK.

3. (1 p.) Using the database schema in Problem 1, write an expression of relational algebra to answer the following query: Find the student ID and name of all students who have completed the course with code CS-A1111.

4. (1 p.) Using the database schema in Problem 1, write an expression of relational algebra to answer the following query: Find the code and name of all courses whose number of credits is 6 and which have been completed by at least one student with grade 4 or 5.

5. (1 p.) Using the database schema in Problem 1, write an expression of relational algebra to answer the following query: Find the students who have started their studies in 2019 in program TIK and who have completed the course with code CS-A1110. The result must contain the ID and the name of the student and his/her grade at this course.

6. (1 p.) Using the database schema in Problem 1, write an expression of relational algebra to answer the following query: Find the ID and name of all students whose study program is TIK and who have completed any course whose number of credits is at least 6.

7. (1 p.) Using the database schema in Problem 1, write an expression of relational algebra to answer the following query: Find the code and name of all courses which has been completed by at least one student whose study program is TFM and who has started his/her studies in 2019.

8. (2 p.) Consider a database schema which consists of four relations, whose schemas are:

```
Product(model, maker, type)
PC(model, speed, ram, hd, price)
Laptop(model, speed, ram, hd, screen, price)
Printer(model, color, type, price)
```

The `Product` relation gives the model number, the manufacturer and the type (PC, laptop or printer) of various products. We assume for convenience that model numbers are unique over all manufactures and product types. The `PC` relation gives for each model number that is a PC the speed (of the processor, in gigahertz), the amount of RAM (in megabytes), the size of the hard disk (in gigabytes), and the price. The `Laptop` relation is similar, except that the screen size (in inches) is also included. The `Printer` relation records for each printer model whether the printer produces color output (1, if so, and 0 otherwise), the printer type (laser, ink-jet, or other), and the price.

Write expressions of relational algebra to answer the following query: Find the model number, price and manufacturer of all PC's and Laptops. Note that the result has to contain both PC and Laptop models.

9. (2 p.) Using the database schema in Problem 8, write an expression of relational algebra to answer the following query: Find those manufactures who make both PC's and laptops whose price is less than 800 euros.

10. (2 p.) Using the database schema in Problem 8, write an expression of relational algebra to answer the following query: Find those manufactures who make ink-jet printers, but no laser printers.

11. (2 p.) Using the database schema in Problem 8, write an expression of relational algebra to answer the following query: Find those pairs of PC models that have both the same amount of ram and the same hard disk size. A pair should be listed only once; for example, list (i, j) but not (j, i). For each pair, the smaller product number must be given first.

12. (3 p.) Using the database schema in Problem 8, write an expression of relational algebra to answer the following query: Find those manufactures who make only one model of ink-jet printers (not two or more models).

13. (3 p.) Using the database schema in Problem 1, write an expression of relational algebra to answer the following query: Find all tuples in `Grades` that are broken, i.e. the value of the attribute *studentID* does not appear in relation `Students` or / and the value of the attribute *courseCode* does not appear in relation `Courses`.

14. (3 p.) What is the difference between the natural join $S \bowtie R$ and the theta-join $S \bowtie_C R$ where the condition C is that $S.A = R.A$ for each attribute A appearing in the schemas of both S and R? For example, if the relations are $S(A, B, C, D)$ and $R(C, D, E)$, what is the difference between $S \bowtie R$ and $S \bowtie_{S.C=R.C \texttt{ AND } S.D=R.D} R$?

15. (8 p.) Suppose relations R and S have n and m tuples, respectively, and $n \neq m$. Give the minimum and maximum number of tuples that the results of the following expressions can have. Give also a short explanation for each result (like "if every tuple of R is equal to some tuple of S").

    a) $R \bowtie S$, assuming that the schemas of the relations have at least one common attribute.

    b) $R \cup S$, assuming that the relations have schemas of identical sets of attributes.

    c) $\sigma_C(R) \times S$, for some condition C.

    d) $\pi_L(R) - S$, for some list of attributes L such that L is identical to the attribute list of S.

16. (8 p.) The semijoin of relations R and S, denoted $R \ltimes S$, is the set of tuples of R that agree with at least one tuple of S on all attributes that are common to the schemas of R and S. Note that the schema of $R \ltimes S$ is the same as the schema of R. Give three different expressions of relational algebra that are equivalent to $R \ltimes S$. You may assume that the relational schemas of R and S are $R(A_1, A_2, \ldots, A_n, B_1, B_2, \ldots, B_m)$ and $S(A_1, A_2, \ldots, A_n, C_1, C_2, \ldots, C_k)$ where the attributes $A_1, A_2, \ldots A_n$ are common for both R and S, the attributes $B_1, B_2, \ldots, B_m$ do not appear in the schema of S, and the attributes $C_1, C_2, \ldots, C_k$ do not appear in the schema of R.