Chair of Applied Computer Science III
Prof. Dr. Guido Moerkotte
Email: moerkotte@uni-mannheim.de

Daniel Flachs
Email: daniel.flachs@uni-mannheim.de

Database Systems II — Solution to Exercise Sheet 9
Spring Semester 2019 — Created May 8, 2019

Exercise 1

Find questions that you would consider appropriate to be asked in a DBS II exam. Also assign points to your task and give an outline of what is expected as the solution. As a reference, note that a typical exam has 90 points and a duration of 90 minutes. A typical exam covers questions of different types like

- *name/list n different properties of Y*,
- *discuss advantages and disadvantages of X*,
- *implement a function that does X*,
- *given the following piece of code, complete the code such that it does X*,
- *given the following piece of code/figure/graph/architecture, answer the following questions: ...*,
- *apply algorithm/method/technique X known from the lecture/exercise to the following problem Y*,
- or other types of questions.

Send your results to `daniel.flachs@uni-mannheim.de`. Appropriate questions (or variants thereof) might appear in the exam.

Exercise 2

Read (at least) Sections 1 to 3 of the paper *Making B^+-Trees Cache Conscious in Main Memory* by Rao and Ross.
`http://ftp.cse.buffalo.edu/users/azhang/disc/disc01/cd1/out/papers/sigmod/p475-rao/p475-rao.pdf`

Exercise 3

Note: The knowledge required to solve the following task is presented in the lecture on May 6, 2019.

Assume you are given the following query:

```
SELECT *
FROM R
WHERE Age > 27 and Income > 30.000 and Weight < 75;
```

That is, we are given a conjunctive query.

We refer to the predicates in the given query by the set

$$P = \{Age > 27, Income > 30\,000, Weight < 75\} = \{p_1, p_2, p_3\}$$

Furthermore, assume you are given the following sample taken from R:

ID	Age	Income	Weight
1	28	40,000	80
2	30	55,000	50
3	27	37,000	75
4	40	60,000	60
5	42	62,000	85
6	22	15,000	55
7	70	20,000	67
8	50	80,000	57
9	55	85,000	86
10	33	42,000	58

Exercise 3 a)

For each $P' \subseteq P$, compute the selectivity $\gamma(P')$ via the formula $F_\gamma(P')$ as described in the script. For instance, for $P' = \{p_1, p_3\}$, we have that $F_\gamma(P') = p_1 \wedge \neg p_2 \wedge p_3$, and, since 3 tuples in the sample qualify this predicate, $\gamma(P') = \frac{3}{10} = 0.3$.

Write the selectivities $\gamma(P'), P' \subseteq P$ as a vector γ where you order the elements in γ by the bitvector representation of P'. For instance, $P' = \{p_1, p_2\}$ has the bitvector representation $(0, 1, 1)$, i. e., the rightmost bit refers to p_1. See `https://en.wikipedia.org/wiki/Power_set#Representing_subsets_as_functions` for a detailed example.

- **Hint 1**: All P' form the power set of P and hence there are $2^{|P|}$ many P'. That is, in this example, 8.

- **Hint 2**: $\sum_{P' \subseteq P} \gamma(P') = 1$. If not, your calculations are wrong.

Solution

$$\gamma = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.1 \\ 0.3 \\ 0.1 \\ 0.1 \\ 0.0 \\ 0.4 \end{pmatrix} \quad \begin{matrix} 000 \\ 100 \\ 010 \\ 110 \\ 001 \\ 101 \\ 011 \\ 111 \end{matrix} \quad \begin{matrix} \neg p_1 \wedge \neg p_2 \wedge \neg p_3 \\ p_1 \wedge \neg p_2 \wedge \neg p_3 \\ \neg p_1 \wedge p_2 \wedge \neg p_3 \\ p_1 \wedge p_2 \wedge \neg p_3 \\ \neg p_1 \wedge \neg p_2 \wedge p_3 \\ p_1 \wedge \neg p_2 \wedge p_3 \\ \neg p_1 \wedge p_2 \wedge p_3 \\ p_1 \wedge p_2 \wedge p_3 \end{matrix}$$

Exercise 3 b)

The *complete design matrix* allows one to derive the selectivities of $\beta(P')$ for all $P' \subseteq P$ from the vector of gamma-selectivities γ. Give the complete design matrix C that is associated with $|P| = 3$.

Solution

The complete design matrix can be found recursively by

$$C_{|P|} = \begin{pmatrix} C_{|P|-1} & C_{|P|-1} \\ \mathbf{0} & C_{|P|-1} \end{pmatrix} \quad \text{and} \quad C_0 = (1)$$

where $\mathbf{0} \in \{0,1\}^{2^{|P|-1} \times 2^{|P|-1}}$, i. e., $\mathbf{0}$ is a zero matrix of dimension $2^{|P|-1} \times 2^{|P|-1}$.

$$C_3 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Exercise 3 c)

Compute $C\gamma$. Note that the result is a vector of selectivities, called the β-selectivities and is denoted by β. What are the predicates that each entry in β refers to?

Solution

$$
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix} 0.0 \\ 0.0 \\ 0.1 \\ 0.3 \\ 0.1 \\ 0.1 \\ 0.0 \\ 0.4 \end{pmatrix}
=
\begin{pmatrix} 1.0 \\ 0.8 \\ 0.8 \\ 0.7 \\ 0.6 \\ 0.5 \\ 0.4 \\ 0.4 \end{pmatrix}
\widehat{=}
\begin{pmatrix} \emptyset \\ p_1 \\ p_2 \\ p_1 \wedge p_2 \\ p_3 \\ p_1 \wedge p_3 \\ p_2 \wedge p_3 \\ p_1 \wedge p_2 \wedge p_3 \end{pmatrix}
$$

Exercise 3 d)

List all possible orderings in a query plan for the predicates in the given query. Ignoring predicate costs and assuming independence of predicates, what is the optimal ordering of predicates based on your results of the previous exercises?

Solution

For n predicates, there are $n!$ possible orderings. In our case:

- $p_1 - p_2 - p_3$
- $p_1 - p_3 - p_2$
- $p_2 - p_1 - p_3$
- $p_2 - p_3 - p_1$
- $p_3 - p_1 - p_2$
- $p_3 - p_2 - p_1$

Ignoring predicate costs and assuming independence of predicates, the optimal ordering is to order predicates by selectivity.

Since $sel(p_1) = 0.8$, $sel(p_2) = 0.8$ and $sel(p_3) = 0.6$ (see last exercise), the optimal orderings, under the aforementioned assumptions, are:

$$p_3 - p_1 - p_2$$

$$p_3 - p_2 - p_1$$