Solutions for Exercise Sheet 3

Richter, Yannick
MTK 03741982
ge78tup@mytum.de

Rodrigues, Diogo
MTK 03770446
diogo.rodrigues@tum.de

TUM – Query Optimization 2022/23
11th November 2022

Our solutions for Exercise Sheet 3.

Exercise 1

Given cardinalities $|R_1|$, $|R_2|$, the domain (the number of distinct values) of $R_1.x$ and $R_2.y$ of R_1 and R_2.

1. How can we estimate the selectivity of $\sigma_{R_1.x=c}$, where c is a constant?
   - If $R_1.x$ is a key of R_1
   - If $R_1.x$ is not a key of R_1
   - *Hint:* The selectivity is given by $\frac{|\sigma_{R_1.x=c}|}{|R_1|}$
2. How can we estimate the selectivity of $\bowtie_{R_1.x=R_2.y}$?
   - If both $R_1.x$ and $R_2.y$ are keys
   - If $R_1.x$ is a key but $R_2.y$ is not a key
   - If both $R_1.x$ and $R_2.y$ are not keys
   - *Hint:* The selectivity is given by $\frac{|\bowtie_{R_1.x=R_2.y}|}{|R_1||R_2|}$
   - *Hint:* You have to make some assumptions here. E.g. decide whether you assume that tuples from one relation always find join partners in the other relation. Note that we don't know the output size of $\sigma_{R_1.x=c}$ ($\bowtie_{R_1.x=R_2.y}$, respectively), so we can't simply use the definition of selectivity.

Item 1

Consider $m(R.c)$ to be the average multiplicity of a possible value of attribute c in relation R.

- If $R_1.x$ is a key
  - $m(R_1.x) = 1$
  - $|\sigma_{R_1.x=c}| = m(R_1.x)$
  - Selectivity $= \frac{|\sigma_{R_1.x=c}|}{|R_1|} = 1/|R_1|$.
- If $R_1.x$ is not a key
  - $m(R_1.x) = |R_1|/dom(R_1)$
  - $|\sigma_{R_1.x=c}| = m(R_1.x) = |R_1|/dom(R_1)$
  - Selectivity $= \frac{|\sigma_{R_1.x=c}|}{|R_1|} = \frac{m(R_1.x)}{|R_1|} = \frac{|R_1|/dom(R_1)}{|R_1|} = \frac{1}{dom(R_1)}$

Item 2

- If both $R_1.x$ and $R_2.y$ are keys
  - If we assume that every key has a join partner: $\frac{\min(|R_1|,|R_2|)}{|R_1|*|R_2|}$
- If $R_1.x$ is a key but $R_2.y$ is not a key
  - If we assume that every key has a join partner
  - $m(R_1.x) = 1$
  - $m(R_2.y) = \frac{|R_2|}{dom(R_2.z)}$
  - Selectivity $= \frac{|R_1|*m(R_2.y)}{|R_1|*|R_2|} = \frac{|R_1|*\frac{|R_2|}{dom(R_2.y)}}{|R_1|*|R_2|} = \frac{\frac{|R_2|}{dom(R_2.y)}}{|R_2|} = \frac{1}{dom(R_2.y)}$
- If both $R_1.x$ and $R_2.y$ are not keys
  - If we assume that $|R_1| < |R_2|$ Key has a join partner:
  - $m(R_1.x) = \frac{|R_1|}{dom(R_1.x)}$
  - $m(R_2.y) = \frac{|R_2|}{dom(R_2.y)}$
  - Selectivity $= \frac{\min(dom(R_1.x),dom(R_2.y))*m(R_1.x)*m(R_2.y)}{|R_1|*|R_2|} = \frac{\min(dom(R_1.x),dom(R_2.y))*\frac{|R_1|}{dom(R_1.x)}*\frac{|R_2|}{dom(R_2.y)}}{|R_1|*|R_2|} = \frac{\min(dom(R_1.x),dom(R_2.y))}{dom(R_1.x)*dom(R_2.y)}$

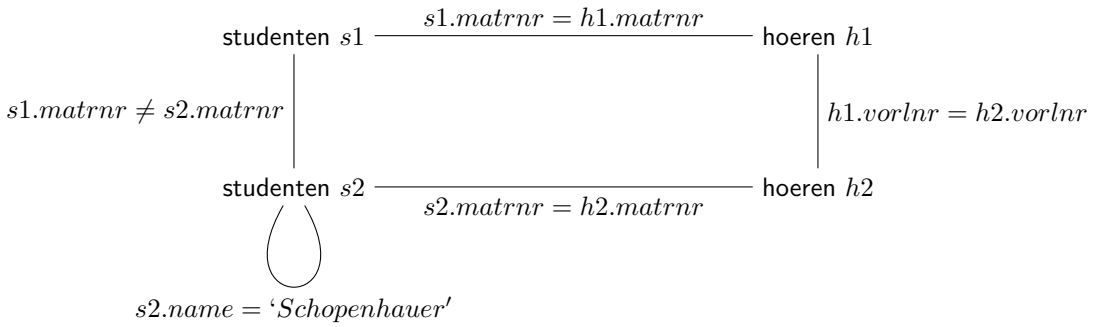
Exercise 2

Give the query graphs for the following two queries:

- Find all students that have ever attended a lecture together with Schopenhauer, excluding Schopenhauer himself.
- Find all professors who gave at least one lecture which was attended by at least two students

---

Item 1



Item 2