4. Exercise

Timo Bergerbusch 344408
Thomas Näveke 311392
Shu Xing 381176

03.12.2017

Exercise 4.1

1.

T_1 and T_2

$T_1 \subseteq T_2$**?**

$$\text{Find mapping } h : T_2 \to T_1, x \rightarrowtail \begin{cases} b_3 & x = b_1 \\ 5 & x = b_2 \\ b_4 & x = b_3 \\ b_3 & x = b_4 \\ a_1 & x = a_1 \\ a_2 & x = a_2 \end{cases} \Rightarrow T_1 \subseteq T_2$$

$T_2 \subseteq T_1$**?**

Impossible since we would have to map the constant 5 to an other constant and T_2 does not contain a constant. So $T_2 \not\subseteq T_1$

Overall

$\Rightarrow T_1 \subset T_2$, but $T_2 \not\subseteq T_1$ and therefore $T_1 \not\equiv T_2$

T_2 and T_3

$T_2 \subseteq T_3$**?**

When we map T_3 to T_2:
$a1 \to a1$
$a2 \to a2$
$b3 \to b4$
$b2 \to b3$
$b1 \to a1$, but then for (b4, b1) in T_3 we cannot find a matched instance in T_1
So there can not exists a function $h : T_3 \to T_2$. $\Rightarrow T_2 \not\subseteq T_3$

$T_3 \subseteq T_2$?

When we map T_2 to T_3:
$a1 \to a1$
$a2 \to a2$
$b4 \to b3$
$b1 \to b4$
$b3 \to b2$
$b2 \to b2$ (or b1)
but then for (b3,b2) in T_2 we cannot find a matched instance in T_3
So there can not exists a function $h : T_2 \to T_3$. $\Rightarrow T_3 \not\subseteq T_2$

Overall

$\Rightarrow T_2 \not\subseteq T_3$, $T_3 \not\subseteq T_2$ and therefore especially $T_2 \not\equiv T_3$

T_1 and T_3

$T_1 \subseteq T_3$?

When we map T_3 to T_1:
$a1 \to a1$
$a2 \to a2$
$b3 \to b3$
$b2 \to b4$
$b1 \to a1$
$b2 \to 5$, because of (b2, b3) in T_3 and b3$\to$ b3, so conflict with b2$\to$ b4
So there can not exists a function $h : T_3 \to T_1$. $\Rightarrow T_1 \not\subseteq T_3$

$T_3 \subseteq T_3$?

Analogously to $T_2 \subseteq T_1$ we would have to map the constant 5 to an other constant and T_3 does not contain a constant. So $T_3 \not\subseteq T_1$

Overall

$\Rightarrow T_1 \not\subseteq T_3$, $T_3 \not\subseteq T_1$ and therefore especially $T_1 \not\equiv T_3$

2.

T_1:

T_1 is minimal, because if we delete any row and so create T' we can not have a function $h : T \to T'$ such that $T' \subseteq T$

T_2:

1. delete row (b_4, a_2), since we have a function h_1, with $h_1(x) = \begin{cases} b_1 & \text{if } x = b_4 \\ x & \text{if } x \neq b_4 \end{cases}$

New Tableau T_1':

a_1	a_2	
b_1	a_2	(R)
a_1	b_3	(R)
b_2	b_4	(R)
b_2	b_1	(R)
b_3	b_2	(R)

2. delete row (a_1, b_3), since we have a function h_2, with $h_2(x) = \begin{cases} b_1 & \text{if } x = a_1 \\ a_2 & \text{if } x = b_3 \\ x & else \end{cases}$

New Tableau $T_1^{''}$:

a_1	a_2	
b_1	a_2	(R)
b_2	b_4	(R)
b_2	b_1	(R)
b_3	b_2	(R)

3. delete row (b_2, b_4), since we have a function h_2, with $h_3(x) = \begin{cases} b_1 & \text{if } x = b_4 \\ x & else \end{cases}$

New Tableau $T_1^{'''}$:

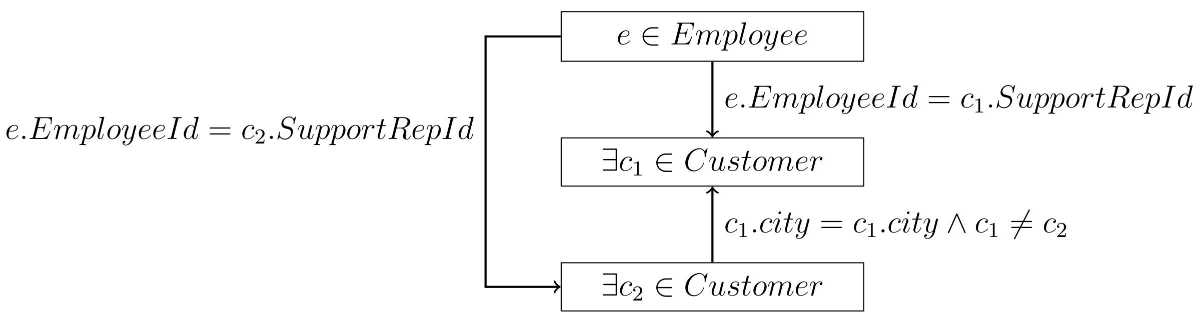
a_1	a_2	
b_1	a_2	(R)
b_2	b_1	(R)
b_3	b_2	(R)

4. This table $T_1^{'''}$, which es derived by $h_3(h_2(h_1(T)))$, is now minimal and equivalent

Exercise 4.2

1.a)

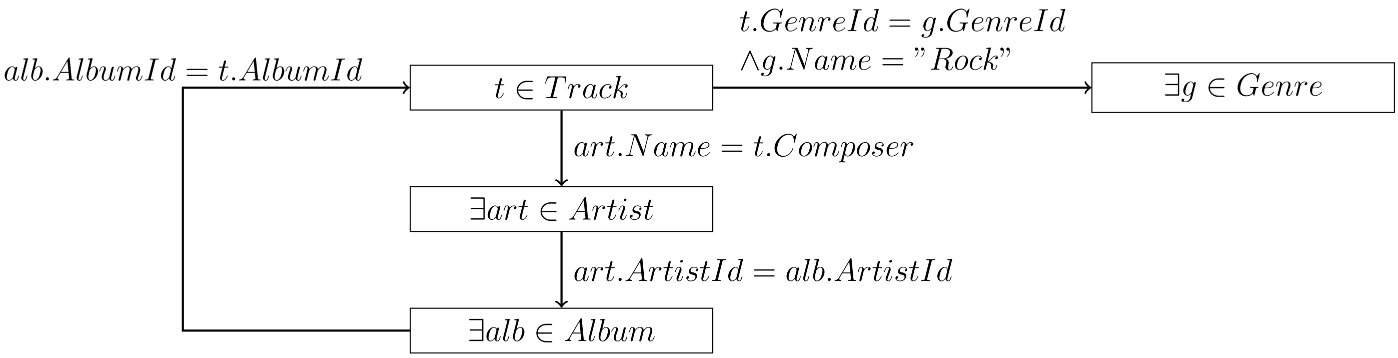
$\{< e.EmployeeId, e.LastName >|\ e \in Employee \wedge \exists c_1 \in Customer \wedge c_1.SupportedRepId = e.EmployeeId \wedge \exists c_2 \in Customer \wedge c_2.SupportRepId = e.EmployeeId \wedge c_1.City = c_2.City \wedge c_1 \neq c_2\}$



This graph is cycle free, which means the query is optimizable using semi-joins.

1.b)

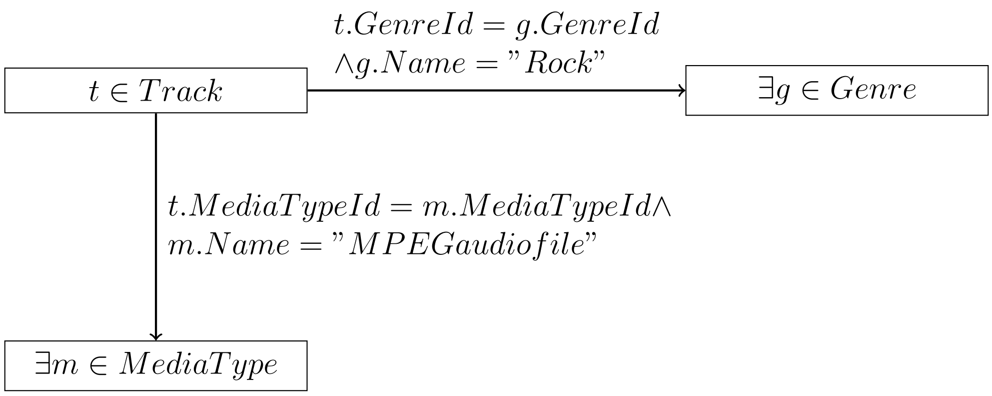
$\{< t.name, t.composer >| \ t \in Track \wedge \exists g \in Genre \wedge g.GenreId = t.GenreId \wedge g.Name = "Rock" \wedge \exists art \in Artist \wedge \exists alb \in Album \wedge alb.ArtistId = art.ArtistId \wedge t.AlbumId = alb.AlbumId \wedge t.Composer = art.Name\}$



This Graph has a cycle. This means the query is not optimizable using semi-joins.

2.

$\{< t.Name >| \ t \in Track \wedge t.Miliseconds \leq 90000 \wedge \exists g \in Genre \wedge t.GenreId = t.GenreId \wedge g.Name = "Rock" \wedge \exists m \in MediaType \wedge t.MediaTypeId = m.MediaTypeid \wedge m.Name = "MPEGaudiofile"\}$



TODO: es fehlt das mit den ≤90000ms

The graph is cycle free, which means it can be optimized using semi-joins.

Exercise 4.3

1.

B = # buffer pages in memory = 4
N = # pages in the input file = $\lceil \frac{7500 \cdot 36}{1024-64} \rceil = 282$
After the initial pass:

runs (subfiles) = $\lceil \frac{N}{B} \rceil = \lceil \frac{282}{4} \rceil = 71$, each subfiles: 4-pages runs

2.

passes = $1 + \lceil \log_{B-1} \lceil \frac{N}{B} \rceil \rceil = 1 + \lceil \log_{4-1} 71 \rceil = 1 + 4 = 5$

3.

Total costs = $2N \cdot (\#passes) = 2 \cdot 282 \cdot 5 = 2820$ I/Os

4.

If B = 4:
$1 + \lceil \log_3 \lceil \frac{N}{B} \rceil \rceil = 2 \Rightarrow (\log_3 \lceil \frac{N}{4} \rceil)_{max} = 1 \Rightarrow (\lceil \frac{N}{4} \rceil)_{max} = 3 \Rightarrow N_{max} = 12$
records in the file = $\frac{12 \cdot (1024-64)}{36} = 320$ records

If B = 42:
$1 + \lceil \log_{41} \lceil \frac{N}{B} \rceil \rceil = 2 \Rightarrow (\log_4 1 \lceil \frac{N}{42} \rceil)_{max} = 1 \Rightarrow (\lceil \frac{N}{42} \rceil)_{max} = 41 \Rightarrow N_{max} = 41 \cdot 42 = 1722$
records in the file = $\frac{1722 \cdot (1024-64)}{36} = 45920$ records