Database Systems, CSCI 4380-01
Homework #2 Solutions
Due Thursday February 10, 2011 at 2 pm

Please submit your answers as a text or PDF file via RPILMS assignment drop box function.

Question 1 (20+10 bonus points). You are given the following simple relations (for simplicity, we are using unique integer ids as identifiers in this example):

Person(id, name, countryOfBirth, yearOfBirth)
Topics(id, title)
Books(id, title, publisher, isbn, versionNo, publicationYear)
BookAuthorbook_id, person_id)
BookEditor(book_id, person_id)
BookTopic(book_id, topic_id)

Write the following queries using relational algebra.

(a) Find people who are both an author and editor of the same book on topic 'Databases'.

Answer:

R1:=$\pi_{person\_id,book\_id}(BookAuthor) \cap \pi_{person\_id,book\_id}(BookEditor)$

R2:=$\pi_{id}(\sigma_{\texttt{title='DataBase'}}Topics)$

R3:=$\pi_{book\_id}(R2 \bowtie_{\texttt{id=topic\_id}} BookTopic)$

$\pi_{person\_id}(R1 \bowtie BookTopic)$

(b) Find the latest version of all books on 'Artificial Intelligence' topic (use versionNo).

Answer:

R1:=$\pi_{id}(\sigma_{\texttt{title='Artificial Intelligence'}}Topics)$

R2:=$\pi_{book\_id}(R1 \bowtie_{\texttt{id=topic\_id}} BookTopic)$

R3:=$\pi_{id,versionNo}(Books \bowtie_{\texttt{id=book\_id}} R2)$

R4:=$\rho_{R4(id2,versionNo2)}(R3)$

R5:=$\pi_{id,versionNo}(R3 \bowtie_{\texttt{versionNo < versionNo2}} R4)$

π_{id} (R3) - π_{id} (R5)

(c) **Bonus.** Find people who authored a book every year between 1990 and 2000. Assume that the database contains at least one book published every year between 1990 and 2000.

Answer:

Solution 1:

R1:=$\pi_{person\_id}(BookAuthor \bowtie_{\texttt{book\_id = id}} (\sigma_{publication\_year>=1990}Books))$

R2:=$\pi_{person\_id}(BookAuthor \bowtie_{\texttt{book\_id = id}} (\sigma_{publication\_year>=1990}Books))$

R3:=$\pi_{person\_id}(BookAuthor \bowtie_{\texttt{book\_id = id}} (\sigma_{publication\_year>=1990} Books))$

R4:=$\pi_{person\_id}(BookAuthor \bowtie_{\texttt{book\_id = id}} (\sigma_{publication\_year>=1990} Books))$

R5:=$\pi_{person\_id}(BookAuthor \bowtie_{\texttt{book\_id = id}} (\sigma_{publication\_year>=1990} Books))$

R6:=$\pi_{person\_id}(BookAuthor \bowtie_{\texttt{book\_id = id}} (\sigma_{publication\_year>=1990} Books))$

R7:=$\pi_{person\_id}(BookAuthor \bowtie_{\texttt{book\_id = id}} (\sigma_{publication\_year>=1990} Books))$

R8:=$\pi_{person\_id}(BookAuthor \bowtie_{\texttt{book\_id = id}} (\sigma_{publication\_year>=1990} Books))$

R9:=$\pi_{person\_id}(BookAuthor \bowtie_{\texttt{book\_id = id}} (\sigma_{publication\_year>=1990} Books))$

R10:=$\pi_{person\_id}(BookAuthor \bowtie_{\texttt{book\_id = id}} (\sigma_{publication\_year>=1990} Books))$

R11:=$\pi_{person\_id}(BookAuthor \bowtie_{\texttt{book\_id = id}} (\sigma_{publication\_year>=1990} Books))$

$R1 \cap R2 \cap R3 \cap R4 \cap R5 \cap R6 \cap R7 \cap R8 \cap R9 \cap R10 \cap R11$

Solution 2:

R1 := $\pi_{publication\_year}(\sigma_{publication\_year>=1990 \texttt{ and } publication\_year<=2000}\ Books)$
Find the books that published between 1990 to 2000.

R2 := $\pi_{person\_id} BookAuthor$
Project the person_id of Bookauthors

R3 := $R2 \times R1$
Use Cartesian join to create a relation that has one attribute contains all the bookauthor and one attribute contains year from 1990 to 2000.

R4 := $\pi_{person\_id,year}(\sigma_{publication\_year>=1990 \texttt{ and } publication\_year<=2000}\ Books)$
Create a relation that all the bookauthor id with the year that they have publication.

R5 := $\pi_{person\_id}(R3 - R4)$
Select the bookauthors who didn't author at least one book published every year between 1990 and 2000.

R6 := $\pi_{person\_id} R4 - R5$
Return the bookauthors who authored a book every year between 1990 and 2000.

Question 2 (10+20+20 points). You are given the following functional dependency set for relation $R(A, B, C, D, E, F, G)$. Assume these dependencies form a minimal basis.

$$\mathcal{F} = \{A \to BC, AD \to F, AF \to E\}$$

(a) Based on these functional dependencies, what are the keys for R?

Answer:

The key is {A,D,G}

(b) Is this relation in BCNF? If not, list all functional dependencies that violate BCNF and convert it to BCNF using BCNF decomposition.

Answer:

No. This relation is not in BCNF.

The functional dependencies that violate BCNF are $\{A \to BC, AD \to F, AF \to E\}$.

Using Algorithm 3.20 on page 92 of the text book gives a set of relations:

Step1:

Because $AD \to$ F violates BCNF. Thus decompose R to:

R1(A,B,C), $\{A \to BC\}$; key:A in BCNF

R2(A,D,E,F,G), $\{AD \rightarrow F, AF \rightarrow E\}$; key:A,D,G violates BCNF;

Step2:

Since R1 is in BCNF, but R2 violates BCNF, thus decompose R2 to:

R21(A,D,F), $\{AD \rightarrow F\}$; key:A,D in BCNF

R22(A,D,E,G), $\{AD \rightarrow E\}$; key:A,D,G violates BCNF

Step3:

Since R21 is in BCNF, but R22 violates BCNF, thus decompose R22 to:

R221 (A,D,E), $\{AD \rightarrow E\}$; key:A,D in BCNF

R222 (A,D,G); key:A,D,G in BCNF

So, after BCNF decomposition, we have:

R1(A,B,C)

R21(A,D,F)

R221 (A,D,E)

R222 (A,D,G)

(c) Is this relation in 3NF? If not, list all functional dependencies that violate 3NF and convert it to 3NF using 3NF decomposition.

Answer:

No. This relation is not in 3NF.

The functional dependencies that violate 3NF are $\{A \rightarrow BC, AD \rightarrow F, AF \rightarrow E\}$.

Using Algorithm 3.26 on page 103 of the text book gives a set of relations:

R1(A,B,C), $\{A \rightarrow BC\}$; key:A in 3NF

R2(A,D,F), $\{AD \rightarrow F\}$; key:A,D in 3NF

R3(A,F,E), $\{AF \rightarrow E\}$; key:A,F in 3NF

Because none of these schema are a key or super key for R, we must add a relation with the key of R as a schema.

R4(A,D,G); key:A,D,G in 3NF

Question 3 (10 points): Find a minimal basis that is equivalent to the following set of functional dependencies:

$$\mathcal{F} = \{AF \rightarrow BD, A \rightarrow F, C \rightarrow D, CD \rightarrow E, AB \rightarrow B\}$$

Answer:

$\mathcal{F} = \{A \rightarrow B, A \rightarrow D, A \rightarrow F, C \rightarrow D, C \rightarrow E\}$

Step 1 of Minimal Basis Algorithm: Form a basis.

$AF \rightarrow B$

$AF \rightarrow D$

$A \rightarrow F$

$C \rightarrow D$

$CD \rightarrow E$

$AB \rightarrow B$

Step 2 of algorithm: Remove trivial functional dependencies.

$AB \to B$ is trivial and can be removed, leaving:

$AF \to B$

$AF \to D$

$A \to F$

$C \to D$

$CD \to E$

Step 3 of Algorithm: if after removing $X \to Y$, the remaining FDs imply $X \to Y$, then we can remove $X \to Y$.

There are no functional dependencies that can be removed in this step.

Step 4 of Algorithm: Given $XZ \to Y$ in F, remove Z if unneccessary.

Because $A \to F$, so the F in $AF \to B$, the F in $AF \to D$ can be removed.

Because $C \to D$, so the D in $CD \to E$ can be removed.

Thus, the final minimal basis is :

$\mathcal{F} = \{A \to B, A \to D, A \to F, C \to D, C \to E\}$