Tutorial 2 - Relational Algebra

CSC343 - Introduction to Databases
Fall 2008

TA: Lei Jiang

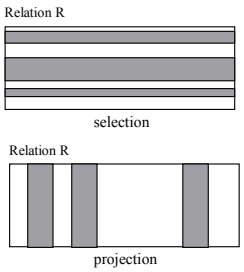
Selection and Projection

$\sigma_c(R) = \{s \in R \mid s \text{ satisfies condition } c\}$

--- selection based on condition c

$\pi_{a_1,\ldots,a_n}(R) = R_{a_1,\ldots,a_n}$

--- projection onto the columes $a_1 \ldots a_n$



Selection and Projection

- **Question**
  - Given following database schema, write following queries in relational algebra.

patients (*pnum, pname, age*)
doctors (*dnum, dname, rank, phone*)
visits (*pnum, dnum, date, diagnosis*)

  - Q1: Who are the patients 10 years old or younger?
    $\sigma_{age \leq 10}(patients)$
  - Q2: Who are the surgeons?
    $\sigma_{rank=surgeon}(doctors)$
  - Q3: What are the phone numbers of doctors?
    $\pi_{dname, phonenum}(doctors)$
  - Q4: What are the phone numbers of surgeons?
    $\pi_{dname, phonenum}(\sigma_{rank=surgeon}(doctors))$

order matters

Set Operations

$R_1 \cap R_2 = \{s \mid s \in R_1 \text{ and } s \in R_2\}$
--- intersection of R_1 and R_2, which must be "schema compatible"
$R_1 \cup R_2 = \{s \mid s \in R_1 \text{ or } s \in R_2\}$
--- union of R_1 and R_2, which must be "schema compatible"
$R_1 - R_2 = \{s \mid s \in R_1 \text{ and } s \notin R_2\}$
--- set difference of tuples from R_1 and R_2, which must be "schema compatible"

R_1

A	B
1	2
3	4

R_2

A	B
1	2
3	4
6	7

$R_1 \cup R_2$

A	B
1	2
3	4
6	7

$R_1 \cap R_2$

A	B
1	2
3	4

$R_1 - R_2$

A	B

$R_2 - R_1$

A	B
6	7

Set Operations

- **Question**

patients (*pnum, pname, age*)
doctors (*dnum, dname, rank*)
visits (*pnum, dnum, date, diagnosis*)

  - Q5: Restate the expression $\sigma_{age\leq 10 \vee age\geq 60}(patients)$ using set operations.

- ***The way of thinking***:
  - find the set A of patients with age >= 60;
  - find the set B of patients with age <=10;
  - take A union B.

- ***Answer***: $\sigma_{age\leq 10}(patients) \bigcup \sigma_{age\geq 60}(patients)$

Set Operations

- **Question**

patients (*pnum, pname, age*)
doctors (*dnum, dname, rank*)
visits (*pnum, dnum, date, diagnosis*)

  - Q6: Restate the expression $\sigma_{rank\neq surgeon \wedge rank\neq \text{oculist}}(doctors)$ using set operations without $\neq$ and $\wedge$.

- ***The way of thinking***:
  - find the set A of oculists;
  - find the set B of surgeon;
  - find the set C of all doctors; (d) take C – (A union B).

- ***Answer***:
$doctors - (\sigma_{rank=surgeon}(doctors) \bigcup \sigma_{rank=\text{oculist}}(doctors))$

Set Operations

- **Question**

patients (*pnum, pname, age*)
doctors (*dnum, dname, rank*)
visits (*pnum, dnum, date, diagnosis*)

  - Q7: Find all the patients who saw doctor 801 but not 802 (i.e. dnum=801, dnum≠802)

- ***The way of thinking***:
  - find the set A of patients who saw doctor 801
  - find the set B of patients who saw doctor 802;
  - take A – B.

- ***Answer***: $\pi_{pnum}\sigma_{dnum=801}(visits) - \pi_{pnum}\sigma_{dnum=802}(visits)$

Cartesian Product and Join

$R_1 \times R_2 = \{<s_1,s_2> \mid s_1 \in R_1 \text{ and } s_2 \in R_2\}$

--- cartesian product of R_1 and R_2

$R_1 \overset{R_1a\ \theta\ R_2b}{\bowtie} R_2 = \sigma_{R_1a\ \theta\ R_2b}(R_1 \times R_2)$

--- θ join, where θ can be $<, \leq, >, \geq, =, \neq$, etc

$R_1 \overset{R_1a = R_2b}{\bowtie} R_2 = \sigma_{R_1a = R_2b}(R_1 \times R_2)$

--- equi-join, same as θ join where θ is =

May also project out $R_2.b_x$

$R_1 \bowtie R_2 = \sigma_{R_1a_1=R_2b_1,\ldots,R_1a_n=R_2b_n}(R_1 \times R_2)$

--- nature join, same as equi-join on all common attributes

R_1

A	B
1	2
3	4

R_2

B	C
2	3
2	4
4	7

$R_1 \bowtie R_2$

A	B	C
1	2	3
1	2	4
3	4	7

Cartesian Product and Join

- **Question**

  patients (*pnum, pname, age*)
  doctors (*dnum, dname, rank*)
  visits (*pnum, dnum, date, diagnosis*)

  – Q8: Form peer groups for patients, where a peer group is a pair of patients where age difference is less than 10 years (can use the rename operator $\rho_A(R)$)

- ***The way of thinking***:
  – form the set A of all patient peer groups;
  – select the pairs from A such that the age difference is less than 10 years

- ***Answer***:

$$\sigma_{A.age \geq B.age \wedge A.age \leq B.age+10 \wedge A.pnum \neq B.pnum} (\rho_A(\pi_{pnum,age}(patients)) \times \rho_B(\pi_{pnum,age}(patients)))$$

Cartesian Product and Join

- **Question**

  patients (*pnum, pname, age*)
  doctors (*dnum, dname, rank*)
  visits (*pnum, dnum, date, diagnosis*)

  – Q9: Who are the surgeons who visited the patient 101 (i.e. pnum = 101)?

- ***The way of thinking***:
  – find the set A of visits by patient 101;
  – find the set B of surgeons;
  – join A and B on *dnum*.

- ***Answer***:

$$\sigma_{visits.dnum=doctors.num \wedge rank=surgeon \wedge pnum=101} (\pi_{pnum,dnum}(visits)) \times \pi_{dnum,rank}(doctors)))$$

Cartesian Product and Join

- **Question**

  patients (*pnum, pname, age*)
  doctors (*dnum, dname, rank*)
  visits (*pnum, dnum, date, diagnosis*)

  – Q10: Who has seen a surgeon after April 11, 2001?

- ***The way of thinking***:
  – find the set A of patient visits after April 11, 2001;
  – find the set B of surgeons;
  – join A and B on *dnum* and project onto *pnum*.

- ***Answer***:
  $\pi_{pnum}\sigma_{visits.dnum=doctors.dnum}(\sigma_{date\geq 2001.4.12}(visits)\times\sigma_{rank=surgeon}(doctors))$

Cartesian Product and Join

- **Question**

  patients (*pnum, pname, age*)
  doctors (*dnum, dname, rank*)
  visits (*pnum, dnum, date, diagnosis*)

  – Q11: Is there any non-surgeon doctors who performed a surgeon (a doctor performed a surgeon if the visit record shows diagnosis="operation" for him)?

  – ***The way of thinking***:
    • find the set A of non-surgeon;
    • find the set B of patient visits where an operation is performed;
    • join A and B on *dnum* and project onto *dnum*.
  – ***Answer***:
    $\pi_{dnum}\sigma_{doctors.dnum=visits.dnum}(\sigma_{rank\neq surgeon}(doctors)\times\sigma_{diagnosis=operation}(visists))$

Division

Given two relation schemas $R_1(A,B)$ and $R_2(A)$ where $R_1.A$ and $R_2.A$ has the same domain, ***R_1 / R_2 = {x | x is a A value such that for every B value y in (a tuple of) R_2 there is a tuple <x,y> in R_1}***

R_1

A	B
1	2
3	2
7	0
5	2
9	0
5	0
9	2
5	6
7	4
1	4
5	4

R_2

B
0
2
4
6

R_1 / R_2

A
5

Divison

- **Question**

patients (*pnum, pname, age*)
doctors (*dnum, dname, rank*)
visits (*pnum, dnum, date, diagnosis*)

  - Q12: Who has seen **all** the surgeons after April 11, 2001?

- ***The way of thinking***:
  - find the set A of patients who has seen some surgeon projected onto *pnum* and *dnum*;
  - find the set B of all surgeon projected onto *dnum*;
  - take A / B

- ***Answer***:

$$\pi_{\text{pnum, dnum}}(\sigma_{date \geq 2001.4.12}(visits)) / \pi_{dnum}(\sigma_{rank=surgeon}(doctors))$$

Division

- **Question**

  patients (*pnum, pname, age*)
  doctors (*dnum, dname, rank*)
  visits (*pnum, dnum, date, diagnosis*)

  – Q13: Find all patients except for the youngest ones?

- ***The way of thinking***:
  – find the Y set of the youngest patients
    - by forming the set P of pairs of patients (A, B) where A.age <= B.age;
    - dividing P by the set of all patients (so we get the set of patients who is younger than or has the same age as all patients)
  – subtract Y from the set of all patients.
- ***Answer***:

$$\pi_{pnum,age}(patients) - (\sigma_{A.age \le B.age}(\rho_A(\pi_{pnum,age}(patients)) \times \rho_B(\pi_{pnum,age}(patients))) / \pi_{pnum,age}(patients))$$

Division

- Given two relation schemas $R_1(A,B)$ and $R_2(A)$, express division R_1/R_2 in terms of other RA operators.

- ***The way of thinking***:
  – find the set S of A values that does not appear with some B values in a tuple of R1
    - by finding the set of P of all possible (A,B) pairs;
    - subtracting P by R1 and projecting onto A
  – find the set of L of all A values in R1
  – taking L – S.

- ***Answer***: $R_1 / R_2 = \pi_A(R_1) - \pi_A((\pi_A(R_1) \times R_2) - R_1)$