CSC 343 – Introduction to Databases

Tutorial #4

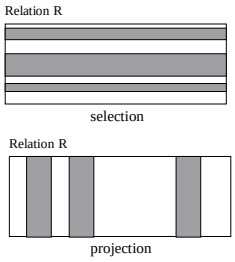
Relational Algebra

We will use following schemas in our discussion,

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

1. Simple selection and projection

$\sigma_c(R) = \{s \in R \mid s \text{ satisfies condition } c\}$
--- selection based on condition c
$\pi_{a_1,...,a_n}(R) = R_{a_1,...,a_n}$
--- projection onto the columes a_1 ... a_n



i. Who are the patients 10 years old or younger?

$s_{age \le 10}(patients)$

ii. Who are the surgeons

$s_{rank=surgeon}(doctors)$

iii. What are the phone numbers of doctors

$p_{dname,phonenum}(doctors)$

iv. What are the phone numbers of surgeons

$p_{dname,phonenum}(s_{rank=surgeon}(doctors))$

note: here the order (i.e. the sequence of applying projection and selection matters), so

$p_{dname,phonenum}(s_{rank=surgeon}(doctors)) \neq s_{rank=surgeon}(p_{dname,phonenum}(doctors))$

2. 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

$R_1 - R_1$

A	B

$R_2 - R_1$

A	B
6	7

R_2

C	D
a	b
c	d

$R_1 \times R_2$

A	B	C	D
1	2	a	b
1	2	c	d
3	4	a	b
3	4	c	d

i. Re-state the expression $s_{age \le 10 \vee age \ge 60}(patients)$ using set operations.

$s_{age \le 10}(patients) \cup s_{age \ge 60}(patients)$

ii. Re-state the expression $s_{rank \ne surgeon \wedge rank \ne \text{oculist}}(doctors)$ using set operations without $\ne$ and $\wedge$

$doctors - (s_{rank = surgeon}(doctors) \cup s_{rank = \text{oculist}}(doctors))$

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

$p_{pnum} s_{dnum=801}(visits) - p_{pnum} s_{dnum=802}(visits)$

3. 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

$R_1 \overset{R_1a\ \theta\ R_2b}{\bowtie} R_2 = s_{R_1a\ \theta\ R_2b}(R_1 \times R_2)$
--- θ join, where θ can be $<$, $\leq$, $>$, $\geq$, $=$, $\neq$, etc

$R_1 \overset{R_1a\ =\ R_2b}{\bowtie} R_2 = s_{R_1a\ =\ R_2b}(R_1 \times R_2)$
--- equi-join, same as θ join where θ is =

$R_1 \bowtie R_2 = s_{R_1a_1=R_2b_1,\ldots,R_1a_n=R_2b_n}(R_1 \times R_2)$
--- nature join, same as equi-join on all common attributes

i. 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 $r_A(R)$).
idea: use cartesian product to form all pairs, then select the appropriate pairs

$$s_{A.age \geq B.age \wedge A.age \leq B.age+10 \wedge A.pnum \neq B.pnum}\ (r_A(p_{pnum,age}(patients)) \times r_B(p_{pnum,age}(patients)))$$

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

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

iii. Who has seen a surgeon in the past two years?

$$p_{pnum} s_{visits.dnum=doctors.dnum}(s_{date \geq 2001.4.12}(visits) \times s_{rank=surgeon}(doctors))$$

note: in relational algebra, we don't need to eliminate duplicates since relations are treated as ***sets*** of tuples.

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

$$p_{dnum} s_{doctors.dnum=visits.dnum}(s_{rank \neq surgeon}(doctors) \times s_{diagnosis=operation}(visists))$$

4. Divison

$R_1 / R_2 = \pi_A(R_1) - \pi_A((\pi_A(R_1) \times R_2) - R_1)$

--- R_1 divided by R_2, where $R_1(A,...)$ and $R_2 \subset R_1$

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

i. Who has seen **all** the surgeons in the past two months?

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