Relational Algebra

Lecture 1 CPTR 319 2010
Scot Anderson, Ph.D.
School of Computing
Southern Adventist University

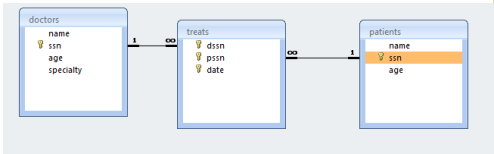
Outline

- Terms Review
- What About sets?
- Relational Algebra Operators
- Additional Operators
- Examples

A Database is…

name	ssn	age	specialty
Anderson	181-1818	39	general
Moss	244-2444	30	gynecology
Paine	266-2666	45	cardiology
Miller	300-0000	60	neurology
O'Brien	333-3736	36	gynecology
McBride	333-3737	36	urology
Nelson	400-0000	36	cardiology
Hardy	454-5454	33	radiology
Snow	500-0000	65	radiology
Peper	555-5555	42	cariology
Cheney	987-6543	50	neurology

dssn	pssn	date
181-1818	111-1111	Aug. 20, 1992
244-2444	454-5454	Oct. 18 1992
266-2666	222-9999	Sept 1, 1990
300-0000	345-6789	May 6, 1989
333-3737	567-0000	Feb 2, 1993
400-0000	222-9999	Dec. 10, 1992
454-5454	567-0000	July 14, 1992
555-5555	111-2222	May 6, 1989
555-5555	111-2222	Sept 5, 1989
555-5555	222-9999	March 3, 1991
987-6543	333-4444	June 23, 1993
987-6543	333-5555	June 23, 1993



Terms

- Tables (Relations)
- Columns
- Rows (Tuples)
- Relationships
- Database consists of
  - a set of tables and a set of relationships between tables.
  - Each of relation has a set of columns and a set of rows (tuples)
  - Each row always has the same columns as every other row in that table.

Set Operations

A = {Apple, Orange, Pear}
B = {Orange, Grape Fruit, Lemon}

Intersection: $A \cap B$ = {Orange}

Union: $A \cup B$ =
{Apple, Orange, Pear, Lemon, Grape Fruit}

$A \times B$ = {(Apple, Orange), (Apple, Grape Fruit), (Apple, Lemon), (Orange, Orange), (Orange, Grape Fruit), (Orange, Lemon), (Pear, Orange), (Pear, Grape Fruit), (Pear, Lemon)}

$A \setminus B$ = {Apple, Pear}

What does this have to do with Databases?

- Databases are made up of tables that have sets of rows.
- The rows often look like the cross product in that they have multiple elements or fields in a row (or tuple).
- As long as we have sets of tuples (fields in each tuple) we should have some organized way of dealing with them.
- Sets operations are only part of the story…

Relational Algebra

- Procedural Language
- Six basic operators
  - Select – rows by restricted the domain of fields
  - Project – restricts the fields by projecting out a subset of the columns.
  - Union – add rows from two different relations provided they have the same type columns.
  - Set difference – we've seen
  - Cartesian product – we've seen
  - Rename – simply renames a relation
- The operators take one or more relations as inputs and give a new relation as a result.

Select Operation

- Notation $\sigma_p(r)$
- p is called the **selection predicate**
- p is a formula in propositional calculus consisting of **terms** connected by: $\wedge$ (and), $\vee$ (or), $\neg$ (not)
- Terms consist of <attribute> OP <attribute> or <constant>
- Example:
- $\sigma_{\text{specialty}=\text{"general"}}(\text{doctors})$

Selection Example

name	ssn	age	specialty
Anderson	181-1818	39	general
Moss	244-2444	30	gynecology
Paine	266-2666	45	cardiology
Miller	300-0000	60	neurology
O'Brien	333-3736	36	gynecology
McBride	333-3737	36	urology
Nelson	400-0000	36	cardiology
Hardy	454-5454	33	radiology
Snow	500-0000	65	radiology
Peper	555-5555	42	cariology
Cheney	987-6543	50	neurology

- $\sigma_{\text{specialty}=\text{"general"}}(\text{doctors})$ =

name	ssn	age	specialty
Paine	266-2666	45	cardiology
Nelson	400-0000	36	cardiology

Project Operation

- $\Pi_{\text{list}}(r)$ – project the field list from the complete list of fields (columns)
- $\Pi_{\text{name, specialty}}(\text{doctors})$ =

name	specialty
Anderson	general
Moss	gynecology
Paine	cardiology
Miller	neurology
O'Brien	gynecology
McBride	urology
Nelson	cardiology
Hardy	radiology
Snow	radiology
Peper	cariology
Cheney	neurology

Composition of operators

- $\sigma_{\text{specialty}=\text{"general"}}(\Pi_{\text{name, specialty}}(\text{doctors}))$

name	specialty
Paine	cardiology
Nelson	cardiology

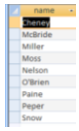
Union Operator

- $\Pi_{\text{name, ssn}}(\text{doctors}) \cup \Pi_{\text{name, ssn}}(\text{patients})$

name	ssn
Anderson	111-2222
Anderson	181-1818
Benson	999-1111
Brown	111-1111
Cheney	987-6543
Davis	777-7777
Edwards	444-3333
Hardy	454-5454
Harrison	123-4444
Irving	567-0000
Johnson	888-8888
McBride	333-3737
Miller	300-0000
Moss	244-2444
Nelson	400-0000
O'Brien	333-3736
Paine	266-2666
Palmer	222-9999
Peper	555-5555
Robertson	333-4444
Smith	345-6789
Snow	500-0000
Truner	666-6666
Williams	333-5555

Set Difference

- Show the name of those doctors that do not have the same name as some patient.
- $\Pi_{name}(doctor) \setminus \Pi_{name}(patient)$



name
Cheney
McBride
Miller
Moss
Nelson
O'Brien
Paine
Peper
Snow

Rename Operation

- $\rho_{x(A1,\ A2,\ A3,\ldots)}(r)$ will rename relation r to x with attributes have the new names A1, A2, A3, …
- $\rho_{d(dname,\ dssn,\ dage,specialty)}(doctor)$

dname	dssn	dage	specialty
Anderson	181-1818		39 general
Moss	244-2444		30 gynecology
Paine	266-2666		45 cardiology
Miller	300-0000		60 neurology
O'Brien	333-3736		36 gynecology
McBride	333-3737		36 urology
Nelson	400-0000		36 cardiology
Hardy	454-5454		33 radiology
Snow	500-0000		65 radiology
Peper	555-5555		42 cariology
Cheney	987-6543		50 neurology

Cartesian-Product

- $\rho_{doctors(DoctorName,DSSN,dage,specialty)}(doctors) \times \rho_{patients(PatientName,PSSN,page)}(patients) = \ldots$
- This is essentially just some renaming of: doctors × patients …

Result:

DoctorName	DSSN	dage	specialty	PatientNam	PSSN	page
Paine	266-2666		45 cardiology	Brown	111-1111	45
McBride	333-3737		36 urology	Brown	111-1111	45
O'Brien	333-3736		36 gynecology	Brown	111-1111	45
Peper	555-5555		42 cariology	Brown	111-1111	45
Cheney	987-6543		50 neurology	Brown	111-1111	45
Moss	244-2444		30 gynecology	Brown	111-1111	45
Miller	300-0000		60 neurology	Brown	111-1111	45
Hardy	454-5454		33 radiology	Brown	111-1111	45
Nelson	400-0000		36 cardiology	Brown	111-1111	45
Snow	500-0000		65 radiology	Brown	111-1111	45
Anderson	181-1818		39 general	Brown	111-1111	45
Paine	266-2666		45 cardiology	Anderson	111-2222	67
McBride	333-3737		36 urology	Anderson	111-2222	67
O'Brien	333-3736		36 gynecology	Anderson	111-2222	67
Peper	555-5555		42 cariology	Anderson	111-2222	67
Cheney	987-6543		50 neurology	Anderson	111-2222	67
Moss	244-2444		30 gynecology	Anderson	111-2222	67
Miller	300-0000		60 neurology	Anderson	111-2222	67
Hardy	454-5454		33 radiology	Anderson	111-2222	67
Nelson	400-0000		36 cardiology	Anderson	111-2222	67
Snow	500-0000		65 radiology	Anderson	111-2222	67
Anderson	181-1818		39 general	Anderson	111-2222	67
Paine	266-2666		45 cardiology	Davis	777-7777	45
McBride	333-3737		36 urology	Davis	777-7777	45
O'Brien	333-3736		36 gynecology	Davis	777-7777	45
Peper	555-5555		42 cariology	Davis	777-7777	45

Additonal Operator ∩

- $r \cap s = r \setminus (r \setminus s)$ or $r - (r - s)$.
- Set intersection is not needed since we can define it with set difference.
- But its useful and we allow its inclusion as **syntactic sugar**.

Additional Operator: Natural Join ⋈

- Show the names of doctors and their patients.
- $\sigma_{ddssn=dssn \wedge pssn=ppssn}(\rho_{d(dname,\ ddssn)}(\Pi_{name,ssn}(doctor)) \times treats \times \rho_{d(pname,ppssn)}(\Pi_{name,ssn}(patient)))$
- OR:
- $\Pi_{\ dname,pname}(\rho_{d(dname,\ dssn)}\ (doctor) \bowtie treats \bowtie \rho_{d(dname,\ dssn)}\ (patient))$