关系代数

讲解人：陆伟 教授

关系模型的特征

关系操作(Relational manipulation)

- The data manipulation on relational model is in fact the manipulation on relation or set.
- The relational algebra and relational calculus are two formal, non-user-friendly languages but they have been used as the basis for other higher-level Data Manipulation Languages (DMLs) for relational database.
- The relational algebra and relational calculus are equivalent to one another.

The relational algebra is a theoretical language with operations that work on one or more relations to define another relation without changing the original relation(s)

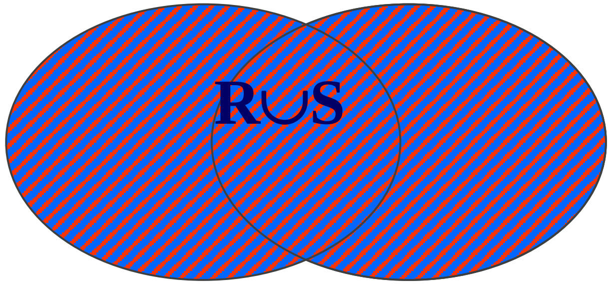
Relations are closed under the algebra

- Both the operands and the results are relations.

关系代数(Relational Algebra)

并(Union)

- The union to two relations R and S defines a relation that contains all the tuples of R, or S, or both R and S, duplicate tuples being eliminated.
- R and S must be union-compatible. – the arities (degree) of the two relations must be equivalent
- $R \cup S = \{t \mid t \in R \lor t \in S\}$



关系代数(Relational Algebra)

R

A	B	C
3	6	7
2	5	7
7	2	3
4	4	3

S

A	B	C
3	4	5
7	2	3

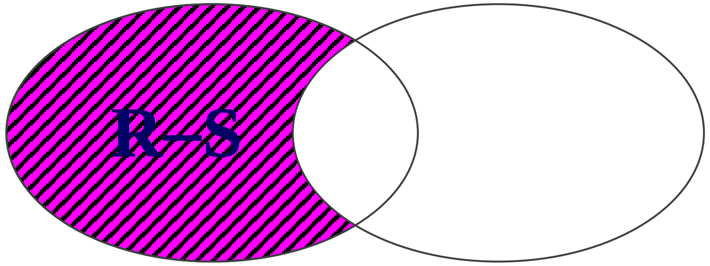
$R \cup S$

A	B	C
3	6	7
2	5	7
7	2	3
4	4	3
3	4	5

关系代数(Relational Algebra)

差(Difference)

- The set difference operation defines a relation consisting of the tuples that are in relation R, but not in S.
- R and S must be union-compatible.
- $R-S=\{t \mid t \in R \wedge t \notin S\}$



关系代数(Relational Algebra)

R

A	B	C
3	6	7
2	5	7
7	2	3
4	4	3

S

A	B	C
3	4	5
7	2	3

R-S

A	B	C
3	6	7
2	5	7
4	4	3

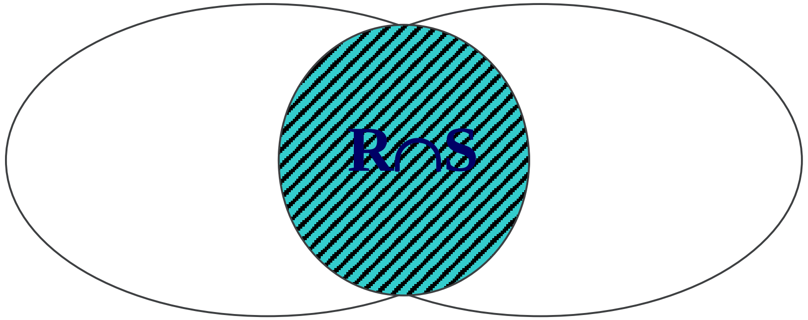
S-R

A	B	C
3	4	5

关系代数(Relational Algebra)

交(Intersection)

- The intersection operation defines a relation consisting of the set of all tuples that are in both R and S.
- R and S must be union-compatible.
- $R \cap S = \{t \mid t \in R \wedge t \in S\}$
- $R \cap S = R - (R - S)$



关系代数(Relational Algebra)

R

A	B	C
3	6	7
2	5	7
7	2	3
4	4	3

S

A	B	C
3	4	5
7	2	3

$R \cap S$

A	B	C
7	2	3

关系代数(Relational Algebra)

笛卡尔积(Cartesian product)

- The Cartesian product operation defines a relation that is the concatenation of every tuple of relation R with every tuple of relation S.

  $$R \times S = \{t \mid t = \langle t_r, t_s \rangle \wedge t_r \in R \wedge t_s \in S\}$$

- Let the arities of R and S be m and n. Let cardinalities of R and S be k1 and k2.
- The arity (degree) of the new relation will be m+n
- The cardinality of the new relation will be k1×k2

关系代数(Relational Algebra)

R

A	B
a	1
b	2

S

C	D	E
a	10	*x*
b	10	*x*
b	20	*y*
c	10	*y*

$R \times S$

A	B	C	D	E
a	1	*a*	10	*x*
a	1	*b*	10	*x*
a	1	*b*	20	*y*
a	1	*c*	10	*y*
b	2	*a*	10	*x*
b	2	*b*	10	*x*
b	2	*b*	20	*y*
b	2	*c*	10	*y*

关系代数(Relational Algebra)

- 选择(Selection or Restriction)
  - The selection operation works on a single relation R and defines a relation that contains only those tuples of R that satisfy the specified condition (predicate).
  - **$\sigma_F(R)=\{t \mid t \in R \wedge F(t)=\text{'true'}\}$**
  - The selection operate the relation from the view of rows.

关系代数(Relational Algebra)

R

A	B	C
3	6	7
2	5	7
7	2	3
4	4	3

$\sigma_{A<5}(R)$

A	B	C
3	6	7
2	5	7
4	4	3

$\sigma_{A<5 \wedge C=7}(R)$

A	B	C
3	6	7
2	5	7

关系代数(Relational Algebra)

投影(Projection)

- The projection operation works on a single relation R and defines a relation that contains a vertical subset of R, extracting the values of specified attributes and eliminating duplicates.
- **$\Pi_A(R) = \{ t[A] \mid t \in R \}$**
- The projection operate the relation from the view of columns.

关系代数(Relational Algebra)

R

A	B	C
a	b	c
d	e	f
c	b	c

$\Pi_{B,C}(R)$

B	C
b	c
e	f

$\Pi_{A}(R)$

A
a
d
c

关系代数(Relational Algebra)

连接(Join)

- The theta join operation defines a relation that contains tuples satisfying the predicate F from the Cartesian product of R and S. The predicate F is of the form R.aiθS.bi where θ may be one of the comparison operators (<, ≤, >, ≥, =,≠).

$$R \underset{A\theta B}{\bowtie} S = \{ \widehat{t_r t_s} \mid t_r \in R \wedge t_s \in S \wedge t_r[A] \theta t_s[B] \}$$

- We can rewrite the Theta join in terms of basic selection and Cartesian product operation.

$$R \underset{A\theta B}{\bowtie} S = \sigma_{r[A]\theta S[B]}(R \times S)$$

关系代数(Relational Algebra)

R

A	B	C
1	2	3
4	5	6
7	8	9

S

D	E
3	1
6	2

$R \underset{B<D}{\bowtie} S$

A	B	C	D	E
1	2	3	3	1
1	2	3	6	2
4	5	6	6	2

关系代数(Relational Algebra)

- 相等连接(Equijoin)
  - When θ is '=' in theta join. A particular type of Theta join

$$R \underset{A=B}{\bowtie} S = \{\widehat{t_r t_s} \mid t_r \in R \wedge t_s \in S \wedge t_r[A] = t_s[B]\}$$

- 自然连接(Natural join)
  - The natural join is an equijoin of two relations R and S over all common attributes x. One occurrence of each common attribute is eliminated from the result.

$$R \bowtie S = \{\widehat{t_r t_s} \mid t_r \in R \wedge t_s \in S \wedge t_r[x] = t_s[x]\}$$

  - When there are no common attributes in R and S, the equijoin is equal to Cartesian product.

关系代数(Relational Algebra)

R

A	B	C
1	2	3
4	5	6
7	8	9

S

C	D
3	1
6	2

$R \bowtie S$

A	B	C	D
1	2	3	1
4	5	6	2

R

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

S

B	D	E
1	a	1
3	a	2
1	a	3
2	b	4
3	b	5

$R \bowtie S$

A	B	C	D	E
1	1	1	a	1
1	1	1	a	3
1	1	3	a	1
1	1	3	a	3
4	2	2	b	5

关系代数(Relational Algebra)

自然连接中存在的一个问题

T

tNo	tName	salary
p01	赵明	8000
p02	钱丰	7000
p03	孙丽	6000
p04	李广	6000

C

cNo	cName	tNo
c01	离散数学	p01
c02	数据结构	p02
c03	数据库系统	p04

如果我们想获取所有教师及其所授课程信息，该如何做?

关系代数(Relational Algebra)

T

tNo	tName	salary
p01	赵明	8000
p02	钱丰	7000
p03	孙丽	6000
p04	李广	6000

$\bowtie$

C

cNo	cName	tNo
c01	离散数学	p01
c02	数据结构	p02
c03	数据库系统	p04

tNo	tName	salary	cNo	cName
p01	赵明	8000	c01	离散数学
p02	钱丰	7000	c02	数据结构
p04	李广	6000	c03	数据库系统

关系代数(Relational Algebra)

外连接(Outer join)

- Often in joining two relations, a tuple in one relation does not have a matching tuple in the other relation. We may want a tuple from one of the relation to appear in the result even when there is no matching value in the other relation.
- The (left) Outer join is a join in which tuples from R that do not have matching values in the common attributes of S are also included in the result relation.
- Missing values in the second relation are set to null.
- Left outer join, right outer join, full outer join

$R ⟕ S$ $R ⟖ S$ $R ⟗ S$

关系代数(Relational Algebra)

T

tNo	tName	salary
p01	赵明	8000
p02	钱丰	7000
p03	孙丽	6000
p04	李广	6000

⟕

C

cNo	cName	tNo
c01	离散数学	p01
c02	数据结构	p02
c03	数据库系统	p04

tNo	tName	salary	cNo	cName
p01	赵明	8000	c01	离散数学
p02	钱丰	7000	c02	数据结构
p04	李广	6000	c03	数据库系统
p03	孙丽	6000	null	null

关系代数(Relational Algebra)

像集(Images set)

- Relation R(X,Z), X and Z are sets of attributes of R. For each t[X] = x, we can define the images set of x Zx in R :
- $Z_x=\{t[Z]\,|\,t\in R, t[X]=x\}$
- 像集表示R中属性组X上值为x的诸元组在Z上分量的集合。

X 姓名	Z 课程
张军	物理
王红	数学
张军	数学

x=张军

Zx=张军

课程
物理
数学

张军同学所选修的全部课程

x=王红

Zx=王红

课程
数学

王红同学所选修的全部课程

关系代数(Relational Algebra)

除法(Division)

- 给定关系R(X,Y)和S(Y,Z)，其中X, Y, Z为属性组。 R中的Y与S中的Y可以有不同的属性名，但必须出自相同的域集。R与S的除运算得到一个新的关系P(X)，P是R中满足下列条件的元组在X属性列上的投影：元组在X上分量值x的象集Y_x包含S在Y上投影的集合。记作:

$$R \div S=\{t_r[X] \mid t_r \in R \wedge \Pi_Y(S) \subseteq Y_x\}$$

- We can rewrite the division expression using difference and Cartesian operation as follow:

$$R \div S=\Pi_X(R)-\Pi_X(\Pi_X(R)\times\Pi_Y(S)-R)$$

关系代数(Relational Algebra)

R

A	B	C	D
a	b	c	d
a	b	e	f
a	b	d	e
b	c	e	f
e	d	c	d
e	d	e	f

S

C	D
c	d
e	f

$R \div S = ?$

$CD_{(a,b)} = \{(c,d),(e,f),(d,e)\}$
$CD_{(b,c)} = \{(e,f)\}$
$CD_{(e,d)} = \{(c,d),(e,f)\}$
$\{(c,d),(e,f)\} \subseteq CD_{(a,b)}$
$\{(c,d),(e,f)\} \subseteq CD_{(e,d)}$

$\Rightarrow R \div S = \{(a,b),(e,d)\}$

关系代数(Relational Algebra)

R

A	B	C	D
a	b	c	d
a	b	e	f
a	b	d	e
b	c	e	f
e	d	c	d
e	d	e	f

S

C	D
c	d
e	f

$\Pi_{AB}(R)$

A	B
a	b
b	c
e	d

$\Pi_{AB}(R) \times \Pi_{CD}(S)$

A	B	C	D
a	b	c	d
a	b	e	f
b	c	c	d
b	c	e	f
e	d	c	d
e	d	e	f

$\Pi_{AB}(R) \times \Pi_{CD}(S)-R$

A	B	C	D
b	c	c	d

$R \div S=$

A	B
a	b
b	c
e	d

-

A	B
b	c

=

A	B
a	b
e	d

关系代数(Relational Algebra)

R

A	B	C
a1	b1	c2
a1	b2	c3
a1	b2	c1
a2	b3	c7
a2	b2	c3
a3	b4	c6
a4	b6	c6

S

B	C	D
b1	c2	d1
b2	c3	d2
b2	c1	d3
b1	c2	d4

R÷S=?

在关系R中，A可以取四个值{a1, a2, a3, a4}。其中：

a1的象集为{(b1,c2), (b2,c3), (b2,c1)}

a2的象集为{(b3,c7), (b2,c3)}

a3的象集为{(b4,c6)}

a4的象集为{(b6,c6)}

S在(B,C)上的投影为{(b1,c2),(b2,c3),(b2,c1)}

显然只有a1的象集$(B,C)_{a1}$包含S在(B,C)属性组上的投影，所以R÷S={a1}

关系代数(Relational Algebra)

除法的内涵（The indication of division）

姓名	课程
张军	物理
王红	数学
张军	数学
王红	物理

÷

课程
物理
数学

=

姓名
张军
王红

选修了全部课程的学生名单

姓名	课程	成绩
张军	物理	93
王红	数学	86
张军	数学	93
王红	物理	92

÷

课程
物理
数学

=

姓名	成绩
张军	93

选修了全部课程并且成绩相同的学生名单

关于本讲内容

祝各位学习愉快！

感谢观看！

讲解人：陆伟 教授