Тема 6. Математические основы реляционных баз данных

Языки запросов позволяют задавать операции над РБД и указывать предикаты для искомых данных.

Языки запросов делятся на два класса:

- Алгебраические языки, выражающие запросы в терминах операций над отношениями;
- Языки исчисления предикатов, в которых запрос является описанием искомых записей, задаваемым в форме предиката.

Рассмотрим реляционную алгебру. Будем предполагать, что доступ к компонентам кортежей возможен как по номерам компонентов, так и по именам атрибутов. Порядок компонентов предполагается существенным, все отношения конечны. Операндами реляционной алгебры являются постоянные или переменные отношения фиксированной арности.

Основные операции

1. Объединение

Объединение отношений R и S (обозначается $R \cup S$) есть множество кортежей, которые принадлежат либо R, либо S, либо им обоим. Арность R и S одинакова. В символической форме объединение записывается так:

$R \cup S = \{ t \mid t \in R \vee t \in S \}$.

2. Разность

Разностью отношений R и S (обозначается R-S) называется множество кортежей, принадлежащих R, но не принадлежащих S. Арность R и S одинакова. Символически: $R-S = \{ t \mid t \in R \ \& \ t \notin S \}$.

3. Декартово произведение

Пусть R и S – отношения арности r и s соответственно. Тогда декартовым произведением $R \times S$ отношений R и S называется множество кортежей длины r+s, первые r компонентов которых образуют кортежи из R, а последние s компонентов - кортежи из S.

$$R \times S = \{(a_1, a_2, \ldots a_r, a_{r+1}, a_{r+2} \ldots a_{r+s}) \mid (a_1, a_2, \ldots a_r) \in R \ \& \ (a_{r+1}, a_{r+2} \ldots a_{r+s}) \in S\}.$$

4. Проекция

Пусть R - отношение арности r. Проекцией R на компоненты $i_1,i_2,\ldots i_m$ называется множество кортежей вида $(a_1,a_2,\ldots a_m)$ таких, что существует кортеж $(b_1,b_2,\ldots b_r)\in R$, удовлетворяющий условию $a_j = b_{ij}$ (j=1 .. m). Проекция обозначается $\pi_{i_1 i_2 \cdots i_m}(R)$.

$$\pi_{i_1 i_2 \cdots i_m}(R) = \{(a_1, a_2, \ldots a_m) \mid \exists(b_1, b_2, \ldots b_r) \in R(a_j = b_{i_j}\ \ (j = \overline{1,m}))\}$$

Существо этой операции состоит в том, что из каждого кортежа отношения R формируется новый кортеж, состоящий из перечисленных компонентов в указанном порядке. Например, для построения $\pi_{2,1}(R)$ нужно из каждого кортежа $t\in R$ сформировать кортеж t', состоящий из второго и первого компонентов кортежа t в указанном порядке.

Если отношение имеет именованные столбцы (атрибуты), то вместо номеров компонентов можно использовать имена атрибутов. Например, для отношения R со схемой R(A,B,C,D) проекция $\pi_{B,A}(R)$ есть то же самое, что и $\pi_{2,1}(R)$.

5. Селекция

Пусть F - формула, операндами которой являются константы или номера компонентов, а операциями – сравнения и логические операции «&», «$\vee$», «$\neg$». Селекция отношения R по формуле F (обозначается $\sigma_F(R)$) есть множество кортежей t таких, что при подстановке i–го компонента t вместо каждого вхождения номера i в формулу F для всех i она станет истинной. Например, $\sigma_{2>3}(R)$ обозначает множество кортежей из R, второй компонент которых больше третьего. Селекция $\sigma_{1=Петя \vee 1=Вася}(R)$ есть множество кортежей из R, первый компонент которых имеет значение Петя или Вася. Формулы могут ссылаться на компоненты по именам атрибутов:

$$\sigma_F(R) = \{t \mid t \in R \,\&\, F(t)\}$$

Примеры:

R		
A	B	C
a	b	C
d	a	F
c	b	D

S		
D	E	F
b	G	a
d	A	f

R∪S		
a	b	c
d	a	f
c	b	d

R-S		
a	b	C
c	b	D

$\sigma_{B=b}$(R)		
A	B	C
a	b	c
c	b	d

R×S					
A	B	C	D	E	F
a	B	c	b	g	A
a	B	c	d	a	F
d	A	f	b	g	A
d	A	f	d	a	F
…	…	…	…	…	…

$\pi_{A,C}$(R)	
A	C
a	c
d	f
c	d

Дополнительные алгебраические операции

1. **Пересечение.**

R∩S = R- (R-S).

2. **Частное.**

Пусть R и S отношения арности r и s соответственно, r>s и S≠0. Тогда частное R и S (записывается R÷S) есть множество кортежей длины r-s таких, что для всех кортежей u∈S кортеж tu принадлежит R.

3. **Соединение**.

θ-соединение R и S по столбцам i и j (записывается $R \underset{i\,\theta\,j}{\bowtie} S$), где θ - операция сравнения, есть краткая запись для $\sigma_{i\theta(r+j)}(R \times S)$ (r – арность R). θ–

соединение R и S есть множество таких кортежей в декартовом произведении R и S, что i–ый компонент R находится в отношении θ с j–ым компонентом S.

4. Естественное соединение.

Естественное соединение R и S обозначается $R \rhd\lhd S$, требует именованных атрибутов. Вычисляется следующим образом:

- Вычисляется R×S;
- Для каждого атрибута A, имеющегося и в отношении R и в отношении S, выбираются те кортежи R×S, у которых совпадают значения в столбцах R.A и S.A;
- Для каждого указанного A удаляется S.A.

Формально, если $A_1, A_2, \ldots A_k$ имеются и в R и в S, то $R \rhd\lhd S$ есть $\pi_{i_1 i_2 \cdots i_m}(\sigma_{R.A_1 = S.A \& R.A_{21} = S, A_2 \& \ldots \& R.A_k = S.A_k}(R \times S))$, где $i_1 i_2 \cdots i_m$ - упорядоченный список всех компонентов R×S, за исключением компонентов $S.A_1, \ldots S.A_k$.

Примеры:

R			
a	b	C	d
a	b	E	f
b	c	E	f
e	d	C	d
e	d	E	f
a	b	D	e

S	
c	d
e	f

R÷S	
a	b
e	d

S	
D	E
3	1

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

6	2

$R \underset{B<D}{\rhd\lhd} S$				
A	B	C	D	E
1	2	3	3	1
1	2	3	6	2
4	5	6	6	2

R		
A	B	C
a	b	C
d	b	C
b	b	F
c	a	D

S		
B	C	D
b	c	d
b	c	e
a	d	b

$R \rhd\lhd S$			
A	B	C	D
a	b	c	d
a	b	c	e
d	b	c	d
d	b	c	e
c	a	d	b

Реляционное исчисление

Исчисление с переменными-кортежами

Выражение реляционного исчисления с переменными-кортежами {t| ψ(t)} представляет множество кортежей t, удовлетворяющих условию ψ(t). Формула ψ строится из атомов и операций.

Атомы:

1. R(s), где R - отношение, s - переменная-кортеж. Атом утверждает, что s - кортеж отношения R.
2. s[i] θ u[j], где s,u - переменные кортежи, θ - операция сравнения. Атом утверждает, что i–ый компонент s находится в отношении θ с j–ым компонентом u. Например, s[1]<u[3] означает, что 1-ый компонент s меньше 3-го компонента u.
3. s[i]θa и aθs[i], где a – константа.

Формулы:

1. Атом – формула;
2. Если φ, ψ - формулы, то φ & ψ, $\varphi \vee \psi$, $\neg\varphi$ - формулы;
3. Если φ – формула, то $\exists s\varphi$, $\forall s\varphi$ - формулы.

Здесь $\exists$ – квантор существования, $\forall$ - квантор общности. Вхождение переменной s в формулу φ в последнем случае считается связанным квантором. Вхождение переменной, не попадающее в область действия никакого квантора, - свободное. Подстановка значений возможна только для свободных вхождений переменных.

4. Порядок старшинства операций: сравнение, $\exists$, $\forall$, $\neg$, $\wedge$, $\vee$.

В выражении реляционного исчисления с переменными-кортежами $\{t|\psi(t)\}$ переменная-кортеж t - единственная свободная в ψ переменная.

Примеры.

1. $R \cup S = \{t|R(t) \vee S(t)\}$, отношения R и S имеют одинаковую арность;
2. $R\text{-}S = \{t|R(t) \& \neg S(t)\}$;
3. Если R и S - отношения арности r и s, то

$R \times S = \{t^{(r+s)}|\exists u^{(r)} \exists v^{(s)}(R(u) \& S(v) \& t[1]=u[1] \& \ldots \& t[r+1]=v[1] \& \ldots \& t[r+s]=v[s])\}$

4. $\pi_{i_1 i_2 \ldots i_k}(R) = \{t^{(k)}|\exists u(R(u) \& t[1]=u[i_1] \& \ldots \& t[k]=u[i_k])\}$;
4. $\sigma_F(R)=\{t|R(t) \& F'\}$, где F' образуется из F заменой номера i на выражение t[i].

Безопасные выражения

Определенное выше исчисление с переменными-кортежами позволяет представлять бесконечные отношения, например $\{t|\neg R(t)\}$, которое невозможно реализовать физически. Для устранения бесконечных отношений множество допустимых выражений реляционного исчисления ограничивают безопасными выражениями $\{ t \mid \psi(t)\}$, для которых можно показать, что каждый компонент любого t, удовлетворяющего $\psi(t)$, должен быть элементом специального множества DOM(ψ). Множество DOM(ψ) определяется как множество символов, которые либо явно появляются в ψ, либо служат компонентами какого-либо кортежа в некотором отношении, упоминаемом в ψ. Множество DOM(ψ) конечно.

Редукция реляционной алгебры к реляционному исчислению с переменными-кортежами.

Справедливо утверждение: для каждого выражения реляционной алгебры существует эквивалентное ему безопасное выражение реляционного исчисления с переменными-кортежами.

Реляционное исчисление с переменными на доменах.

Исчисление с переменными на доменах строится аналогично исчислению с переменными-кортежами:

1. В исчислении с переменными на доменах вместо переменных-кортежей используются переменные на доменах, представляющие компоненты кортежей.
2. Атомы:
   a. $R(x_1,x_2,\ldots x_k)$, где R – отношение, x_i - константа или переменная на домене. Атом утверждает, что $(x_1,x_2,\ldots x_k)$ – кортеж R.
   b. $x\theta y$, где x,y – константы или переменные на доменах, θ - сравнение. Выражение утверждает, что x находится в отношении θ к y.
3. Формулы используют связки &, $\vee$, $\neg$ и кванторы $(\exists x)(\forall x)$, где x – переменная на домене.

Понятия свободного и связанного вхождения переменных, безопасного выражения определяются так же, как и в исчислении с переменными-кортежами. Выражения исчисления с переменными на доменах имеют вид $\{x_1,x_2,\ldots x_k|\psi(x_1,x_2,\ldots x_k)\}$, где ψ - формула со свободными переменными $x_1,x_2,\ldots x_k$.

Сведение исчисления с переменными–кортежами к исчислению с переменными на доменах.

Пусть задано выражение с переменными-кортежами $\{t|\psi(t)\}$. Переход к эквивалентному выражению с переменными на доменах производится по следующим правилам.

Если t имеет арность k, введем k новых переменных на доменах $t_1,t_2,\dots t_k$ и заменим исходное выражение следующим: $\{t_1,t_2,\dots t_k|\psi'(t_1,t_2,\dots t_k)\}$. Формула ψ' получается из формулы ψ, в которой любой атом R(t) заменяется атомом $R(t_1,t_2,\dots t_k)$, а каждое свободное вхождение t[i] – переменной t_i.

Далее для каждого квантора $(\exists u)(\forall u)$ вводим m новых переменных на доменах $u_1,u_2,\dots u_m$ (m - арность u). В области действия квантора заменяем u[i]на u_i и R(u) на $R(u_1,u_2,\dots u_m)$. Заменяем знаки $(\exists u)(\forall u)$ на $(\exists u_1)\dots$ $(\exists u_m)$ $((\forall u_1)$ $\dots(\forall u_m))$. В результате получаем выражение исчисления с переменными на доменах, эквивалентное исходному выражению.

Справедливо утверждение: для каждого безопасного выражения реляционного исчисления с переменными кортежами существует эквивалентное ему безопасное выражение исчисления с переменными на доменах.

Можно также показать, что для каждого безопасного выражения реляционного исчисления с переменными на доменах существует эквивалентное ему выражение реляционной алгебры.

Практические задания для самостоятельного выполнения

Выполнить следующие операции над двумя совместными отношениями R1, R2 с идентичной структурой.

R1

ФИО	Год рождения	Должность	Кафедра
Иванов И.И.	1968	Зав. кафедрой	22
Сидоров С.С.	1973	Доцент	22
Козлов К.К.	1980	Ассистент	23

R2

ФИО	Год рождения	Должность	Кафедра
Цветкова Н.Н.	1965	Доцент	23
Петрова П.П.	1973	Ст. преподаватель	22
Козлов К.К.	1980	Ассистент	23

1. Объединение. В результате операции построить новое отношение **R = R1 U R2**
2. Пересечение. Результирующее отношение **RP = R1 ∩ R2**
3. Вычитание. В результате строится новое отношение **RV = R1 - R2**