

Phép tính quan hệ

Nguyễn Khắc Văn
vannk@hcmup.edu.vn

Nội dung chi tiết

- Giới thiệu
- Phép tính quan hệ trên bộ
- Phép tính quan hệ trên miền

Giới thiệu

❖ Đại số quan hệ (Relational Algebra)

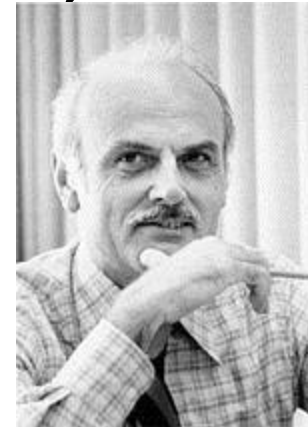
- Là tập hợp các phép toán cơ sở của mô hình dữ liệu quan hệ, đóng vai trò là cơ sở hình thức cho các phép toán của mô hình quan hệ.
- Đại số quan hệ được dùng để giải thích các truy vấn SQL được đánh giá như thế nào. DBMS thường dùng đại số quan hệ như ngôn ngữ trung gian bậc cao dùng để dịch query trước khi tối ưu hóa thực thi

❖ Xét về mặt khái niệm, thì SQL lại dựa vào 1 ngôn ngữ truy vấn chính quy hoàn toàn khác (formal query language)

Relational Calculus (phép tính quan hệ)

Giới thiệu (tt)

- Là ngôn ngữ truy vấn hình thức
- Do Codd đề nghị vào năm 1972, “Data Base Systems”, Prentice Hall, p33-98
- Đặc điểm
 - Phi thủ tục, gần với ngôn ngữ tự nhiên
 - Dựa vào lý thuyết logic
 - Rút trích cái gì (what) \neq rút trích như thế nào (how)
 - Khả năng diễn đạt tương đương với ĐSQH



Giới thiệu (tt)

Đại số quan hệ (relational algebra) có tính thủ tục, gần với ngôn ngữ lập trình

vs

Phép tính quan hệ (relational calculus) không có tính thủ tục và gần với ngôn ngữ tự nhiên hơn

Ví dụ: xét câu truy vấn sau sau “liệt kê các sinh viên học khoa CNTT”

Giới thiệu (tt)

Nếu theo đại số quan hệ ta thực hiện theo các bước sau:

1. Tạo mối kết nối tự nhiên của 2 quan hệ SINHVIEN và KHOA trên thuộc tính #KHOA;
2. Thu hẹp kết quả của kết nối này chỉ còn các bộ liên quan đến khoa 'Công Nghệ Thông Tin';
3. Dùng phép chiếu (project) để kết quả chỉ còn lại các thuộc tính cần lấy (#SINHVIEN).

Nếu theo phép tính quan hệ thì:

1. Tìm các sinh viên trong quan hệ SINHVIEN sao cho tồn tại liên kết đến khoa 'Công Nghệ Thông Tin'.

Giới thiệu (tt)

❖ Một số khái niệm logic toán học

- **Mệnh đề:** là các khẳng định có giá trị chân lý xác định
- **Vị từ:** Là một khẳng định $P(x, y, \dots)$, với x, y là các biến trên miền xác định A, B, \dots
 - $P(x, y, \dots)$ không phải là mệnh đề
 - Thay x, y, \dots bằng các giá trị cụ thể ta được một mệnh đề
 - x, y, \dots là các biến tự do
- **Lượng từ**
 - Mệnh đề " $\forall x \in A, P(x)$ " và " $\exists x \in A, P(x)$ " là các lượng từ hóa của vị từ $P(x)$.
 - Trong đó \forall là *lượng từ phổ dụng*, \exists là *lượng từ tồn tại*
 - x không còn là biến tự do, nó bị buộc bởi các lượng từ \forall, \exists

Giới thiệu (tt)

- Có 2 loại
 - Phép tính quan hệ trên bộ (Tuple Rational Calculus)
 - SQL
 - Phép tính quan hệ trên miền (Domain Rational Calculus)
 - QBE (Query By Example)

Nội dung chi tiết

- Giới thiệu
- **Phép tính quan hệ trên bộ**
- Phép tính quan hệ trên miền

Phép tính quan hệ trên bộ

- Biểu thức phép tính quan hệ trên bộ có dạng

$$\{ t.A \mid P(t) \}$$

- t là biến bộ
 - Biến nhận giá trị là một bộ của quan hệ trong CSDL
 - $t.A$ là giá trị của bộ t tại thuộc tính A
- P là công thức có liên quan đến t
 - $P(t)$ có giá trị ĐÚNG hoặc SAI phụ thuộc vào t
- Kết quả trả về là tập các bộ t sao cho $P(t)$ đúng

Ví dụ 1

- Tìm các nhân viên có lương trên 30000

$$\{ t \mid \underbrace{t \in \text{NHANVIEN}}_{P(t)} \wedge \underbrace{t.\text{LUONG} > 30000}_{P(t)} \}$$

- $t \in \text{NHANVIEN}$ đúng
 - Nếu t là một thể hiện của quan hệ NHANVIEN
- $t.\text{LUONG} > 30000$ đúng
 - Nếu thuộc tính LUONG của t có giá trị trên 30000

Ví dụ 2

- Cho biết mã và tên nhân viên có lương trên 30000
- Tìm những bộ t thuộc NHANVIEN có thuộc tính lương lớn hơn 30000
- Lấy ra các giá trị tại thuộc tính MANV và TENNV

$\{ t.MANV, t.TENNV \mid t \in NHANVIEN \wedge t.LUONG > 30000 \}$

- Tập các MANV và TENNV của những bộ t sao cho t là một thể hiện của NHANVIEN và t có giá trị lớn hơn 30000 tại thuộc tính LUONG

Ví dụ 3

- Cho biết các nhân viên (MANV) làm việc ở phòng ‘Nghien cuu’

$t.MANV \mid t \in NHANVIEN$

$s \in PHONGBAN \wedge s.TENPHG = \text{'Nghien cuu'}$

- Lấy ra những bộ t thuộc $NHANVIEN$
- So sánh t với một bộ s nào đó để tìm ra những nhân viên làm việc ở phòng ‘Nghien cuu’
- Cấu trúc “tồn tại” của phép toán logic

$$\exists t \in R (Q(t))$$

Tồn tại 1 bộ t thuộc quan hệ R sao cho vị từ $Q(t)$ đúng

Ví dụ 3

- Cho biết các nhân viên (MANV) làm việc ở phòng 'Nghiên cứu'

$\{ t.MANV \mid t \in NHANVIEN \wedge$

$\exists s \in PHONGBAN ($

$s.TENPHG = \text{'Nghiên cứu'} \wedge$
 $s.MAPHG = t.PHG) \}$

Q(s)

Ví dụ 4

- Cho biết tên các nhân viên (TENNV) tham gia làm đề án hoặc có thân nhân

$$\{ t.TENNV \mid t \in NHANVIEN \wedge (\\ \exists s \in PHANCONG (t.MANV = s.MA_NVIEN) \vee \\ \exists u \in THANNHAN (t.MANV = u.MA_NVIEN)) \}$$

Ví dụ 5

- Cho biết tên các nhân viên (TENNV) vừa tham gia làm đề án vừa có thân nhân

$$\{ t.TENNV \mid t \in NHANVIEN \wedge (\\ \exists s \in PHANCONG (t.MANV = s.MA_NVIEN) \wedge \\ \exists u \in THANNHAN (t.MANV = u.MA_NVIEN)) \}$$

Ví dụ 6

- Cho biết tên các nhân viên (TENNV) tham gia làm đề án mà không có thân nhân nào

$$\{ t.TENNV \mid t \in NHANVIEN \wedge \\ \exists s \in PHANCONG (t.MANV = s.MA_NVIEN) \wedge \\ \neg \exists u \in THANNHAN (t.MANV = u.MA_NVIEN) \}$$

Ví dụ 7

- Với mỗi đề án ở 'TP HCM' cho biết mã đề án, mã phòng ban chủ trì và tên người trưởng phòng

$$\{ s.MADA, s.PHONG, t.TENNV \mid s \in DEAN \wedge t \in NHANVIEN \wedge \\ s.DDIEM_DA = 'TP\ HCM' \wedge \\ \exists u \in PHONGBAN (s.PHONG = u.MAPHG \wedge \\ u.TRPHG = t.MANV) \}$$

Ví dụ 8

- Tìm các nhân viên (MA_NVIEN) tham gia vào tất cả các đề án
- Cấu trúc “với mọi” của phép toán logic

$$\forall t \in R (Q(t))$$

Q đúng với mọi bộ t thuộc quan hệ R

Ví dụ 8 (tt)

- Tìm các nhân viên (MANV, HONV, TENNV) tham gia vào tất cả các đề án

$$\{ t.MANV, t.HONV, t.TENNV \mid t \in NHANVIEN \wedge$$
$$\forall s \in DEAN (\exists u \in PHANCONG ($$
$$u.SODA = s.MADA \wedge$$
$$t.MANV = u.MA_NVIEN)) \}$$

Ví dụ 9

- Tìm các nhân viên (MANV, HONV, TENNV) tham gia vào tất cả các đề án do phòng số 4 phụ trách
- Cấu trúc “kéo theo” của phép tính logic

$$P \Rightarrow Q$$

Nếu P thì Q

Ví dụ 9 (tt)

- Tìm các nhân viên (MANV, HONV, TENNV) tham gia vào tất cả các đề án do phòng số 4 phụ trách

$$\{ t.MANV, t.HONV, t.TENNV \mid t \in NHANVIEN \wedge$$
$$\forall s \in DEAN ($$
$$s.PHONG = 4 \Rightarrow (\exists u \in PHANCONG ($$
$$u.SODA = s.MADA \wedge$$
$$t.MANV = u.MA_NVIEN)))) \}$$

Định nghĩa hình thức

- Một công thức truy vấn tổng quát có dạng

$$\{ t_1.A_i, t_2.A_j, \dots t_n.A_k \mid P(t_1, t_2, \dots, t_n) \}$$

- t_1, t_2, \dots, t_n là các biến bộ
- A_i, A_j, \dots, A_k là các thuộc tính trong các bộ t tương ứng
- P là công thức
 - P được hình thành từ những công thức nguyên tố

Biến bộ


- Biến tự do (free variable)

$\{ t \mid t \in \text{NHANVIEN} \wedge t.\text{LUONG} > 30000 \}$


t là biến tự do

- Biến kết buộc (bound variable)

$\{ t \mid t \in \text{NHANVIEN} \wedge \exists s \in \text{PHONGBAN} (s.\text{MAPHG} = t.\text{PHG}) \}$


Biến tự do


Biến kết buộc

Công thức nguyên tố

- (i) $t \in R$

- t là biến bộ
- R là quan hệ

$$t \in \text{NHANVIEN}$$

- (ii) $t.A \theta s.B$

- A là thuộc tính của biến bộ t
- B là thuộc tính của biến bộ s
- θ là các phép so sánh $<, >, \leq, \geq, \neq, =$

$$t.\text{MANV} = s.\text{MANV}$$

- (iii) $t.A \theta c$

- c là hằng số
- A là thuộc tính của biến bộ t
- θ là các phép so sánh $<, >, \leq, \geq, \neq, =$

$$s.\text{LUONG} > 30000$$

Công thức nguyên tố (tt)

- Mỗi công thức nguyên tố đều mang giá trị ĐÚNG hoặc SAI
 - Gọi là chân trị của công thức nguyên tố
- Công thức (i)
 - Chân trị ĐÚNG nếu t là một bộ thuộc R
 - Chân trị SAI nếu t không thuộc R

R	A	B	C
	α	10	1
	α	20	1

$t1 = \langle \alpha, 10, 1 \rangle$

$t2 = \langle \alpha, 20, 2 \rangle$

$t1 \in R$ có chân trị ĐÚNG

$t2 \in R$ có chân trị SAI

Công thức nguyên tố (tt)

- Công thức (ii) và (iii)
 - Chân trị tùy thuộc vào việc thay thế giá trị thật sự của bộ vào vị trí biến bộ

R	A	B	C
	α	10	1
	α	20	1

Nếu t là bộ $\langle \alpha, 10, 1 \rangle$

Thì $t.B > 5$ có chân trị ĐÚNG ($10 > 5$)

Qui tắc

- (1) Mọi công thức nguyên tố là công thức
- (2) Nếu P là công thức thì
 - $\neg P$ là công thức
 - (P) là công thức
- (3) Nếu P_1 và P_2 là các công thức thì
 - $P_1 \vee P_2$ là công thức
 - $P_1 \wedge P_2$ là công thức
 - $P_1 \Rightarrow P_2$ là công thức

Qui tắc (tt)

- (4) Nếu $P(t)$ là công thức thì
 - $\forall t \in R (P(t))$ là công thức
 - Chân trị ĐÚNG khi $P(t)$ ĐÚNG với mọi bộ t trong R
 - Chân trị SAI khi có ít nhất 1 bộ làm cho $P(t)$ SAI
 - $\exists t \in R (P(t))$ là công thức
 - Chân trị ĐÚNG khi có ít nhất 1 bộ làm cho $P(t)$ ĐÚNG
 - Chân trị SAI khi $P(t)$ SAI với mọi bộ t trong R

Qui tắc (tt)

- (5) Nếu P là công thức nguyên tố thì
 - Các biến bộ t trong P là biến tự do
- (6) Công thức $P = P_1 \wedge P_2$, $P = P_1 \vee P_2$, $P = P_1 \Rightarrow P_2$
 - Sự xuất hiện của biến t trong P là tự do hay kết buộc phụ thuộc vào việc nó là tự do hay kết buộc trong P_1, P_2

Một số biến đổi

- (i) $P_1 \wedge P_2 = \neg (\neg P_1 \vee \neg P_2)$
- (ii) $\forall t \in R (P(t)) = \neg \exists t \in R (\neg P(t))$
- (iii) $\exists t \in R (P(t)) = \neg \forall t \in R (\neg P(t))$
- (iv) $P \Rightarrow Q = \neg P \vee Q$

Công thức an toàn

- Xét công thức

$$\{ t \mid \neg(t \in \text{NHANVIEN}) \}$$

- Có rất nhiều bộ t không thuộc quan hệ NHANVIEN
 - Thậm chí không có trong CSDL
 - Kết quả trả về không xác định
-
- Một công thức P gọi là an toàn nếu các giá trị trong kết quả đều lấy từ miền giá trị của P
 - Dom(P)
 - Tập các giá trị được đề cập trong P

Công thức an toàn (tt)

- Ví dụ

$$\{ t \mid t \in \text{NHANVIEN} \wedge t.\text{LUONG} > 30000 \}$$

- $\text{Dom}(t \in \text{NHANVIEN} \wedge t.\text{LUONG} > 30000)$
- Là tập các giá trị trong đó
 - Có giá trị trên 30000 tại thuộc tính LUONG
 - Và các giá trị khác tại những thuộc tính còn lại
- Công thức trên là an toàn

Nội dung chi tiết

- Giới thiệu
- Phép tính quan hệ trên bộ
- **Phép tính quan hệ trên miền**

Phép tính quan hệ trên miền

- Biểu thức phép tính quan hệ trên miền có dạng

$$\{ x_1, x_2, \dots, x_n \mid P(x_1, x_2, \dots, x_{n+m}) \}$$

- x_1, x_2, \dots, x_{n+m} là các biến miền
 - Biến nhận giá trị là một miền giá trị của một thuộc tính
- P là công thức theo x_1, x_2, \dots, x_n
 - P được hình thành từ những công thức nguyên tố
- Kết quả trả về là tập các giá trị x_1, x_2, \dots, x_n sao cho khi các giá trị được thay thế cho các x_i thì P đúng

Ví dụ 3

- Cho biết mã và tên nhân viên có lương trên 30000

$$\{ r, s \mid \exists x (\\ \langle p, q, r, s, t, u, v, x, y, z \rangle \in \text{NHANVIEN} \wedge \\ x > 30000) \}$$

Ví dụ 4

- Cho biết các nhân viên (MANV) làm việc ở phòng 'Nghien cuu'

$$\{ s \mid \exists z ($$
$$\langle p, q, r, s, t, u, v, x, y, z \rangle \in \text{NHANVIEN} \wedge$$
$$\exists a, b (\langle a, b, c, d \rangle \in \text{PHONGBAN} \wedge$$
$$a = \text{'Nghien cuu'} \wedge b = z)) \}$$

Ví dụ 10

- Cho biết các nhân viên (MANV, HONV, TENNV) không có thân nhân nào

$$\{ p, r, s \mid \exists s (\\ \langle p, q, r, s, t, u, v, x, y, z \rangle \in \text{NHANVIEN} \wedge \\ \neg \exists a (\langle a, b, c, d, e \rangle \in \text{THANNHAN} \wedge a = s)) \}$$

Công thức nguyên tố

- (i) $\langle x_1, x_2, \dots, x_n \rangle \in R$
 - x_i là biến miền
 - R là quan hệ có n thuộc tính

- (ii) $x \theta y$
 - x, y là các biến miền
 - Miền giá trị của x và y phải giống nhau
 - θ là các phép so sánh $<, >, \leq, \geq, \neq, =$

- (iii) $x \theta c$
 - c là hằng số
 - x là biến miền
 - θ là các phép so sánh $<, >, \leq, \geq, \neq, =$

Nhận xét

- Một công thức nguyên tố mang giá trị ĐÚNG hoặc SAI với một tập giá trị cụ thể tương ứng với các biến miền
 - Gọi là chân trị của công thức nguyên tố
- Một số qui tắc và biến đổi tương tự với phép tính quan hệ trên bộ

Công thức an toàn

- Xét công thức

$$\{ p, r, s \mid \neg (<p, q, r, s, t, u, v, x, y, z> \in \text{NHANVIEN}) \}$$

- Các giá trị trong kết quả trả về không thuộc miền giá trị của biểu thức
- Công thức không an toàn

Công thức an toàn (tt)

- Xét công thức

$$\{ x \mid \underbrace{\exists y (<x, y> \in R)}_{\text{Công thức 1}} \wedge \underbrace{\exists z (\neg <x, z> \in R \wedge P(x, z))}_{\text{Công thức 2}} \}$$

Công thức 1

Công thức 2

- R là quan hệ có tập các giá trị hữu hạn
- Cũng có 1 tập hữu hạn các giá trị không thuộc R
- Công thức 1: chỉ xem xét các giá trị trong R
- Công thức 2: không thể kiểm tra khi không biết tập giá trị hữu hạn của z

Công thức an toàn (tt)

- **Nhận xét**

Ở phép tính quan hệ có biến là bộ, để biết được biểu thức có an toàn hay không ta giới hạn các giá trị phải thuộc 1 miền giá nào đó.

VS

Nhưng ở phép tính quan hệ có biến là miền, ta không thể giới hạn miền, mà phải thêm một số qui tắc để định nghĩa an toàn.

Công thức an toàn (tt)

- Biểu thức

$$\{ x_1, x_2, \dots, x_n \mid P(x_1, x_2, \dots, x_n) \}$$

được gọi là an toàn nếu:

- Những giá trị xuất hiện trong các bộ của biểu thức phải thuộc về miền giá trị của P
- Lượng từ \exists : biểu thức $\exists x (Q(x))$ đúng khi và chỉ khi xác định được giá trị của x thuộc $\text{dom}(Q)$ làm cho $Q(x)$ đúng
- Lượng từ \forall : biểu thức $\forall x (Q(x))$ đúng khi và chỉ khi $Q(x)$ đúng với mọi giá trị của x thuộc $\text{dom}(Q)$



Những nội
dung chính
cần nhớ?

Nguyễn Khắc Văn
vannk@hcmup.edu.vn

Nội dung tự học

- Tối ưu hóa câu truy vấn (thầy gửi)
- Trigger (sql – khoa học tự nhiên)