THƯ VIỆN
ĐẠI HỌC NHA TRANG
M
005.74
Ng 527 H

GUYỄN XUÂN HUY - LÊ HOÀI BẮC

Bài tập CƠ SỞ DỮ LIỆU



Chào mùng bạn đã đến với thư viện của chúng tôi

Xin vui lòng:

- Không xé sách
- Không gạch, viết, vẽ lên sách



NHÀ XUẤT BẢN THỐNG KÊ - 2003

NGUYỄN XUÂN HUY - LÊ HOÀI BẮC

BÀI TẬP CƠ SỞ DỮ LIỆU

TRUONG DAI HOC MHA TRUME

M124 39

NHÀ XUẤT BẢN THỐNG KÊ HÀ NỘI - 2003

WÁC TÁC

Lời nói đầu	5
Phần thứ nhất: <i>Tóm tắt lý thuyết và bài tậ</i> p	9
Chương 1: Quan hệ và đại số quan hệ	11
Chương 2: Các thao tác trên bộ và quan hệ	27
Chương 3: Ngôn ngữ hỏi SQL	31
Chương 4: <i>Phụ thuộc hàm</i>	35
Chương 5: Chuẩn hoá	51
Một số đề thi	57
Phần thứ hai: <i>Bài giải</i>	71
Bài giải chương 1: Quan hệ và đại số quan hệ	73
Bài giải chương 2: Các thao tác trên bộ và quan hệ	79
Bài giải chương 3: <i>Ngôn ngữ hỏi SQL</i>	91
Bài giải chương 4: <i>Phụ thuộc hàm</i>	101
Bài giải chương 5: <i>Chuẩn hoá</i>	127
Bài giải các đề thi	135

LỜI NÓI ĐẦU

Khác với toán học, trong tủ sách tin học nước nhà, ta chỉ thấy một số sách bài tập lập trình. Đó chắc chắn là một thiệt thời cho sinh viên và các bạn tự học.

Cuốn Bài tập cơ sở dữ liệu này là một thử nghiệm nhằm trợ giúp các bạn trẻ một phương thức tự kiểm tra và đánh giá tri thức ban đầu, mức nhập môn, về một lĩnh vực chiếm vị trí đáng nói trong quá trình phát triển của công nghệ thông tin.

Những năm gần đây, trong các kỳ thi tốt nghiệp đại học, thi chuyển đổi, thi tuyển cao học và nghiên cứu sinh đều có mảng về cơ sở dữ liệu. Đó là điều dễ hiểu, vì cơ sở dữ liệu la phần không thể thiếu trong các hệ thống tin học hoá.

Trong phương án đầu tiên của cuốn sách chúng tôi chọn lọc và đề xuất một số bài tập thuộc năm mảng tri thức sau đây: đại số quan hệ, các phép toán trên bộ, ngôn ngữ hỏi SQL, phụ thuộc hàm và chuẩn hoá. Mỗi mảng tri thức được trình bày thành ba phần: Phần thứ nhất bao gồm một số điều tóm tắt về lý thuyết. Phần tiếp theo là các bài tập, cuối cùng là các bài giải. Dấu * được dùng để ghi chú các bài tập ở mức nâng cao.

Phần cuối sách chúng tôi tuyển chọn và giới thiệu một số đề thi tuyển cao học và nghiên cứu sinh để bạn đọc làm quen với các nội dung tổng hợp.

Mục tiêu cuối cùng của việc ra bài tập là giúp cho người học hiểu sâu và kỹ hơn về các khái niệm đã học. Để đạt được điều này mong bạn đọc đừng bỏ qua bài tập nào. Với các bài dễ, bạn có thể giải trong một vài phút. Với các bài khó, trong lần luyện tập thứ nhất bạn có thể bỏ qua. Sau một vài lần thử sức, tin rằng bạn sẽ hoàn toàn làm chủ được các khái niệm liên quan đến cơ sở dữ liệu.

Chúng tôi cho rằng các tài liệu sau đây sẽ giúp ích bạn đọc tra cứu các nguồn trí thức cơ sở:

- 1. Date C. J., *Nhập môn các hệ cơ sở dữ liệu*, Những người dịch: Hồ Thuần, Nguyễn Quang Vinh, Nguyễn Xuân Huy, NXB Thống Kê, Hà Nội, Tập I (1985), Tập II (1986).
 - 2. Nguyễn Xuân Huy, Thuật toán, NXB Thống Kê, Hà Nội, 1987.
- 3. Vũ Đức Thi, *Cơ sở dữ liệu:Kiến thức và thực hành*, NXB Thống Kê, Hà Nội, 1997.
- 4. Lê Tiến Vương, *Nhập môn cơ sở dữ liệu quan hệ*, Tái bản lần thứ 4, NXB Thống Kê, Hà Nội, 1999.
- 5. Garcia-Molina H., Ullman J., Widom J., Database System: The Complete Book, Prentice Hall, 2002.
- 6. Maier D., *The Theory of Relational Database*, Computer Science Press, Rockville, Md, 1983.
- 7. Ullman, J., *Principles of Data-base and Knowledge-base Systems*, (Second Edition), Computer Science Press, Potomac, Md., 1982, (Có bản dịch tiếng Việt của Trần Đức Quang.)

Người đầu tiên định hướng cho chúng tôi tìm hiểu về cơ sở dữ liệu và luôn luôn khuyến khích chúng tôi học tập và trao đổi kiến thức là giáo sư Hồ Thuần, Viện Công nghệ Thông tin.

Cuốn sách này được khởi thảo và hoàn thành theo phương án đầu tiên là nhờ nhiệt tinh đóng góp về ý tưởng, nội dung và thẩm định của các đồng nghiệp của chúng tôi. Giáo sư Lê Tiến Vương, Tổng cục Địa chính, giáo sư Hoàng Kiếm, giáo sư Trần Vĩnh Phước, Đại học Quốc gia thành phố Hồ Chí Minh đã thảo luận chi tiết về những nội dung cơ bản và kiến trúc cho tập sách.

Đặc biệt, các đồng nghiệp trẻ, giáo sư Vũ Ngọc Loãn, Đại học Quốc gia Hà Nội, giáo sư Nguyễn Thanh Thuỷ, Đại học Bách khoa Hà Nội, tiến sỹ Trình Đình Thắng. Đại học Sư pham Hà Nội II, tiến sỹ Dương Anh Đức tiến sỹ Đỗ Văn Nhơn, thac sỹ Nguyễn Tấn Trần Minh Khang, Đại học Quốc gia thành phố Hồ Chí Minh, thac sỹ Nguyễn Xuân Tùng, Trung tâm Tin học Bưu điện Hà Nội, thạc sỹ Nguyễn Ngọc Hà, Trung tâm Tin học Bưu điện Hải Phòng, thạc sỹ Trịnh Thanh Lâm, Intel, thạc sỹ Nguyễn Xuân Hoàng, Misa Group đã có những góp ý cụ thể về nội dung chương trình đào tạo và các yêu cầu thực tiễn của cơ sở dữ liệu. Các cử nhân Bùi Thuý Hằng và Trần Quốc Dũng, Viện Công nghệ Thông tin đã giúp chúng tôi đọc lại và chỉnh sửa các trang bản thảo.

Chúng tôi chân thành cảm ơn những đóng góp vô giá của các đổng nghiệp.

Chúng tôi mong rằng sẽ tiếp tục nhận được những ý kiến chỉ giáo của bạn đọc gần xa về nội dung và cấu trúc của tập sách.

Cát Bà, Mùa Hoa Phượng, 2003 Các tác giả NGUYỄN XUÂN HUY - LÊ HOÀI BẮC

TÓM TẮT LÝ THUYẾT VÀ BÀI TẬP

Chương 1 QUAN HỆ VÀ ĐẠI SỐ QUAN HỆ

TÓM TẮT LÝ THUYẾT

Quan hê

Cho tập hữu hạn $U = \{A_1, A_2, ..., A_n\}$ khác trống $(n \ge 1)$. Các phần tử của U được gọi là thuộc tính. Ứng với mỗi thuộc tính $A_i \in U$, i = 1,2,..., n có một tập không rỗng $dom(A_i)$ được gọi là miển trị của thuộc tính A_i .

Một quan hệ R với các thuộc tính $U = \{ A_1, A_2, ..., A_n \}$, ký hiệu là R(U), là một tập các ánh xạ $t: U \to D$ sao cho với mỗi $A_i \in U$ ta có $t(A_i) \in dom(A_i)$. Mỗi ánh xạ được gọi là một bộ của quan hệ R.

Mỗi quan hệ R(U) có hình ảnh là một bảng, mỗi cột ứng với một thuộc tính, mỗi dòng là một bộ.

Ta ký hiệu t / U là một bộ trên tập thuộc tính U.

Một *quan hệ trống*, kỷ hiệu Ø, là quan hệ không chứa bộ nào.

Chú ý

Vì mỗi quan hệ là một tập các bộ nên trong quan hệ *không có hai bộ* trùng lặp.

Các ký hiệu cơ bản

Theo truyền thống của lý thuyết cơ sở dữ liệu chúng ta chấp nhận các quy định sau đây:

- Các thuộc tính được ký hiệu bằng các chữ LATIN HOA đầu bảng chữ A, B, C,...
- Tập thuộc tính được ký hiệu bằng các chữ LATIN HOA cuối bảng chữ X, Y, Z,...
- Các thuộc tính trong một tập được liệt kê như một xâu ký tự, không có các dấu biểu diễn tập, chẳng hạn ta viết X = ABC thay vì viết X = { A, B, C }. XY biểu diễn hợp của hai tập thuộc tính X và Y, X \cup Y. Phép trừ hai tập hợp X và Y được ký hiệu là X-Y hoặc X \ Y.
- Các bộ được biểu diễn bằng các chữ Latin thường có thể kèm chỉ số t, u, v, t₁...
- Với mỗi bộ t trong quan hệ R(U) và mỗi tập con các thuộc tính
 X ⊆ U ta ký hiệu t[X] hoặc t.X là hạn chế của bộ t trên tập thuộc tính X.
- Hàm Attr(R) cho tập thuộc tính của quan hệ R.
- Hàm Card(R) cho lực lượng (số bộ) của quan hệ R.
- Trong trường hợp tập thuộc tính U đã cho trước ta có thể viết đơn giản R thay cho R(U).
- Ký hiệu REL(U) là tập toàn thể các quan hệ trên tập thuộc tính U.

Hai quan hệ R và S được gọi là tương thích nếu chúng có cùng một tập thuộc tính, tức là nếu Attr(R) = Attr(S).

Với mỗi bộ u trong quan hệ R(U) và mỗi bộ v trong quan hệ S(V) ta ký hiệu u*v là phép dán bộ. u*v cho ta bộ t trên tập thuộc tính UV thoả điều kiên: t.U = u và t.V = v.

Với mỗi bộ u trong quan hệ R(U) và với mỗi quan hệ S(V) ta ký hiệu u*S là phép dán bộ u với quan hệ S. u*S cho ta quan hệ

$$P(UV) = \{ u*v \mid v \in S \}$$

Để thể hiện các phép toán quan hệ ta sẽ dùng các ký pháp tựa như ký pháp của hệ ISBL (Information System Base Language).

Đại số quan hệ

Phép chọn (phép lọc)

Cho quan hệ R(U) và biểu thức điều kiện (còn gọi là biểu thức lọc hay biểu thức chọn) e. Phép chọn trên quan hệ R theo điều kiện e, ký hiệu R(e) cho ta quan hệ:

$$P(U) = R(e) = \{ t \in R \mid Sat(t, e) \}.$$

trong đó hàm logic Sat(t, e) kiểm tra bộ t thoả điều kiện e được xác định như sau:

- Thay mọi xuất hiện của mỗi thuộc tính A trong biểu thức chọn e bằng trị tương ứng của A trong bộ t, t.A, ta thu được một mênh để logic b.
- Tính trị của b. Nếu là đúng (True) thì bộ t thoả điều kiện e; ngược lại, nếu trị của b là sai (False) thì bộ t không thoả điều kiện e.

Trong các biểu thức chọn ta sử dụng ký hiệu cho các phép toán logic như sau:

- Tích: & hoặc AND
- Tổng: | hoặc OR

Phủ định: ! hoặc NOT

Kéo theo: ⇒ hoặc IMPLY

Phép chiếu

Phép chiếu quan hệ R(U) trên tập con thuộc tính $X \subseteq U$, ký hiệu R[X], cho ta quan hệ

$$P(X) = R[X] = \{t.X \mid t \in R\}$$

R[X] được tính theo 2 bước như sau:

- Xoá các cột không thuộc X của bảng R,
- Xoá bớt các dòng giống nhau trong bảng kết quả: chỉ giữ lại một dòng trong số các dòng giống nhau.

Phép kết nối tự nhiên

Phép kết nối (tự nhiên) hai quan hệ R(U) và S(V), ký hiệu $R \star S$, cho ta quan hệ chứa các bộ được dán từ các bộ u của quan hệ R với mỗi bộ v của quan hệ S sao cho các trị trên miền thuộc tính chung (nếu có) của hai bộ này giống nhau.

$$P(UV) = R*S = \{u*V \mid u \in R, v \in S, u.M = v.M, M = U \cap V\}$$

Nếu $M = U \cap V = \emptyset$, R*S sẽ cho ta tích Descartes trong đó mỗi bộ của quan hệ R sẽ được ghép với mọi bộ của quan hệ S.

Ph**ép cộng** (hợp)

Phép cộng (hợp theo lý thuyết tập hợp hoặc kết nối dọc) hai quan hệ tương thích R(U) và S(U), ký hiệu R+S, cho ta quan hệ chứa các bộ của mỗi quan hệ thành phần,

$$P(U) = R + S = \{t \mid t \in R \lor t \in S\}$$

Phép trừ

Phép trừ (theo lý thuyết tập hợp hoặc *lấy phần riêng*) hai quan hệ tương thích R(U) và S(U), ký hiệu R-S, cho ta quan hệ chứa các bộ của quan hệ R không có trong quan hệ S,

$$P(U) = R-S = \{t \mid t \in R, t \notin S\}$$

Phép giao

Phép giao (theo lý thuyết tập hợp hoặc lấy phần chung) hai quan hệ tương thích R(U) và S(U), ký hiệu R&S, cho ta quan hệ chứa các bộ xuất hiện đồng thời trong cả hai quan hệ thành phần,

$$P(U) = R\&S = \{t \mid t \in R, t \in S\}$$

Các phép toán cộng, trừ và giao đựcc gọi là các phép toán tập hợp trên các quan hệ (tương thích).

Phép chia

Cho hai quan hệ R(U) và S(V). Phép chia quan hệ R cho quan hệ S, ký hiệu R:S, cho ta quan hệ

$$P(M) = R : S = \{ \underline{t} M \mid \underline{t} \in \mathbb{R}, (\underline{t} M) * S \subseteq R, M = U - V \}$$

Thứ tự thực hiện các phép toán quan hệ

Trong một biểu thức quan hệ các phép toán *một ngôi* có độ *ưu tiên* cao hơn (do đó được thực hiện sớm hơn) các phép toán hai ngôi. Tiếp đến là nhóm các phép toán kết nổi, giao và chia, cuối cùng là nhóm các phép toán cộng và trừ. Thứ tự ưu tiên từ cao đến thấp của các phép toán quan hệ được liệt kê như sau:

(), []

* ,&, :

Dãy các phép toán cùng thứ tự ưu tiên được thực hiện lần lượt từ trái qua phải. Nếu biểu thức quan hệ có chứa các cặp ngoặc () thì các biểu thức con trong các cặp ngoặc được thực hiện trước.

Một số hàm tiện ích

- 1. Sum(R,A): cho tổng các giá trị số trong thuộc tính (cột) A của quan hệ R, $Sum(R,A) = \Sigma (t.A \mid t \in R)$
- Avg(R,A): cho trung bình cộng các giá trị trong thuộc tính (cột)
 A của quan hệ R, Avg(R,A) = Sum(R,A) / Card(R) nếu
 Card(R) ≠ 0
- 3. Max(R,A): cho giá trị lớn nhất trong thuộc tính (côt) A của quan hệ R.
- 4. Min(R,A): cho giá trị nhỏ nhất trong thuộc tính (cột) A của quan hệ R.

Nếu trong biểu thức quan hệ có chứa các hàm tiện ích thì các hàm này được thực hiện *sớm nhất* trong ngữ cảnh cho phép.

Thí dụ

Biểu thức quan hệ P = S*R(A > Avg(S,A))[AB] sẽ được thực hiện theo trật tự sau đây:

- 1. Tính hàm c = Avg(S,A)
- 2. Thực hiện phép chọn $P_1 = R(A > c)$
- 3. Thực hiện phép chiếu $P_2 = P_1 [AB]$

Thực hiện phép kết nối P = S*P₂

Chú ý: Trong một số tài liêu có sử dụng ký pháp khác cho các phép toán quan hệ như sau

Phép toán	Ký hiệu	Ký hiệu khác
chọn	R(e)	σ _e (<i>R</i>)
chiếu	R[X]	$\pi_{X}(R)$
kết nối tự nhiên	R * S	R⋈S
cộng	R+S	R∪S
giao	R&S	R∩S
trừ	R-S R\S	
chia	R:S R÷S	

Cơ sở dữ tiệu minh họa: CSDL Thực tập

Hầu hết bài tập trong chương này liên quan đến CSDL Thực tập gồm ba quan hệ sau đây:

SV(SV#, HT, NS, QUE, HL)

DT(DT#, TDT, CN, KP)

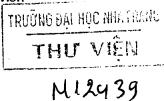
SD(SV#, DT#, NTT, KM, KQ)

Quan hê SV(SV#, HT, NS, QUE, HL) chứa thông tin về các sinh viên trong một lớp của một trường đại học,

SV - tên quan hệ sinh viên

SV# - mã số sinh viên

HT - ho và tên sinh viên



NS - năm sinh của sinh viên

QUE - quê (tỉnh)

HL - học lực thể hiện qua điểm trung bình

 Quan hệ DT(DT#, TDT, CN, KP) chứa thông tin về các đề tài nhà trường quản lý,

DT - tên quan hệ để tài

DT# - mã số đề tài

TDT - tên đề tài

CN - họ và tên chủ nhiệm đề tài

KP - kinh phí cấp cho để tài (triệu đồng).

 Quan hệ SD(SV#, DT#, NTT, KM, KQ) chứa thông tin về tình hình thực tập của các sinh viên theo các để tài,

SD - tên quan hệ sinh viên - để tài

SV# - mã số sinh viên

DT# - mã số để tài mà sinh viên đó tham gia

NTT - nơi thực tập để triển khai đề tài (tỉnh)

KM - khoảng cách từ nơi thực tập đến trường

KQ - kết quả thực tập theo đề tài đã chon

- Giả thiết là một sinh viên có thể tham gia nhiều để tài, mỗi để tài sinh viên đó thực tập tại một địa điểm.
- Với mỗi câu hỏi, yêu cầu trả lời bằng một biểu thức của đại số quan hệ. Tuổi được tính đến năm 2003. Thí du,

Câu hỏi

Cho danh sách các sinh viên trẻ (dưới 18 tuổi tính đến năm 2003), học và thực tập đều đạt loại khá/giỏi (điểm không dưới 8.5)

Trả lời

BÀI TẬP

1.1. Cho hai quan hệ R(A,B,C,D) và S(C,D) như sau

Hãy xác định:

- a) R[AB]
- b) R(3-B+D>1)
- c) R(B < 4) + R(D > 3)
- d) $R(B >= 1 \& B \le 5)$
- e) R*S[C]
- f) R(B < 4) R(D > 3)
- g) R(B < 4) & R(D > 3).
- h) R:S

- 1.2. Cho thông tin về những sinh viên sinh trước năm 1983, quê ở Hải Phòng?
- 1.3. Cho danh sách các tỉnh có sinh viên đến thực tập?
- 1.4. Cho biết các địa điểm thực tập xa trường (KM > 100) của đề tài số 7?
- 1.5. Cho thông tin về việc thực tập tại Nha Trang của các sinh viên?
- *1.6. Cho danh sách sinh viên thực tập tại quê nhà?
- 1.7. Cho thông tin về các để tài có sinh viên thực tập?
- 1.8. Cho biết mã của các đề tài không có sinh viên nào tham gia?
- 1.9. Cho biết mã của những đề tài có kinh phí 1.5 triệu và những đề tài có kinh phí trên 2 triệu?
- **1.10.** Cho biết mã của những sinh viên dưới 20 tuổi, thực tập khá (có điểm kết quả trên 7)?
- *1.11. Cho biết mã của những để tài có địa bàn thực tập ít ra là như đề tài 1.
- *1.12. Cho danh sách những đề tài được triển khai thực tập ở tất cả các tỉnh có sinh viên thực tập.
- *1.13. Cho danh sách những sinh viên thực tập theo đề tài có kinh phí lớn hơn một phần năm tổng kinh phí cấp cho các đề tài.
- *1.14. Cho danh sách các sinh viên có điểm học tập cao hơn điểm thực tập trung bình của để tài mã số 4.
- 1.15. Một phép toán 2 ngôi θ có tính chất giao hoán nếu:

$$(\forall R, S)$$
: $R\theta S = S\theta R$

hoặc cả hai vế đồng thời không có nghĩa

Chứng minh rằng các phép toán quan hệ kết nối, cộng và giao có tính giao hoán.

- 1.16. Tìm thí dụ chứng tỏ các phép toán trừ và chia không có tính giao hoán.
- 1.17. Cho quan hệ R(U) và hai biểu thức chọn e và h. Chứng minh
 - a) R(e & h) = R(h & e)
 - b) $R(e \& h) = R(e) \& R(h) \subseteq R(e)$
 - c) $R(e \& h) = R(h) \& R(e) \subseteq R(h)$
 - d) R(e & h) = R(e)(h)
 - e) R(e | h) = R(h | e)
 - f) $R(e \mid h) = R(e) + R(h)$
 - g) R(! e) = R R(e)
 - h) R(True) = R
 - i) $R(False) = \emptyset$
- **1.18.** Cho quan hệ R(U), các biểu thức chọn e, h trên U và tập con các thuộc tính $X \subseteq U$. Ký hiệu Attr(e) là tập các thuộc tính của U có mặt trong e. Chứng minh, nếu Attr $(e) \subseteq X$ thì
 - (a) R(e)[X] = R[X](e)
 - b) R / h & e)[X] = R(h)(e)[X] = R(h)[X](e)
- *1.19. Chứng minh rằng phép chia có thể được biểu diễn qua các phép chiếu, kết nối và trừ như sau,

$$R: S = R[M] - (R[M]*S - R)[M]$$

trong đó M = Attr(R) - Attr(S).

1.20. Phép toán quan hệ được gọi là đóng nếu với mọi quan hệ đầu vào ta đều thu được đầu ra là một quan hệ. Cho biết tính đóng (ghi có / không) của các phép toán quan hệ

Phép toán	Ký hiệu	Tính đóng	
chọn	()		
chiếu	[]		
kết nối tự nhiên	*		
cộng	+		
giao	&		
trừ	-		
chia	:		

1.21. Phép toán 2 ngôi θ có tính chất kết hợp nếu:

$$(\forall R, S, T)$$
: $(R \theta S) \theta T = R \theta (S \theta T)$

hoặc cả hai vế đồng thời không có nghĩa.

Chứng minh rằng các phép toán *kết nối*, *cộng* và *gi*ao của đại số quan hệ có tính chất *kết hợp*.

- 1.22. Tìm thí du chứng tỏ các phép toán trừ và chia không có tính kết hợp.
- 1.23. Chứng minh rằng với mọi cặp quan hệ tương thích R và S ta có

$$R - (R - S) = R \& S$$

1.24. Chứng minh rằng với mọi quan hệ R(U), mọi tập con X trong U và mọi biểu thức điều kiện e ta có

a)
$$R(e)(e) = R(e)$$

b)
$$R[X][X] = R[X]$$

1.25. Chứng minh rằng với mọi quan hệ R(U) ta có

a)
$$R * R = R$$

b)
$$R + R = R$$

c)
$$R \& R = R$$

d)
$$R - R = \emptyset$$

e)
$$R: R = \emptyset$$
, $Attr(R: R) = \emptyset$.

1.26. Phép toán quan hệ được gọi là nở (co) ngang nếu quan hệ kết quả có số thuộc tính nhiều hơn (ít hơn) các quan hệ đầu vào, được gọi là nở (co) dọc nếu quan hệ kết quả có số bộ nhiều hơn (ít hơn) các quan hệ đầu vào. Hãy đánh dấu (+), (-) hoặc (=) để khẳng định tính nở hoặc co hoặc không nở/co của mỗi phép toán tương ứng.

Phép toán	Ký hiệu	Nở/Co ngang	Nở/Co dọc
chọn	()		
chiếu	[]		
kết nối tự nhiên	*		
cộng	+		
giao	&	,	
trừ	-		
chia	:-		