MTA SZTAKI Közlemények 34/1986 pp.123-149

ЭКВИВАЛЕНТНОСТЬ ОДНОГО КЛАССА РЕЛЯЦИОННЫХ ВЫРАЖЕНИЙ

До Суан Тхо

ВВЕДЕНИЕ

На основе реляционной алгебры /relational algebra/ введенной доктором Коддом /1.3./, запросы использователей базы данных /БД/ могут быть выражены в виде реляционных выражений. Поэтому в области оптимизации обработки запросов в БД большое внимание уделялось вопросу представления реляционных выражений и создания эквивалентных преобразований реляционных выражений.

В /4/, /5/ было использовано табло /tableaux/ как двухверное представление одного класса запросов использователей соответствующего ограниченным реляционным выражениям /restricted relational expressions/. Ограниченным реляционным выражением называется выражение обладающее тремя операциями реляционной алгебры: выбор /Select/, проектирование /Project/ и соединение /Join/. /В /4/ его называют выражением SPJ./

Табло /обозначенное T/ - двухмерная матрица, каждый столбец которой соответствует одному из атрибутов универсального отношения в фиксированном порядке. Первая строка называется сводкой. Символы в табло выражены из:

1/ - Характерных переменных /distinguished variables/,

2/ - Нехарактерных переменных /nondistinguished variables/,

3/ - Констант,

4/ - Бланков.

Каждое ограниченное реляционное выражение может быть представлено в виде соответствующего табло. В /5/ были предложены методы эквивалентных преобразований любого табло к виду табло T_O с минимальным числом строк. Как известно, число строк в табло больше чем число операций соединения в ограниченных реляционных выражениях на единицу и, осуществление операции соединения в вычислительной машине требует много времени и объема памяти. Поэтому этим методам было уделено большое значение для оптимальной обработки реляционного вырежения. Несмотря на то, понятие табло написано в /4/, /5/ имеет некоторое ограничение. С его помощью лишь только представит в виде табло реляционное выражение обладающее операцией выбора вида $\sigma_{A=c}(r)$.

Практически нам приходилось обрабатывать наиболее сложные запросы. Рассмотрим следующие примеры /см. /6//.

Пусть в БД хранятся сведения о служащих и их учреждениях. Универсальное отношение R определено на атрибутах Н /номер служащего/, З /зарплата/, У /номер учреждения/, Ф /фамилия служащего/, Д /должность/, В /название учреждения/, А /адрес/, т.е.

R = R (Н,Ф,Д,З,У,В,А)
R1=R1 (Н,Ф,Д,З,У) - схема отношения СЛУЖАЩИЕ
R2=R2 (У,В,А) - схема отношения УЧРЕЖДЕНИЕ.

Первый запрос: получить номер, фамилию, должность тех служащих, которые работают в учреждении № 19 и имеют зарплату больше 1500. 1-й запрос представим в виде реляционного выражения:

$$\Pi_{Н,Ф,Д}(\sigma_{З>1500 \wedge У=19}(R1))$$

Второй вопрос: получить номер, фамилию, должность тех служащих, работающих в учреждениях № 17, 19, 32.

2-ой запрос представим в виде:

$$\Pi_{Н,Ф,Д}(\sigma_{У=17 \vee У=19 \vee У=32}(R1))$$

Легко видеть, что 1-й, 2-ой запросы не могут буть представлены в виде табло.

В этой работе мы введем новое понятие - обобщенное табло /обозначение τ /.

Введенное нами понятие τ позволяет развивать результаты опубликованные в /4/, /5/ для ограниченного реляционного выражения обладающего операцией выбора таких видов:

а/ $\sigma_{A\theta c}(r)$ б/ $G_{A\in} \{a_1, a_2, \ldots, a_k\}$

где A - атрибут, c - константа, $\theta \in \{=, \neq, <, \leq, >, \geq\}$, $a_i \in \mathrm{dom}(A)$

Будем обозначать в общем виде $\sigma_{A\in\xi}$ для случаев а/ и б/, ξ - множество значений и $\xi \subseteq \mathrm{dom}(A)$.

Работа состоит из 4 основных разделов. В пером разделе повторены основные понятия реляционной модели. Во втором разделе введены понятия обобщенного табло τ и правила построения табло соответствующего ограниченному реляционному выражению E с операцией выбора $\sigma_{A\in\xi}$. В третьем разделе будут описаны понятия эквивалентности и эквивалентное преобразование двух обобщенных табло. В четвертом разделе будем показывать связь между табло T и обобщенным табло τ .

1. Основные понятия реляционной модели

В реляционном БД объекты описываются при помощи характеризующих признаков, которые называются атрибутами. Каждый атрибут имеет имя и множество значений. Множество значений атрибута A называется его доменом и обозначается dom(A).

$R = \{A_1, A_2, \ldots, A_n\}$ - множество имен атрибутов отношения r. R называют схемой отношения r. Под понятием отношения понимают любое подмножество декартова произведения доменов, т.е.

$$r(A_1, \ldots, A_n) \subseteq \mathrm{dom}(A_1) \times \ldots \times \mathrm{dom}(A_n)$$

Элементы отношения $(a_1, \ldots, a_n)$, $a_i \in \mathrm{dom}(A_i)$ является кортежем.

Значение атрибута A в кортежи μ обозначено $\mu[A]$.

Будем обозначать операцию выбора через $\sigma_{A\in\xi}$, A – атрибут схемы отношения r, $\xi \subseteq dom(A)$.

$$\sigma(r)_{A\in\xi} = \{\mu\in r \mid \mu[A] \in\xi\}.$$

Пусть $X \subseteq R$. Операция проектирования отношения r на X /обозначена $\Pi_X(r)$/ определяется следующим образом:

$$\Pi_X(r) = \{\mu[X] \mid \mu\in r\}.$$

Если отношения r_1 и r_2 имеются, схемы R_1, R_2 соответственно. Тогда операция естественного соединения /natural join/ отношения r_1 и r_2 обозначена $r_1 \times r_2$.

$$r_1 \times r_2 = \{\mu \mid /\mu \text{ - кортеж отношения со схемой } R_1 \text{ и } R_2/$$

$$\wedge(\exists \upsilon_1\in r_1 \wedge \upsilon_2\in r_2 : \upsilon_1[R_1] = \mu[R_1]\wedge\upsilon_2[R_2] = \mu[R_2])\}$$

Пусть отношения r_1 и r_2 имеют общую схему отношения. Будем обозначать операцию объединения двух отношений r_1 и r_2:

$$r_1 \cup r_2 = \{\mu \mid \mu\in r_1 \vee \mu\in r_2\}.$$

Операция пересечения отношений r_1 и r_2 обозначается $r_1 \cap r_2$

$$r_1 \cap r_2 = \{\mu \mid \mu\in r_1 \wedge \mu\in r_2\}.$$

<u>З А М Е Ч А Н И Е</u>: Если r_1 и r_2 имеют общую схему, то $r_1 \times r_2 = r_1 \cap r_2$. Поэтому считаем, что операция пересечения – частным случаем операции естественного соединения.

Операция вычитания отношения r_1 и r_2 /обозначим $r_1 - r_2$/ определяется:

$$r_1 - r_2 = \{\mu \mid \mu\in r_1 \wedge \mu\bar{\in} r_2\}.$$

Под понятием реляционного выражения понимают выражение, обладающее операндами - реляционными схемами, операторами - операциями реляционной алгебры. В /3/ Кодд определил, что множество операции реляционной алгебры обладает реляционной полнотой.

Нам известно, что реляционное выражение может описывать много довольно сложных и разнообразных запросов пользователей, в которых операции выбора, проектирования и соединения имеют охватывающее значение.

Поэтому в этой работе мы будем рассматривать ограниченное реляционное выражение, в котором имеются только операции выбора, проектирования и соединения.

В /4/, /5/ были введены понятия о усиленной эквивалентности /strong equivalence/ и слабой эквивалентности /weak equivalence/ двух реляционных выражений.

Пусть E реляционное выражение с операндами-схемами отношений $R_1, R_2, \ldots, R_n$. Присваиванием α называется замена каждой схемы R_i $(i = \overline{1,n})$ в E ее соответствующим содержанием r_i, т.е.

$$\alpha = R_i \rightarrow r_i, \quad i = (\overline{1,n}).$$

Будем обозначать $V_\alpha(E)$ - значение реляционного выражения E соответствующее присваиванию α . А V(E) есть отображение присваивания α для операндов в выражении E на значение выражения $V_\alpha(E)$, т.е.

$$V(E) : \alpha \rightarrow V_\alpha(E).$$

Тогда два реляционных выражения E_1 и E_2 считаются усиленно эквивалентными, если $V_\alpha(E_1) = V_\alpha(E_2)$, $\forall\alpha$ или $V(E_1) = V(E_2)$.

Понятие о слабой эквивалентности основано на предположении существования универсального отношения на множестве атрубутов $\bigcup_{i=1}^{n} R_i$, и для определенного состояния I универсального отношения каждой схеме R_i присваивает соответственным значениям $r_i = \Pi_{R_i}(I)$.

Обозначить $V_I(E)$ - значение реляционного выражения E при состоянии I. $V_I(E)$ определяется индуктивно по числу реляционной операции в выражении E следующим образом:

1. Если в E имеется только схема отношения R_i, то:

$$V_I(E) = \Pi_{R_i}(I)$$

2а. Если $E = \sigma_{A_i \in \xi}(E_1)$, то

$$V_I(E) = G_{A_i \in \xi}(V_I(E_1)).$$

2б. Если $E = \Pi_X(E_1)$, то

$$V_I(E) = \Pi_X(V_I(E_1)).$$

2в. Если $E = E_1 \times E_2$, то

$$V_I(E) = V_I(E_1) \times V_I(E_2).$$

2г. Если $E = E_1 \cup E_2$, то

$$V_I(E) = V_I(E_1) \cup V_I(E_2).$$

2д. Если $E = E_1 - E_2$, то

$$V_I(E) = V_I(E_1) - V_I(E_2).$$

Рассматривается как отображение из состояний I универсального отношения в отношение $V_I(E)$.

Если $\forall I$, $V_I(E_1) = V_I(E_2)$, то говоря, что E_1 слабо эквивалентно /или эквивалентно/ E_2. Обозначают $E_1 \equiv E_2$.

В этой работе исследуется случай слабой эквивалентности. Но полученные результаты могут обобщать и на случай усиленной эквивалентности.

2. Понятие обобщенного табло

Обобщенным табло /обозн. τ / называется двухмерная матрица, каждый столбец которой соответствует одному из атрибутов универсального отношения в фиксированном порядке. Первая строка называется сводкой.

Символы матрицы могут быть:

1. Характерные переменные /обозначаются a_i/
2. Нехарактерные переменные /обозначены b_i/
3. Символы множеств значений ξ , принадлежащих доменам атрибутов универсального отношения /обозн. $\tilde{\xi}$ /
4. Бланки.

<u>З А М Е Ч А Н И Е</u>: В случае, когда каждое множество значений ξ имеет только один элемент, то получается табло T. Поэтому можно рассматривать T как частный обобщенного табло τ .

Пусть τ обобщенное табло , множество символов появляющихся в τ . Под понятием оценки /Evaluation/ ρ для τ понимают присваивание соответственного значения каждому символу из , и это присваивание значения удовлетворяет слудующим условиям:

а/ Если $\tilde{\xi} \in$ - символ множества значения, то $\tilde{\xi}$ присваивается одно значение $c \in \xi$. Будем обозначать $\rho(\tilde{\xi}) = c$.

Если $\xi = \emptyset$, то $\rho(\tilde{\xi}) = \emptyset$ и табло τ обладающее пустым множеством - пустое. Пустое табло /обозн. $\emptyset$/ отображает из любого состояния I универсального отношения в пустое отношение.

б/ Если $w = v_1 v_2 \ldots v_n$ - одна строка табло τ , то $\rho(w) = \rho(v_1) \ldots \rho(v_n)$. Множество атрибутов соответствующих столбцам не имеющим бланк на сводке τ называются целевой схемой отношения /target relation scheme/.

Значение табло τ соответствующее одному состоянию I универсального отношения /обозн. τ (I)/ опредяется следующим образом:

$\tau(I) = \{ \rho(w_o) \mid$ для некоторой оценки ρ имеют $\rho(w_i) \in I$, $i = \overline{1,m}\}$, где w_o - сводка, w_i - строки табло τ.

<u>Правила построения обобщенного табло</u>

Пусть E ограниченное реляционное выражение. Тогда обобщенное табло τ для E строится индуктивно по числу операции в E следующим образом:

1. Если в E нет никакой операции, то E - схема отношения R. Тогда табло τ состоит из сводки и одной строки.
   а/ Если A_i - атрибут схемы R, то в столбце A_i на сводке и на строке имеется одинаковая характерная переменная a_i.
   б/ Если A_i не является атрибутом схемы R, то в столбце на сводке состоит бланк, и на строке ставить новую нехарактерную переменную.

2. Пусть $E = \sigma_{A_i \in \xi}(E_1)$. τ_1 обобщенное табло построенное для E_1. Тогда $\tau = \rho_{A_i \in \xi}(\tau_1)$ получается из τ_1 следующим образом:
   а/ Если в столбце A_i сводки τ_1 имеется бланк, то выражение E не имеет никакой смысл, и табло τ для E не определено.
   б/ Если в столбце A_i сводки τ_1 имеется символ множества значения $\bar{\xi}_1$ тогда:
      (i) Если $\xi_1 \cap \xi = \emptyset$, то $\forall I$:
      $V_I(E) = G_{A_i \in \xi}(V_I(E_1)) = \emptyset$, отсюда $\tau = \emptyset$.
      (ii) Если $\xi_1 \cap \xi \neq \emptyset$, то табло τ получается из τ_1 заменой $\bar{\xi}_1$ на $\bar{\xi}_2$ ($\xi_2 = \xi_1 \cap \xi$) в местах где $\bar{\xi}_1$ появляется в столбце A_i.
   в/ Если τ_1 имеет характерную переменную a_i в столбце A_i сводки, то τ получается из τ_1 заменой a_i на $\bar{\xi}$ в местах, где a_i появляется в столбце A_i.

3. Пусть имеется $E = \Pi_x(E_1)$ и τ_1 - обобщённое табло для E_1. Табло $\tau = \Pi_x(\tau_1)$ получается из τ_1 следующим образом: в столбцах соответствующих атрибутам невходящих в X все символы сводки заменяются бланками, а характерные переменные на строках заменяются новыми нехарактерными переменными.

4. Пусть $E = E_1 \times E_2$ и τ_1, τ_2 два обобщенных табло для E_1, E_2 соответственно. Тогда $\tau = \tau_1 \times \tau_2$ для E строится следующим образом:

а/ Если в столбце A_i сводки табло τ_1 имеет символ множества $\bar{\xi}_1$, а табло τ_2 имеет $\tilde{\xi}_2$ и $\xi_1 \cap \xi_2 = \emptyset$. Тогда $\forall I$, $V_I(E) = \Phi$, отсюда $\tau = \emptyset$.

б/ Пусть $_1$ и $_2$ множества символов двух табло τ_1 и τ_2 соответственно. Ненарушая общность мы предполагаем, что τ_1 и τ_2 имеют одинаковые характерные переменные в столбцах соответствующих одним атрибутам, τ_1 и τ_2 имеют непересекающие множества нехарактерных переменных. Тогда табло τ включает в себя все строки табло τ_1 и τ_2 с символами определяющими по следующим правилам:

(i) Если в столбце A_i сводки обе τ_1, τ_2 или одно из них имеется символ множества значений $\tilde{\xi}$ то в столбце A_i сводки табло τ ставит символ $\tilde{\xi}$. Все характерные переменные в столбце A_i табло τ заменяются символом $\tilde{\xi}$.

(ii) Если в столбце A_i сводки τ_1 имеет символ множества значений $\tilde{\xi}_1$, τ_2 имеет символ множества $\tilde{\xi}_2 \cdot \xi_1 \cap \xi_2 \neq \emptyset$. Тогда в соответственном столбце A_i табло τ на сводке поставить символ множества $\tilde{\xi}$, $\xi = \xi_1 \cap \xi_2$, и во всех местах где появляются $\tilde{\xi}_1$, $\tilde{\xi}_2$ заменяют символом $\tilde{\xi}$.

(iii) Если в столбце A_i обе τ_1, τ_2 или одно из них имеет характерную переменную a_i и два предыдущих правила не используется, то на сводке τ поставить характерную переменную a_i.

(iv) Для остальных случаев на сводке τ поставить бланки.

Пример:

1-й запрос /см. введение/ представить в виде реляционного выражения

$$E = \Pi_{Н,Ф,Д}(\sigma_{З\in\xi_1 \wedge У\in\xi_2}(R_1)),$$

где $\xi_1 = \{C/C > 1500 \wedge C \in dom(З)\}$, $\xi_2 = \{19\}$, тогда по правилам построения τ для E имеем:

$$\tau = \begin{array}{|ccccccc|} \multicolumn{1}{c}{Н} & Ф & Д & З & У & В & \multicolumn{1}{c}{А} \\ \hline a_1 & a_2 & a_3 & & & & \\ \hline a_1 & a_2 & a_3 & \tilde{\xi}_1 & \tilde{\xi}_2 & b_1 & b_2 \\ \hline \end{array}$$

Лемма 1: Пусть τ_1, τ_2 два обобщенных табло. Для любого состояния I универсального отношения имеют место следующие равенства:

1^o/ $\sigma_{A_i\in\xi}(\tau(I)) = (\sigma_{A_i\in\xi}(\tau))(I)$,

2^o/ $\Pi_X(\tau(I)) = (\Pi_X(\tau))(I)$,

3^o/ $\tau(I) \times \tau_1(I) = (\tau \times \tau_1)(I)$.

Доказательство:

1. Доказательство 1-ого равенства производится по двух шагам:

a/ $\sigma_{A_i\in\xi}(\tau(I)) \subseteq (\sigma_{A_i\in\xi}(\tau))(I)$.

Пусть $\sigma_{A_i\in\xi}(\tau(I)) \neq \emptyset$. В этом случае надо доказать, что $\forall\mu \neq \emptyset$, если $\mu\in \sigma_{A_i\in\xi}(\tau(I))$, то $\mu\in(\sigma_{A_i\in\xi}(\tau))(I)$.

Обозначаем множество символов табло τ и $\mu = (\mu_1 \ldots \mu_i \ldots \mu_n)$. Тогда существует оценка ρ для τ :

$$\rho(w_o) = \rho(v_1)\ldots \rho(v_i)\ldots \rho(v_n) = (\mu_1 \ldots \mu_i \ldots \mu_n),$$

$\rho(w_j) \in I, \quad j = \overline{1,m},$

$w_o = (v_1 \ldots v_i \ldots v_n)$ - сводка, w_j - строка табло τ.

По определению табло символы появляющиеся на сводке могут быть характерными переменными, символами множества значений или бланками. Так как $\sigma_{A_i \in \xi}(\tau(I)) \neq \emptyset$, поэтому v_i не может быть бланком. Рассмотрим два случая: (i) v_i - характерная переменная a_i . Тогда:

$$\rho(v_i) = \mu[A_i] = \mu_i \in \xi.$$

(ii) v_i - символ множества значений $\tilde{\xi}_1$. Тогда

$$\rho(v_i) = \rho(\tilde{\xi}_1) = \mu[A_i] = \mu_i \in \xi \cap \xi_1.$$

Пусть - множество символов табло $\tau' = \sigma_{A_i \in \xi}(\tau)$,

$w'_o = (u_1 \ldots u_i \ldots u_n)$ - сводка табло τ' .

По правилам построения τ' из τ очевидно что У' отличается от У только одним символом v_i и u_i. Будем выбирать оценку ρ' для τ' следующим образом:

-$\forall d \in У'$ и $d \in У$ /т.е. $d \neq u_i$/, тогда $\rho'(d) = \rho(d)$

$$\rho'(u_i) = \begin{cases} \rho'(\tilde{\xi}) = \rho(v_i) = \mu[A_i] = \mu_i \in \xi, & \text{если } v_i \text{ характерная переменная } a_i \\ \rho'(\tilde{\xi}_2) = \rho(\tilde{\xi}_1) = \mu[A_i] = \mu_i \in \xi_2, & \text{где } \xi_2 = \xi \cap \xi_1, \text{ если } v_i \text{ символ множества } \tilde{\xi}_1. \end{cases}$$

По способу построения ρ' и ρ имеем:

$$\rho'(w'_o) = \rho(w_o) = \mu \; ; \quad \rho'(w'_j) = \rho(w_j) \in I, \quad j = \overline{1,m}$$

Следует $\mu \in \sigma_{A_i \in \xi}(\tau))(I)$.

Если $\sigma_{A_i \in \xi}(\tau(I)) \neq \emptyset$, видно, что а/ - удовлетворено.

б/ - аналогично можем доказать, что

$$\sigma_{A_i \in \xi}(\tau(I)) \supseteq (\sigma_{A_i \in \xi}(\tau))(I) \qquad (1.б)$$

Таким образом равенство 1^{o}/ доказано.

2/ По определению обобщенного табло и правила построения $\Pi_X(\tau)$ из τ легко показать, что: $\Pi_X(\tau(I)) = (\Pi_X(\tau))(I)$, $\forall I$.

3/ Для доказательства 3-его равенства, сначала мы покажем, что

а/ $\tau(I) \times \tau_1(I) \subseteq (\tau \times \tau_1)(I)$, $\forall I$.

Пусть $\tau(I) \times \tau_1(I) \neq \emptyset$. Надо доказать, что

$\forall \nu \neq \emptyset$, $\nu \in \tau(I) \times \tau_1(I) \rightarrow \nu \in (\tau \times \tau_1)(I)$.

По предложению $\nu = \mu \bowtie \mu$, где $\mu \in \tau(I)$, $\mu_1 \in \tau_1(I)$.

Это означает, что $\exists$ оценка ρ для τ и ρ_1 для τ_1 такие, что:

$$\rho(w_o) = \mu, \quad \rho(w_j) \in I, \quad j = \overline{1,m}$$

w_o - сводка, w_j - строка табло τ .

$$_1(w_o^1) = \mu_1, \quad \rho_1(w_j^1) \in I, \quad j = \overline{1,m_1}$$

w_o^1 - сводка, w_j^1 - строки табло τ_1 .

Пусть У, $У_1$ множества символов двух табло τ, τ_1 соответственно.

$У'$ множество символов $\tau' = \tau \times \tau_1$. Выбираем оценку ρ' для τ' следующим образом:

если $d \in У'$, $d \in У$, $d \notin У_1$, то $\rho'(d) = \rho(d)$

если $d \in У'$, $d \notin У$, $d \in У_1$, то $\rho'(d) = \rho_1(d)$

если $d \in У'$, $d \in У$ $d \in У_1$, то $\rho'(d) = \rho(d) = \rho_1(d) = \mu[A_i] = \mu_1[A_i]$.

/Это имеет место, когда d является характерной переменной или символом множества $\tilde{\xi}$ в У и $У_1$./

Если $d \in У'$, $d \notin У$, $d \notin У_1$, то по правилу построения $\tau \times \tau_1$, d может быть только символом множества значений в каком-нибудь столбце A_j сводки табло $\tau \bowtie \tau_1$. В соответственном столбце сводки табло τ имеет символ множества $\tilde{\xi}$, а табло $\tau_1 - \tilde{\xi}_1$, $\xi \cap \xi_1 \neq \xi$, $\xi \cap \xi_1 \neq \xi_1$.

Мы имеем:

$$\rho'(d) = \rho'(\tilde{\xi}_2) = \rho(\tilde{\xi}) = \rho_1(\tilde{\xi}_1) = \mu[A_i] = \mu_1[A_j], \ \xi_2 = \xi \cap \xi_1.$$

На основе выбирания ρ' для $\tau \times \tau_1$ имеем:

$$\rho'(w'_o) = \nu \ , \quad \rho'(w'_j) \in I, \quad j = \overline{1, m+m_1} \ ,$$

w'_o - сводка, w'_j - строка табло $\tau \times \tau_1$.

Отсюда следует, что $\nu \in (\tau \bowtie \tau_1)(I)$.

Если $\tau(I) \bowtie \tau_1(I) = \emptyset$, видно, что а/ - удовлетворено.

Так и требовалось доказать.

в. Аналогично можно доказать:

$$\tau(I) \bowtie \tau_1(I) \supseteq (\tau \times \tau_1)(I) .$$

Таким образом равенство 3/ доказано.

Т е о р е м а 1

Пусть E ограниченное реляционное выражение, использование правила 1 - 4 позволяет создать соответствующее обобщенное табло τ дли E:

$$V_I(E) = \tau(I) \quad \forall I .$$

Д о к а з а т е л ь с т в о:

Доказательство теоремы проводится индуктивно по числу операции в E.

1/ Пусть E - схема отношения R. Тогда из определения $\tau(I)$ и $V_I(E)$ непосредственно следует, что $V_I(E) = \tau(I)$.

2/ Пусть второе правило было использовано. $E = \sigma_{A_i \in \xi}(E_1)$ и τ_1 обобщённое табло E_1. Индуктивным образом имеем:

$V_I(E_1) = \tau_1(I)$. Обозначаем $\tau = \sigma_{A_i \in \xi}(\tau_1)$ - обобщенное табло для E построено с использованием правила 2/. Тогда на основе определения отображения V_I /правило 2а/ и леммы 1 получается:

$$V_I(E) = V_I\ \sigma_{A_i \in \xi}(E_1)) = \sigma_{A_i \in \xi}(V_I(E_1))$$
$$= \sigma_{A_i \in \xi}(\tau_1(I)) = (\sigma_{A_i \in \xi}(\tau_1))\ (I)$$
$$= \tau(I).$$

3/ Пусть третье правило было использовано $E = \Pi_X(E_1)$. По индуктивно, то $V_I(E_1) = \tau_1(I)\ \forall I$. Обозначение $\tau = \Pi_X(\tau_1)$ табло построенное для E использованием 3-его правила. Тогда на основе определения V_I /правило 2б/ и леммы 1 имеем:

$$V_I(E) = V_I(\Pi_X(E_1)) = \Pi_X(V_I(E_1))$$
$$= \Pi_X(\tau_1(I)) = (\Pi_X(\tau_1))\ (I)$$
$$= \tau(I).$$

4/ Пусть $E = E_1 \bowtie E_2$, τ_1 и τ_2 два обобщенных табло для E_1, E_2 соответственно. По предположению индукции

$$V_I(E_1) = \tau_1(I), \quad V_I(E_2) = \tau_2(I).$$

Обозначаем $\tau = \tau_1 \bowtie \tau_2$ табло построенное использованием 4-ого правила.

Тогда: $V_I(E) = V_I(E_1 \bowtie E_2)$
$$= V_I(E_1) \bowtie V_I(E_2) = \tau_1(I)\ x\ \ _2(I)$$
$$= (\tau_1 \bowtie \tau_2)\ (I)$$
$$= \tau(I) \text{ и т.д. и т.п.}$$

Таким образом для любого ограниченного реляционного выражения E можем построить его соответствующее табло τ. Поэтому τ используется как способ представления реляционных выражений. В третьем разделе будем использовать эквивалентности обобщенных табло.

3. Эквивалентность обобщенных табло

Табло τ_1 и τ_2 называются эквивалентными /обозн. $\tau_1 \equiv \tau_2$/, если при любом состоянии универсального отношения I:

$$\tau_1(I) = \tau_2(I).$$

Табло τ_1 называется включенным в τ_2 /обозн. $\tau_1 \subsetneq \tau_2$/, если $\forall I$

$$\tau_1(I) \subseteq \tau_2(I).$$

З а м е ч а н и е : Необходимым условием для $\tau_1 \equiv \tau_2$ или $\tau_1 \subseteq \tau_2$ является то, что τ_1 и τ_2 имеют общую целевую схему отношения.

Пусть τ_1 и τ_2 два обобщенных табло имеющих множество символов $У_1$ и $У_2$ соответственно. Отображение $\Psi : У_1 \to У_2$ называют гомоморфизмом, если Ψ удовлетворяет слудующим условиям:

1/ Есди $\tilde{\xi}_1 \in У_1$, $\tilde{\xi}_1$ - символ множества значения, то
$\Psi(\tilde{\xi}_1) = \tilde{\xi}_2$ - символ множества значений и $\xi_2 \subsetneq \xi_1$

2/ Если $a_i \in У_1$ характерная переменная, то $\Psi(a_i)$ является или характерной переменной или символом множества значений $\tilde{\xi}_k$, который появляется на сводке в соответственном столбце табло τ_2.

3/ Если w_j какая-то строка табло τ_1, то $\Psi(w_j)$ тоже строка τ_2.

Аналогично в /4/ для обобщенного табло мы имеем место следующей теоремы.

Т е о р е м а 2 :

Пусть τ_1 и τ_2 два обобщенных табло с общей целевой схемой отношения, имеющего множества символов $У_1$ и $У_2$ соответственно. Необходимыми достаточными условиями для $\tau_2 \subseteq \tau_1$ является существование гомоморфизма

$$\Psi : У_1 \to У_2.$$

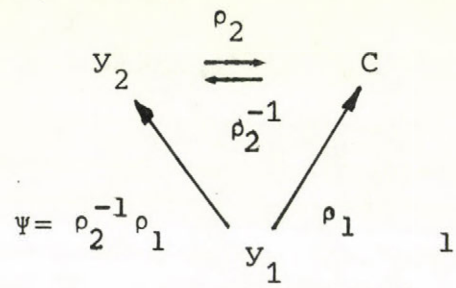
Д о к а з а т е л ь с т в о :

Необходимость - Если $\tau_2 \subseteq \tau_1$, то существует гомоморфизм. Пусть ρ_2 оценка табло τ_2, которая взаимооднозначно отображает $У_2$ в множество значений С. Выберем I состояние универсального отношения составляющегося из таких кортежей: $\rho_2(w_j^2)$, $j=\overline{1,m_2}$, w_j^2 - строки табло τ_2.

С таким выбором I и ρ_2 - оценка для τ_2 имеем:

$$\rho_2(w_o^2) = \mu \in \tau_2(I), \quad w_o^2 \text{ - сводка } \tau_2.$$

По предположению $\tau_2 \subseteq \tau_1 \to \mu \in {}_1(I)$, отсюда следует, что существование оценки ρ_1 для τ_1 таково, что:


(р.1)

$$\rho_1(w_o^1) = \mu$$

$$\rho_1(w_j^1) \in I, \quad j = \overline{1,m_1}$$

где w_o^1 - сводка, w_j^1 - строки табло τ_1.

Построим отображение /рис. 2./ $\psi = \rho_2^{-1}\rho_1$, и докажем, что ψ гомоморфизм. То есть ψ нужно удовлетворять условиям 1-3.

Пусть $w_o^1 = v_1 \ldots v_i \ldots v_n$

$$w_o^2 = u_1 \ldots u_i \ldots u_n.$$

Исходя из того, что τ_1 и τ_2 имеют общую целевую схему отношения и $\rho_1(w_o^1) = \rho_2(w_o^2) = \mu$, поэтому $\psi : v_i \to u_i$, где v_i, u_i являются либо характерными переменными либо символами множества значений. Таким образом ψ удовлетворяет 2-ому условию гомоморфизма.

Если v_i символ множества значений $\tilde{\xi}_1$, то $\psi(v_i)$ не может быть характерной переменной. Действительно в обратном случае $\psi(\tilde{\xi}_1) = a_i$ и с - значение присвоенное характерной переменной, $c \in dom(A_i)$, $c \notin \xi_1$.

Строим I' из I с заменой $\rho_2(a_i)$ в столбце i значением с и выберем оценку ρ_2' : $\rho_2'(d) = \rho_2(d)$, для $d \in У_2$ и $d \neq a_i$; $\rho_2'(a_i) = c$.

Тогда $\mu' = \rho_2'(w_o^2) \in \tau_2\ (I')$, но $\mu' \notin \tau_1\ (I')$. Это противоречит предположению $\tau_2 \subseteq \tau_1$.

И так $\Psi(\tilde{\xi}_1)$ является символом множества значений $\tilde{\xi}_2$. Более этого $\xi_2 \subseteq \xi_1$. Действительно в обратном случае $\exists c \in \xi_2 - \xi_1 \neq \emptyset$ и аналогичным рассуждением предыдущему пришли к противоречию $\tau_2 \nsubseteq \tau_1$.

Таким образом Ψ удовлетворяет 1-ому условию гомоморфизма.

Сейчас нам нужно доказать, что Ψ удовлетворят 3-ему условию гомоморфизма, где $\Psi = \rho_2^{-1}\ \rho_1$.

Если w_j^1 - строка τ_1, то $\rho_1(w_j^1)$ - один кортеж состояния I. ρ_2 - взаимнооднозначное отображение из $У_2$ в I. Тогда $\rho_2^{-1}(\rho_1(w_j^1))$ - строка τ_2.

<u>Д о с т а т о ч н о с т ь :</u> Пусть существует гомоморфизм $\Psi : У_1 \to У_2$, I - любое состояние универсального отношения, $\mu \in \tau_2(I)$. Тогда оценка ρ_2 для τ_2: ρ_2 отображает множество $У_2$ табло τ_2 в множество значений С, т.е.

$$\rho_2 : У_2 \to C$$

$$\rho_2(w_o^2) = \mu; \quad \rho_2(w_j^2) \in I\ ,\ j = \overline{1,m_2}$$

w_o^2 - сводка, w_j^2 - строки τ_2.

Для иллюстрации $\mu \in \tau_1(I)$ выберем оценку ρ_1 для τ_1 таким образом /р.2/

$$\forall d \in У_1 \to \rho_1(d) = \rho_2(\Psi(d))$$

$$\begin{tikzcd} У_1 \arrow[r, "\Psi"] \arrow[dr, "\rho_1 = \rho_2(\Psi)"'] & У_2 \arrow[d, "\rho_2"] \\ & C \end{tikzcd}$$

/р.2/

По определению гомоморфизма, если $\tilde{\xi}_1 \in Y_2$ - символ множества значений, то

$$\rho_1(\tilde{\xi}_1) = \rho_2(\Psi(\tilde{\xi}_1)) = \rho_2(\tilde{\xi}_2) = c \in \xi_2$$

С другой стороны $\tilde{\xi}_2 = \Psi(\tilde{\xi}_1)$, $\xi_2 \subseteq \xi_1 \rightarrow c \in \xi_1$.

Из условия гомоморфизма следует

$$\rho_1(w_o^1) = \rho_2(\Psi(w_o^1)) = \rho_2(w_o^2) = \mu$$

w_o^1, w_o^2 - сводки τ_1 и τ_2 соответственно.

$$\rho_1(w_j^1) = \rho_2(\Psi(w_j^1)) = \rho_2(w_k^2) \in I$$

w_j^1, w_k^2 - строки τ_1 и τ_2 соответственно.

Таким образом $\mu \in \tau_1(I)$ и $\tau_2 \subseteq \tau_1$. Теорема доказана.

Пусть τ_1 и τ_2 обобщенные табло, Θ отобразает строки табло τ_1 в строки τ_2.

Θ называется включенным отображением, если оно удовлетворяет следующим условиям:

а/ Если в столбце A_k на строке i табло τ_1 имеется характерная переменная, то в столбце A_k на строке $\Theta(i)$ табло τ_2 также имеется характерная переменная.

б/ Если в столбце A_k на строке i табло τ_1 имеется символ множества значений $\tilde{\xi}$, то в столбце A_k на строке $\Theta(i)$ табло τ_2 имеется символ множества значений $\tilde{\xi} : \xi' \subseteq \xi$.

в/ Если столбец A_k на строках i и j табло τ_1 имеют одинаковые нехарактерные переменные, то в столбце A_k на строках $\Theta(i)$ и $\Theta(j)$ табло τ_2 имеют одинаковые символы. Этот символ может быть символом множества значений, характерной или нехарактерной переменной.

Включенное отображение является основой для проверки эквива-

4. Связь между табло и обобщенным табло

в /7/ введены понятие монотонного реляционного выражения и объединения табло. Под понятием монотонного реляционного выражения понимают вырашения обладающие операциями выбора, проектирования, соединения и объединения.

Объединением табло называется выражение $\bigcup_{i=1}^{n} T_i$, где $T_1, T_2, \ldots, T_n$ табло имеющие общую целевую схему отношения. Общая целевая схема отношения табло T_i, $i = \overline{1,n}$ является целевой схемой объединения табло $\bigcup_{i=1}^{n} T_i$. Объединение табло - форма представления монотонного реляционного выражения.

При одном состоянии I универсального отношения значения объединения табло определеяются следующим образом: $(\bigcup_{i=1}^{n} T_i)(I) = \bigcup_{i=1}^{n} T_i)(I)$.

Аналогично предыдущему мы имеем понятие объединения обобщенных табло. Пусть $\tau_1, \ldots, \tau_n$ обобщенные табло имеющие общую целевую схему отношения.

Тогда $\bigcup_{i=1}^{n} \tau_i$ объединение обобщенных табло. $\forall$I - значение объединения обобщенных табло определяется таким же образом:
$(\bigcup_{i=1}^{n} \tau_i)(I) = \bigcup_{i=1}^{n} \tau_i(I)$.

Для выявления связи между табло (T) и обобщенным табло (τ) мы рассмотрим правило разложения обобщенного табло τ на объединение табло.

Пусть У - множество символов τ и табло τ имеет только один символ множества значений $\tilde{\xi}$ в У /обозначаем $\tau = \tau\ |\xi|$/. Мы будем обозначать $S_c^{\tilde{\xi}}\ \tau(|\xi|)$ замена $\tilde{\xi}$ в табло τ одним элементом $c \in \xi$ в местах, где $\tilde{\xi}$ появляется. $\tau[c]$ - табло, полученное после замены. Видно, что $\tau[c] = T[c]$ так как табло (T) является частным случаем обобщенного табло когда все множества имеют один элемент, т.е.

$$S_{c}^{\tilde{\xi}}\ (\tau\,[\xi]) = \tau\,[c] = T[c].$$

Если τ имеет k символов множества значений $\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_k$, т.е. $\tau = \tau\ [\xi_1, \ldots, \xi_k]$. Обозначаем через $S_{c_1 \ldots c_k}^{\tilde{\xi}_1 \ldots \tilde{\xi}_k}(\tau\,[\xi_1 \ldots \xi_k])$ - замены в обобщенном табло каждого символа значения $\tilde{\xi}_i$ одним элементом $c_1 \in \xi_i$.

Табло полченное после замен имеет вид:

$$S_{c_1 \ldots c_k}^{\tilde{\xi}_1 \ldots \tilde{\xi}_k}(\tau\,[\xi_1 \ldots \xi_k]) = \tau\,[c_1 \ldots c_k] = T[c_1 \ldots c_k].$$

<u>Л е м м а 2 :</u> Пусть τ - обобщенное табло, имеющее только один символ множества значений $\tilde{\xi}$ в У, т.е. $\tau = \tau\ [\xi]$. Тогда $\tau\,[\xi] = \bigcup\limits_{c \in \xi} T[c]$.

<u>Д о к а з а т е л ь с т в о :</u>

Надо доказать, что $\forall I$ имеет место равенства:

$$(\tau\,[\xi])\ (I) = (\bigcup_{c \in \xi} T[c])$$

пусть $\mu \in ([\xi])(I)$ и $\mu \neq \emptyset$. Это означает существование оценки ρ для $\tau\,[\xi]$, такой, что: $\rho\,(w_o) = \mu$, $\rho\,(w_j) \in I$, $j = \overline{1,m}$; $\rho\,(\tilde{\xi}) = c_k \in \xi$,

w_o - сводка, w_j - строки табло τ .

По правилу разложения табло τ, в $\bigcup\limits_{c \in \xi} T(c)$ имеется табло $T[c_k] = S_{c_k}^{\xi}\ \tau\,[\xi]$. Если S - множество символов табло T, то S и У отличаются только одним символом c_k и $\tilde{\xi}$.

Выберем оценку ρ' для $T[c_k]$ следующим образом:

$$\rho'(d) = \rho(d),\ \forall d \in S,\ \ d \neq c_k;\ \ \rho'(c_k) = \rho(\tilde{\xi}) = c_k$$

Тогда $\rho'(w_o^T) = \mu$, $\rho'(w_j^T) \in I$, $j = \overline{1,m}$

w_o^T - сводка, w_j^T - строки табло T.

лентности табло и установления эквивалентных преобразований.

Т е о р е м а 3: Пусть τ_1, τ_2 - обобщенные табло имеющие общую целевую схему отношения. Необходимым и достаточным условиями для $\tau_2 \subseteq \tau_1$ является существование включенного отображения $\theta : \tau_1 \to \tau_2$.

Д о к а з а т е л ь с т в о :

Необходимость: Если $\tau_2 \subseteq \tau_1$, то по теореме 3 существует гомоморфизм $\Psi : У_1 \to У_2$. Надо доказать, что Ψ является включенным отображением.

Из 1-ого и 2-ого условия гомоморфизма следует, что Ψ удовлетворяет условиям а/ и б/ включенного отображения. Из 3-его условия гомоморфизма следует, что Ψ отображает строки табло τ_1 в строки табло τ_2, одновременно Ψ удовлетворяет условию с/ включенного отображения.

Д о с т а т о ч н о с т ь :

Пусть существует включенное отобразение θ надо доказать, что $\tau_2 \subseteq \tau_1$.

Пусть отобразение $\Psi : У_1 \to У_2$ удовлетворяет следующему условию: если d_1 - символ в столбце A_k на строке w_j^1 табло τ_1 и d_2 символ в столбце A_k на строке $\theta(w_j^1)$ табло τ_2, то $\Psi(d_1) = d_2$.

Докажем, что Ψ - гомоморфизм.

Известно, что θ - включенное отображение. Из условия б/ отображения θ следует Ψ удовлетворяет 1-ому условию. Из условия а/ отображения θ следует, что Ψ удовлетворяет 2-ому условию. Ψ тоже удовлетворяет 3-ему условию, так как Ψ было определено на основе отображения θ. Таким образом Ψ гомоморфизм. По теореме 3, $\tau_2 \subseteq \tau_1$.

С л е д с т в и е 1: τ_1, τ_2 два обобщенных табло имеющих общую целевую схему отношения. Необходимым и достаточным условием для $\tau_1 \equiv \tau_2$ является существование включенного отображения из τ_1 в τ_2 и включенного отображения из τ_2 в τ_1 .

И так $\mu \in (T[c_k])(I) \Rightarrow \mu \in (\bigcup_{c \in \xi} T[c])(I)$.

Аналогичным образом мы можем доказать, что если

$$\mu \in (\bigcup_{c \in \xi} T[c])(I) \Rightarrow \mu \in (\tau[\xi])(I) .$$

Лемма доказана.

Из леммы 2 доказать индуктивно по числу множество значений в табло τ имеем место следующей теоремы:

Т е о р е м а 4 : Пусть τ обобщенное табло, имеющее k символов множеств значений $\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_k$ в k столбцах соответствующих k атрибутам табло τ . Тогда имеет место следующее равенство:

$$\tau[\xi_1, \ldots, \xi_k] = \bigcup_{c_1 \in \xi_1} \ldots \bigcup_{c_k \in \xi_k} T[c_1 \ldots c_k] .$$

З а м е ч а н и е : Не любое произвольное реляционное выражение представленное в виде объединения табло может быть представлено в виде обобщенного табло.

П р и м е р : Реляционное выражение:

$$\Pi_{Н,Ф,Д,З}(\sigma_{У=19 \wedge (\text{`ДИРЕКТОР'} \vee З=2000)}(R_1))$$

Может быть представлено в виде объединения, но не может быть представлено в виде эквивалентного обобщенного табло.

З а к л ю ч е н и е: С понятием табло T в [4], [5], [7] запросы обладающие операцией выбора вида $\sigma_{A \in \{a_1 \ldots a_n\}}(r)$ обычно подлежат разложению на объединения n подзапросов и будут представлены в виде объединения табло, каждое из которых соответствует одному запросу. Например, выражение $\Pi_{Н,Ф,У}(\sigma_{У \in \{17,19,32\}}(R_1))$ может быть представлено в виде объединения трех подвыражений:

$$\Pi_{Н,Ф,У}(\sigma_{У=17}(R_1)) \cup \Pi_{Н,Ф,У}(\sigma_{У=19}(R_1)) \cup \Pi_{Н,Ф,У}(\sigma_{У=32}(R_1)) .$$

Тем самым одно может быть представлено в виде $\bigcup_{i=1} T_i$, каждое T_i соответствует подвыражению. С точки зрения осуществления на практике, такой подход будет дорогостоящим из-за большой запоминающей емкости и вычислительного времени ЭВМ. Это объясняется тем, что количество табло, находящихся в памяти, и, с которыми все время должны обращаться, равно n. В случае достаточно большого n работа будет слишком сложной. В то время как показали выше, такие запросы могут быть представлены в виде эквивалентного обобщенного табло τ . Поэтому обработка таких выражений будет более удобной и эффективной. Кроме этого введение понятия обобщенного табло еще позволяет расширить результаты в /4/, /5/, /7/ для реляционных выражений с операцией выбора $\sigma_{A\theta C}$, где $\theta = \{\neq,<,\leq,>,\geq\}$.

В общем для достижения этих расширенных результатов требуются более сложные и тонкие техники доказательства.

Автор благодарит доц. Хо Тхуан, Нгуен Кат Хо и Ле Тиен Вьюнг за полезные обсуждения.

ЛИТЕРАТУРА

[1] Codd, E.F., A Relational Model of Data for Large Shared Data Banks, Comm. Ac., 1970, v.13, No.6, p. 337-387.

[2] Codd, E.F., Further Normalization of the Database Relation Relational Model. In: Database Systems, Englewood Cliffs, N.Y.: Prentice-Hall, 1972, p. 33-64.

[3] Codd, E.F., Relational Completeness of Database Sublanguages. In: Database Systems. Englewood Cliffs, N.Y. :Prentice-Hall, 1972, p. 79-90.

[4] Aho A.V., Sagiv, Y., and Ullman, J.D., Efficient Optimization of a class of Relational Expressions. ACM Trans. Database Syst., 4,4 :Dec, 1979 p. 435-454.

[5] Aho, A.V., Sagiv, Y., and Ullman, Y.D., Efficient Optimization of a Relational Expressions. ACM Trans. Database Syst., 4,4 :Dec. 1979. p. 435-454.

[6] Astrahan, M.M., Chamberlin, D.D., Implementation of a Structured English Query Language. Communication of the ACM Oct. 1975 Vol. 18, Number 10.

[7] Sagiv, A., and Yannakakis, M., Equivalence Among Relational Expressions with the Union and Difference Operators., Y, if the Association for Computing Machinery, Vol. 27, No. 4, Oct. 1980.

[8] Aho A.V., Beeri, C., and Ullman, Y.D., The Theory of Joins in Relational Database Proc. 18-th IEEE Symposium on Foundations of Computer Science, 1977, p. 107-113.

[9] Деит К.: Введение в системы баз данных. М. Наука. 1980

[10] Мартин Дж.: Организация баз данных в вычислительных сисметах. М. Мир. 1980.

[11] Горчинская О.Ю.: Теоретический аспект построения реляционных моделей. А. и Т. № 1, 1983.

The equivalence of a class of relational expressions

Do Suan Tho

Summary

Many database queries can be formulated in terms of relational expressions using the relational operations select, project, join. The equivalence problem for these queries is studied with query optimization in mind.

In this paper we introduce the so-called generalized tableaux, two-dimensional representations of queries. Every relational expression over the operations select of the $\sigma_{A\in\xi}(r)$ form (A-attribute of the relation r, $\xi \subseteq \mathrm{dom}(A)$), project and join can be represented by generalized tableaux. We consider the equivalence problem for generalized tableaux, then we discuss the relationship between tableaux and generalized ones.

A relációs kifejezések egy osztályának ekvivalenciájáról

Do Suan Tho

Összefoglaló

Több adat-bázis lekérdezést relációs kifejezések formájában lehet megfogalmazni, használva a kiválasztás, projekció és összekapcsolás operációkat. Ezen lekérdezések ekvivalencia problémájáról van szó, különös tekintettel az optimalizációra. A cikkben a szerző egy u.n. általánositott táblát vezet be, a lekérdezések kétdimenziós reprezentációját.
Minden olyan relációs kifejezés, amelynek $\sigma_{A\in\xi}(r)$ formája van, reprezentálható általánositott táblával. A cikkben az ekviva-

lencia problémát a szerző egyrészt az általánositott táblákra és az általánositott és közönséges táblákra vonatkozóan vizsgálja.