

Krassimir Markov, Vitalii Velychko, Oleksy Voloshin
(editors)

**Information Models
of
Knowledge**

**ITHEA[®]
KIEV – SOFIA
2010**

Krassimir Markov, Vitalii Velychko, Oleksy Voloshin (ed.)

Information Models of Knowledge

ITHEA®

Kiev, Ukraine – Sofia, Bulgaria, 2010

ISBN 978-954-16-0048-1

First edition

Recommended for publication by The Scientific Council of the Institute of Information Theories and Applications FOI ITHEA

ITHEA IBS ISC: 19.

This book maintains articles on actual problems of research and application of information technologies, especially the new approaches, models, algorithms and methods for information modeling of knowledge in: Intelligence metasynthesis and knowledge processing in intelligent systems; Formalisms and methods of knowledge representation; Connectionism and neural nets; System analysis and synthesis; Modelling of the complex artificial systems; Image Processing and Computer Vision; Computer virtual reality; Virtual laboratories for computer-aided design; Decision support systems; Information models of knowledge of and for education; Open social info-educational platforms; Web-based educational information systems; Semantic Web Technologies; Mathematical foundations for information modeling of knowledge; Discrete mathematics; Mathematical methods for research of complex systems.

It is represented that book articles will be interesting for experts in the field of information technologies as well as for practical users.

General Sponsor: Consortium FOI Bulgaria (www.foibg.com).

Printed in Ukraine

Copyright © 2010 All rights reserved

© 2010 ITHEA® – Publisher; Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org ; e-mail: info@foibg.com

© 2010 Krassimir Markov, Vitalii Velychko, Oleksy Voloshin – Editors

© 2010 Ina Markova – Technical editor

© 2010 For all authors in the book.

® ITHEA is a registered trade mark of FOI-COMMERCE Co., Bulgaria

ISBN 978-954-16-0048-1

C/o Jusautor, Sofia, 2010

ОБОБЩЕННОЕ ИСЧИСЛЕНИЕ НА ДОМЕНЕ

Дмитрий Буй, Ирина Глушко

Аннотация: В данной статье доказано, что обобщенная табличная алгебра, исчисление строк (кортежей) и исчисление на домене, пополненные произвольными предикатными и функциональными сигнатурами на универсальном домене, одинаково выразительны. Проведена редукция исчисления строк к исчислению на домене. Доказано, что при этом обобщении табличная алгебра остается не менее выразительной, чем исчисление на домене.

Ключевые слова: реляционные (табличные) базы данных, исчисление на домене, табличная алгебра, исчисление строк (кортежей).

АСМ классификация ключевых слов: H.2 Database Management (E.5) H.2.4 Systems – Relational databases

Введение

В предыдущих работах [Буй, Глушко, 2010], [Buy, Glushko, 2010] проведено обобщение табличной алгебры, построенной на основе реляционных алгебр Кодда; обобщение состоит в том, что вместо таблиц рассматриваются пары, состоящие из таблиц и схем таблиц. Сделан первый шаг в доказательстве эквивалентности табличной (реляционной) алгебры Кодда и исчисления строк (кортежей), пополненного произвольными предикатными и функциональными сигнатурами на универсальном домене: доказано, что исчисление строк при этом обобщении не менее выразительно, чем табличная алгебра.

Задача данной работы показать, что обобщенная табличная алгебра, исчисление строк (кортежей) и исчисление на домене, пополненные произвольными предикатными и функциональными сигнатурами на универсальном домене, одинаково выразительны.

Обобщенное исчисление на домене

М. Лакруа (M. Lacroix) и А. Пиротте (A. Pirotte) [Lacroix, Pirotte, 1977] предложили альтернативную к исчислению кортежей версию исчисления – реляционное исчисление с переменными на доменах. Различие их в том, что вместо переменных строк рассматриваются переменные значения из домена, которые представляют компоненты строки. Кроме того, исчисление на домене поддерживает условие принадлежности (membership condition) [Дейт, 2005], [Хомоненко, Цыганков, Мальцев, 2004]. В общем виде условие принадлежности записывается как $\langle t(\langle A_1, d_1 \rangle, \langle A_2, d_2 \rangle, \dots), R \rangle$, где R – схема, A_i – атрибут таблицы t , а d_i – переменная домена или литерал (константа). Проверяемое условие истинно тогда и только тогда, когда в таблице t существует строка, которая имеет указанные значения из универсального домена D для указанных атрибутов. Причем порядок атрибутов в таблице фиксированный.

Все неопределенные в статье понятия и обозначение понимаем в смысле [Редько, Брона, Буй, Поляков, 2001].

Введем множество так называемых разрешенных формул исчисления на домене при использовании:

- множества атрибутов A и универсального домена D ;
- множества предметных констант d_1, d_2, \dots ;
- множества предметных переменных x_1, x_2, \dots ;
- множества функциональных символов $f_1^{n_1}, f_2^{n_2}, \dots$;
- множества предикатных символов $p_1^{m_1}, p_2^{m_2}, \dots$.

Областью интерпретации предметных констант и предметных переменных есть универсальный домен D . Применяем x как синтаксическую переменную, областью изменения которой являются переменные; $f(p)$ как синтаксическую переменную, областью изменения которой являются функциональные (соответственно предикатные) символы; d как синтаксическую переменную, областью изменения которой являются константы; наконец, \mathcal{A} как синтаксическую переменную, областью изменения которой являются атрибуты.

Следующие выражения являются термами (индукция по длине термов):

- a) всякая предметная константа или предметная переменная есть терм;
- b) если u_1, \dots, u_n – термы, f – n -арный функциональный символ, то $f(u_1, \dots, u_n)$ – терм;
- c) выражение является термом только в том случае, если это следует из правил a), b).

Будем применять u как синтаксическую переменную, областью изменения которой являются термы.

Атомарные формулы (атомы) имеют вид:

- a1. пусть $\langle t, R \rangle$ – таблица схемы $R = \{A_1, \dots, A_n\} \subseteq A$. Тогда $t(a_1, \dots, a_n)$ – атом, где a_i – термы, $i = 1, \dots, n$;
- a2. пусть u_1, \dots, u_m – термы, а p – m -арный предикат на универсальном домене D . Тогда $p(u_1, \dots, u_m)$ – атом.

Используем логические связки \neg, \wedge, \vee , кванторы \exists, \forall и скобки $(,)$ для построения формул из атомов.

- f1. Каждый атом – формула
- f2. Если P – формула, то $\neg P$ – формула.
- f3. Если P и Q – формулы, то $P \wedge Q, P \vee Q$ – формулы.
- f4. Пусть x – переменная, P – формула, \mathcal{A} – атрибут из множества A , тогда $\exists x (\mathcal{A})P$ – формула.
- f5. Пусть x – переменная, P – формула, $\neg \mathcal{A}$ атрибут из множества A , тогда $\forall x (\mathcal{A})P$ – формула.
- f6. Если P – формула, то (P) – формула.
- f7. Других формул нет.

Понятия свободных и связанных переменных на домене и область действия связанной переменной определяются в исчислении на домене так же, как и в исчислении строк. Схема переменной x в формуле P $scheme(x, P)$ – это или универсальный домен D , или неопределенная величина. Формально определять понятие разрешенной формулы не будем. Как и в исчислении строк формула разрешена, если выполняется требование согласования схем переменной в подформулах и требование того, что переменная, связанная квантором, должна входить свободно в формулу, которая следует за квантором.

Выражение исчисления с переменными на домене имеет вид: $\{x_1, \dots, x_n \mid P(x_1, \dots, x_n)\}$, где

1. формула P – разрешена, а x_1, \dots, x_n – все свободные переменные, входящие в формулу P ;
2. $R = \{A_1, \dots, A_n\}, R \subseteq A$ – схема, порядок атрибутов фиксированный;
3. $scheme(x_i, P) = D, i = 1, \dots, n$.

Пусть $P(x)$ – разрешенная формула, \mathcal{A} – атрибут из множества A и если $scheme(x, P)$ определена, то $scheme(x, P) = D$. В результате подстановки константы d из универсального домена D вместо переменной x в формулу P получим формулу обозначенную $P(d/x)$. Определим сначала истинностные значения атомов:

a1. пусть переменная x в подформуле $t(x)$ свободна в P , $R = \{A\}$ – схема таблицы $\langle t, R \rangle$. Атом $t(x)$ приобретает значение истина, при подстановке константы $d \in D$ вместо x , если $\exists s(s \in t \wedge \langle A, d \rangle \in s)$, где s – строка таблицы $\langle t, R \rangle$, иначе атом $t(x)$ приобретает значение ложь;

a2. пусть переменная x в подформуле $p(u_1, \dots, u_m)$ свободна в P , тогда при подстановке константы $d \in D$ вместо переменной x , заменим соответствующую предметную переменную на d . Атом $p(u_1, \dots, u_m)$, где u_i – термы (предметные константы), приобретает значение истина, если предикат p истинный на соответственных значениях, иначе атом приобретает значение ложь.

Пусть формула P разрешенная формула без свободных переменных. Интерпретация формулы P определяется следующим образом.

f2. Если $P = \neg G$, то в G нет свободных переменных. Формула P истинная, когда G ложная, и ложная, когда формула G истинная.

f3. Если $P = G \wedge Q$ или $P = G \vee Q$, то в G и Q нет свободных переменных. Если $P = G \wedge Q$, то формула P истинная тогда, когда G и Q обе истинные, в противном случае формула P ложная. Если $P = G \vee Q$, то формула P ложная тогда, когда G и Q обе ложные, в противном случае формула P истинная.

f4. Если $P = \exists x(A)G$, то x – единственная свободно входящая в формулу G переменная. Формула P истинная, если существует константа $d \in D$, такая, что формула $G(d/x)$ – истинная, иначе формула P ложная.

f5. Если $P = \forall x(A)G$, то x – единственная свободно входящая в формулу G переменная. Формула P истинная, если для каждой константы $d \in D$ формула $G(d/x)$ истинная, иначе формула P ложная.

f6. Если $P = (G)$, то формула P истинная, если формула G истинная, и ложная, если формула G ложная.

Пусть $E = \{x_1, \dots, x_n \mid P(x_1, \dots, x_n)\}$ – выражение исчисления на домене. Значением выражения E назовем таблицу схемы $R = \{A_1, \dots, A_n\}$, содержащую все строки $s \in S(R)$, $s = \{\langle A_1, d_1 \rangle, \dots, \langle A_n, d_n \rangle\}$, такие, что формула $P(d_1/x_1, \dots, d_n/x_n)$ истинная.

Сведение исчисления строк к исчислению на домене

Пусть $F = \{y(R) \mid P(y)\}$ – выражение исчисления строк, переменная y – единственная переменная, входящая в формулу P свободно, а другие переменные связаны не более, чем одним квантором, $R = \{A_1, \dots, A_n\}$ – схема. Применяем y как синтаксическую переменную, областью изменения которой являются переменные исчисления строк. Областью интерпретации предметных переменных исчисления строк есть множество всех строк. Построим выражение исчисления на домене, эквивалентное заданному выражению исчисления строк F . Каждая переменная z исчисления строк, входящая в формулу P , соответствует некоторой схеме R_2 таблицы, причем либо z входит в $\exists z(R_2)$ или $\forall z(R_2)$, либо $z = y$, а $R_2 = R$. Пусть $R_2 = \{A_1, \dots, A_m\}$. Выполним следующие замены.

- Переменную z исчисления строк заменим m переменными z_1, \dots, z_m исчисления на домене.
- Каждый атом $t(z)$ исчисления строк заменим на $t(z_1, \dots, z_m)$.

- Каждый атом $p(\mathbf{v}_1, \dots, \mathbf{v}_m)$, где \mathbf{v}_i – термы исчисления строк, $i = 1, \dots, m$, заменим на $p(\mathbf{u}_1, \dots, \mathbf{u}_m)$, где \mathbf{u}_j – термы исчисления на домене, $j = 1, \dots, m$.
- Формулу с квантором $\exists \mathbf{z}(R_2)G$ преобразуем в $\exists \mathbf{z}_1(A_1) \dots \exists \mathbf{z}_m(A_m)G$.
- Формулу с квантором $\forall \mathbf{z}(R_2)G$ преобразуем в $\forall \mathbf{z}_1(A_1) \dots \forall \mathbf{z}_m(A_m)G$.
- Если $\mathbf{z} = \mathbf{y}$, а $R_2 = R$, то заменим $\mathbf{y}(R)$ в начале выражения F на $\mathbf{z}_1, \dots, \mathbf{z}_m$.

В результате проведенных замен получим, очевидно, выражение исчисления с переменными на домене, эквивалентное заданному выражению исчисления строк. Поэтому можем сформулировать следующую теорему.

Теорема. Если F – произвольное разрешенное выражение исчисления строк, то можно эффективно (и равномерно) построить эквивалентное ему выражение E исчисления на домене.

Сведение исчисления на домене к табличной алгебре

Рассматриваем два множества: \mathbf{A} – множество атрибутов и \mathbf{D} – универсальный домен. Под табличной алгеброй понимаем алгебру $\langle \mathbf{T}, \Omega_{P, \Xi} \rangle$, где \mathbf{T} – множество всех таблиц, $\Omega_{P, \Xi} = \{ \cup_R, \cap_R, \setminus_R, \sigma_{p, R}, \pi_{X, R}, \otimes_{R_1, R_2}, \div_{R_2}, Rt_{\xi, R}, \sim_R \}_{p \in P, \xi \in \Xi, X \subseteq R, R_1, R_2 \subseteq A}$ – сигнатура, P, Ξ – множества параметров. Выражением табличной алгебры называется любое выражение, построенное из таблиц множества \mathbf{T} при использовании операций множества $\Omega_{P, \Xi}$.

Теорема. Для каждого выражения исчисления на домене E можно эффективно (и равномерно) построить эквивалентное ему выражение F табличной алгебры.

Доказательство. Пусть $E = \{ \mathbf{x}_1, \dots, \mathbf{x}_n \mid P(\mathbf{x}_1, \dots, \mathbf{x}_n) \}$ – выражение исчисления на домене, значением которого есть таблица схемы $R = \{ A_1, \dots, A_n \}$. Построим для каждой подформулы G формулы P эквивалентное алгебраическое выражение F_G . Докажем индукцией по числу операторов в подформуле G из P , что если G содержит свободные переменные на универсальном домене $\mathbf{y}_1, \dots, \mathbf{y}_m$, то для выражения $\{ \mathbf{y}_1, \dots, \mathbf{y}_m \mid G(\mathbf{y}_1, \dots, \mathbf{y}_m) \}$ существует эквивалентное выражение табличной алгебры F_G . Тогда, когда G есть само P , получим алгебраическое выражение для $\{ \mathbf{x}_1, \dots, \mathbf{x}_n \mid P(\mathbf{x}_1, \dots, \mathbf{x}_n) \}$. Предполагается, что $\mathbf{y}_1, \dots, \mathbf{y}_m$ – все свободные переменные в G , и значением выражения $\{ \mathbf{y}_1, \dots, \mathbf{y}_m \mid G(\mathbf{y}_1, \dots, \mathbf{y}_m) \}$ является таблица схемы $R_G = \{ B_1, \dots, B_m \} \subseteq \mathbf{A}$.

Заменим связанные переменные в формуле P так, чтобы ни одна переменная не входила одновременно и свободно, и связано в P , а также не связывалась квантором в двух разных местах. Каждая переменная ассоциирована теперь с атрибутом, либо входящим в квантор $\forall \mathbf{x}(\mathcal{A})$ или $\exists \mathbf{x}(\mathcal{A})$, когда переменная связана в формуле P , либо сопоставленным с переменной, стоящей слева от вертикальной черты.

Для каждой формулы P исчисления на домене можно выписать алгебраическое выражение со значением \mathbf{D} (значением данного алгебраического выражения есть одноатрибутная таблица $\langle t, \{A\} \rangle$, содержащая все строки $s = \langle \langle A, d \rangle \rangle$, $d \in \mathbf{D}$ – все элементы множества \mathbf{D}). Например, $\langle \langle A, d \rangle \rangle \cup \overline{\langle \langle A, d \rangle \rangle}$, где d – произвольный элемент из \mathbf{D} . Далее $[D]$ будет обозначать алгебраическое выражение со значением \mathbf{D} .

База индукции (нет операторов, т.е. логических связей и кванторов). Подформула G – атом вида $p(\mathbf{u}_1, \dots, \mathbf{u}_m)$ или $t(\mathbf{a}_1, \dots, \mathbf{a}_n)$.

1) Пусть G имеет вид $p(u_1, \dots, u_m)$, где u_i – термы исчисления на домене. Положим F_G равным $\sigma_{\tilde{p}}([D]_1 \otimes_{R_1, R_2} \dots \otimes_{R_{k-1}, R_k} [D]_k)$, где R_i – одноатрибутные схемы таблиц, которые есть значениями алгебраических выражений $[D]_i$, $i = 1, \dots, k$. Соединение в данном случае является декартовым произведением. А предикат-параметр селекции задан как $\tilde{p}(s) = T \Leftrightarrow p(s(A_1), \dots, s(A_m)) = T$, $s \in S(R)$, где p – m -арный сигнатурный предикатный символ. \square

2) Пусть G имеет вид $t(a_1, \dots, a_n)$, где a_i – константа или переменная на универсальном домене D . Пусть $R = \{C_1, \dots, C_k\}$ – схема таблицы $\langle t, R \rangle$. Положим алгебраическое выражение F_G равным $\pi_X(\sigma_{\tilde{p}}(t))$, где \tilde{p} – предикат-параметр селекции, который является конъюнкцией сравнений $C_i = a_i$ по всех i , для которых a_i – константа; X – множество атрибутов $\{C_j \mid a_j \text{ – переменная}\}$. \square

Шаг индукции. Предположим, что подформула G содержит по крайней мере один оператор и что гипотеза индукции справедлива для всех подформул формулы P , имеющих меньше операторов, чем G .

Случай 1. $G = \neg Q$. Пусть F_Q – алгебраическое выражение для Q . Положим $F_G = \neg F_Q$. \square

Случай 2. $G = Q \vee Q'$. Пусть свободные переменные из Q – это $z_1, \dots, z_k, v_1, \dots, v_p$, а из Q' – это $z_1, \dots, z_k, w_1, \dots, w_q$, где переменные v_1, \dots, v_p и w_1, \dots, w_q попарно не совпадают. Пусть F_Q и $F_{Q'}$ – алгебраические выражения для Q и Q' соответственно, причем C_i , $i = 1, \dots, p$ – атрибуты ассоциированные с переменными v_1, \dots, v_p , а K_j , $j = 1, \dots, q$ – атрибуты ассоциированные с переменными w_1, \dots, w_q . Положим $F_1 = F_Q \otimes_{R_Q, \{K_1\}} [D]_1 \otimes_{\{K_1\}, \{K_2\}} \dots \otimes_{\{K_{q-1}\}, \{K_q\}} [D]_q$, а

$F_2 = F_{Q'} \otimes_{R_{Q'}, \{C_1\}} [D]_1 \otimes_{\{C_1\}, \{C_2\}} \dots \otimes_{\{C_{p-1}\}, \{C_p\}} [D]_p$, где R_Q и $R_{Q'}$ – схемы таблиц, которые являются значениями алгебраических выражений F_Q и $F_{Q'}$ соответственно, а C_i и K_j – одноатрибутные схемы таблиц, которые являются значениями алгебраических выражений $[D]_i$ и $[D]_j$ соответственно, $i = 1, \dots, p$, $j = 1, \dots, q$. Обозначим схемы таблиц, которые являются значениями алгебраических выражений F_1 и F_2 как R_{F_1} и R_{F_2} соответственно. Для соответствующим образом построенных выражений F_1 и F_2 , их схемы R_{F_1} и R_{F_2} равны. Поэтому положим $F_G = F_1 \cup_{R_{F_1}} F_2$. \square

Случай 3. $G = Q \wedge Q'$. Это выражение исчисления на домене можно выразить через операции отрицания и дизъюнкции. Следуя закону де Моргана $G = \neg(\neg Q \vee \neg Q')$. \square

Случай 4. $G = \exists x(A)Q$. Пусть F_Q – алгебраическое выражение для Q . Положим F_G равным $\pi_{X \setminus A}(F_Q)$, где X – схема таблицы, которая является значением алгебраического выражения F_Q . \square

Случай 5. $G = \forall x(A)Q$. Пусть F_Q – алгебраическое выражение для Q . Положим F_G равным $F_Q \div_{R_1, R_2} [D]$, где R_1 – схема таблицы, которая является значением алгебраического выражения F_Q , R_2 – схема таблицы, которая является значением алгебраического выражения $[D]$. \square

Выводы

Классическое исчисления на домене пополнено произвольными предикатными и функциональными сигнатурами на универсальном домене (в то время как обычно рассматривают лишь бинарные предикаты, а функциональная сигнатура вообще пуста). Определено синтаксис термов, атомов и формул исчисления на домене; выделен класс разрешенных формул. Проведена редукция исчисления строк к исчислению на домене. Доказано, что при этом обобщении табличная алгебра остается не менее

выразительной, чем исчисление на домене. Учитывая, что ранее доказано, что исчисление кортежей не менее выразительно, чем табличная алгебра, можно сделать вывод, что табличная алгебра, исчисление строк и исчисление на домене одинаково выразительны.

Література

[Буй, Глушко, 2010] Д.Б. Буй, І.М. Глушко. Узагальнене числення рядків // Наукові записки НаУКМА. Серія: Комп'ютерні науки. Київ, 2010.

[Buy, Glushko, 2010] Generalized table algebra, generalized tuple calculus and theirs equivalence // Proceeding of the CSE'2010 International Scientific Conference on Computer Science and Engineering. – Košice - Stará Ľubovňa, Slovakia. – 2010, September 20-22.

[Lacroix, Pirotte, 1977] M. Lacroix, A. Pirotte. Domain-oriented relational languages / M. Lacroix, A. Pirotte. In: Proc. 3rd Int. Conf on Very Large Data Bases. Tokyo, October, 1977. – P. 370-378.

[Дейт, 2005] К. Дж. Дейт. Введение в системы баз данных: [8-е изд.: пер. с англ.]. Москва: Издательский дом «Вильямс», 2005.

[Хомоненко, Цыганков, Мальцев, 2004] А. Д. Хомоненко, В. М. Цыганков, М. Г. Мальцев. Базы данных. Санкт-Петербург: КОРОНА принт, 2004.

[Редько, Брона, Буй, Поляков, 2001] Реляційні бази даних: табличні алгебри та SQL-подібні мови / В. Н. Редько, Ю. Й. Брона, Д. Б. Буй, С. А. Поляков. Київ: Видавничий дім «Академперіодика», 2001.

[Мейєр, 1987] Д. Мейєр. Теория реляционных баз данных. Москва: Мир, 1987.

[Ульман, 1983] Дж. Ульман. Основы систем баз данных. Москва: Финансы и статистика, 1983.

[Коннолли, Бегг, 2003] Т. Коннолли, К. Бегг. Базы данных. Проектирование, реализация и сопровождение. Теория и практика: [3-изд.: пер. с англ.]. Москва: Издательский дом «Вильямс», 2003.

Информация об авторах

Дмитрий Буй – доктор физико-математических наук, старший научный сотрудник, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики: Украина, Киев, 03680, пр. Глушкова 2, корп. 6; e-mail: buy@unicyb.kiev.ua.

Ирина Глушко – аспирантка, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики: Украина, Киев, 03680, пр. Глушкова 2, корп. 6; e-mail: glushkoim@gmail.com.