УДК 004.652

Математические основания реляционных баз данных. Часть 2: свойства обобщенных табличных операций

Н. Д. Кахута

Университет экономики и права «КРОК», Украина

Статья посвящена созданию фрагмента теории табличных алгебр, построенных на основе классических реляционных алгебр Кодда. Отличительной особенностью применяемой техники является использование свойств теоретико-множественных конструкций и их перенесение на табличный случай. Такое перенесение свойств возможно ввиду наличия простых представлений сигнатурных операций в терминах указанных теоретико-множественных конструкций. Наличие указанных представлений делает возможным рассмотрение обобщенной табличной алгебры, получаемой снятием требований, во-первых, конечности таблиц и, во-вторых, односхемности строк таблицы.

Ключевые слова: *реляционные алгебры Кодда, полный образ, ограничение, прямое произведение, совместность, коинициальность, обобщенные табличные алгебры.*

Стаття присвячена створенню фрагмента теорії табличних алгебр, побудованих на основі класичних реляційних алгебр Кодда. Особливістю застосовуваної техніки є використання властивостей певних теоретико-множинних конструкцій та їх перенесення на табличний випадок. Таке перенесення властивостей можливо через наявність простих зображень сигнатурних операцій в термінах теоретико-множинних конструкцій. Наявність зазначених зображень робить можливим розгляд узагальненої табличної алгебри, яка одержується зняттям вимог, по-перше, скінченності таблиць і, по-друге, односхемності рядків таблиці.

Ключові слова: *реляційні алгебри Кодда, повний образ, обмеження, прямий добуток, сумісність, коініціальность, узагальнені табличні алгебри.*

The article is devoted to creation of a fragment of the theory for table algebras, which constitute a generalization of classical Codd's relational algebras. A distinctive feature of this technique is that the set-theoretic properties of some specific constructions are used and transferred to the table case. Such transfer is possible because there exist simple representations of signature operations in terms of these set-theoretic constructions. The fact of these representations existence allows to get rid of some requirements to generalized table algebra: firstly, tables have not to be the finite set of rows, and secondly, rows in a table can have different schemas.

Key words: *Codd's relational algebras, a whole image, restriction, Cartesian product, consistency, coinitiality, generalized table algebras.*

1. Введение

Статья продолжает работу [1] и посвящена созданию фрагмента теории табличных алгебр [2-11], построенных на основе классических реляционных алгебр Кодда [12-18]. Отличительной особенностью применяемой техники является использование свойств теоретико-множественных конструкций полного образа, ограничения, обобщенного прямого произведения, бинарных отношений совместности и коинициальности (конфинальности) [19-22] и перенесение свойств указанных конструкций для табличного случая на основе представлений табличных операций, полученных в первой части работы [1].

2. Свойства табличных операций, полурешетка таблиц

В следующих утверждениях приведены свойства операций табличных алгебр, для установления которых явно используются соответствующие свойства указанных выше теоретико-множественных конструкций.

2.1. Свойства проекции, соединения, насыщения и переименования

Предложение 1 (монотонность проекции, соединения, насыщения, переименования; дистрибутивность проекции, соединения, переименования относительно объединений). Операции проекции, соединения, переименования, насыщения монотонны по теоретико-множественному включению:

1) $t_1 \subseteq t_2 \Rightarrow \pi_X(t_1) \subseteq \pi_X(t_2)$ (монотонность проекции);

2) $t_1 \subseteq t_1' \,\&\, t_2 \subseteq t_2' \Rightarrow t_1 \otimes t_2 \subseteq t_1' \otimes t_2'$ (монотонность соединения по совокупности аргументов);

3) $t_1 \subseteq t_2 \Leftrightarrow Rt_\xi(t_1) \subseteq Rt_\xi(t_2)$, где t_1, t_2 имеют одинаковую ξ-допустимую схему (монотонность переименования);

4) $t_1 \subseteq t_2 \Rightarrow D_{A,t_1} \subseteq D_{A,t_2}$ (монотонность активного домена); $t_1 \subseteq t_2 \Rightarrow C(t_1) \subseteq C(t_2)$ (монотонность насыщения).

Операции проекции, соединения, переименования дистрибутивны относительно теоретико-множественного объединения:

5) $\pi_X(\bigcup\limits_i t_i) = \bigcup\limits_i \pi_X(t_i)$, здесь схемы таблиц t_i одинаковы (дистрибутивность проекции относительно объединений);

6) $t \otimes (\bigcup\limits_i t_i) = \bigcup\limits_i t \otimes t_i$ (дистрибутивность соединения относительно объединений по второму аргументу), здесь схемы таблиц t_i одинаковы (для первого аргумента аналогично);

7) $Rt_\xi(\bigcup\limits_i t_i) = \bigcup\limits_{\eta[R]} Rt_\xi(t_i)$, где таблицы t_i имеют одинаковую ξ-допустимую схему R (дистрибутивность переименования относительно объединений). □

Доказательство. Используя представления соответствующих операций ([1, предложение 1]), монотонность полного образа относительно объединений ([11, подраздел 1.3, предложение 1.3.1; 19]) и обобщенного прямого произведения ([20], это свойство применяется при доказательстве монотонности насыщения) доказываются пп. 1-3; при доказательстве дистрибутивности операций в пп. 5-7 надо также использовать представления операций и дистрибутивность полного образа относительно объединений ([11, подраздел 1.3, предложение 1.3.1;19]). ▫

Установим монотонность активного домена и насыщения. Из представления насыщения (через активный домен и обобщенное прямое произведение, см. [1, предложение 1, п. 3]) и монотонности обобщенного прямого произведения [20] следует, что из монотонности активного домена вытекает монотонность насыщения; поэтому проверим монотонность активного домена.

Для этого воспользуемся следующим очевидным представлением активного домена

$$D_{A,t} = I_A[t], \tag{2.1}$$

где $I_A : S \tilde{\rightarrow} D$, $I_A(s) \simeq s(A)$ для $s \in S$.

Содержательно говоря, частичная операция I_A строке s сопоставляет значение атрибута A в ней. Очевидно, что

$$dom\, I_A = \bigcup_{R:A\in R} S(R)\,. \qquad (2.2)$$

Операция I_A по сути является (обобщенным) селектором.

Из представления (2.1), а также монотонности полного образа и вытекает монотонность активного домена относительно теоретико-множественного включения. ▫□

Установим еще некоторые интересные свойства данных операций. Ниже понятие (первичного, primary key) ключа таблицы понимается стандартно (см. например [11, подраздел 2.4, с. 43; 23]).

Предложение 2 (свойства проекции). Пусть R – схема таблиц t, t_i. Выполняются следующие соотношения:

1) $\pi_X(t) \in T(R \cap X)$; причем любая таблица схемы $R \cap X$ равна проекции по множеству атрибутов X некоторой таблицы схемы R (т.е.отображение $\pi_X : T(R) \to T(R \cap X)$ является сюръективным);

2) $R \subseteq X \Rightarrow \pi_X(t) = t$, причем для непустых таблиц эта импликация переходит в эквивалентность: $t \neq t_{\varnothing} \Rightarrow (R \subseteq X \Leftrightarrow \pi_X(t) = t)$; в частности, $\pi_R(t) = t$;

3) $\pi_X(t) = t_{\varnothing} \Leftrightarrow t = t_{\varnothing}$ (критерий пустоты проекции, сохранение пустой таблицы $t_{\varnothing}$);

4) $\pi_X(t) = t_{\varepsilon} \Leftrightarrow t \neq t_{\varnothing}\ \&\ X \cap R = \varnothing$; в частности, $\pi_X(t_{\varepsilon}) = t_{\varepsilon}$ (сохранение специальной таблицы t_{ε});

5) $\pi_X(\bigcap_i t_i) \subseteq \bigcap_i \pi_X(t_i)$ (верхняя оценка проекции пересечений); X – ключ таблиці $\bigcup_i t_i \Rightarrow \pi_X(\bigcap_i t_i) = \bigcap_i \pi_X(t_i)$ (достаточное условие дистрибутивности проекции относительно пересечений);

6) $\pi_X(t_1 \setminus_R t_2) \supseteq \pi_X(t_1) \setminus \pi_X(t_2)$ (нижняя оценка проекции разности);

7) $\pi_X(\pi_Y(t)) = \pi_{X \cap Y}(t)$, или в операторном виде $\pi_X \circ \pi_Y = \pi_Y \circ \pi_X = \pi_{X \cap Y}$ (композиция проекций); в частности, $\pi_X(\pi_X(t)) = \pi_X(t)$ (идемпотентность проекции) и $\pi_X(t) = \pi_{X \cap R}(t)$.□

Доказательство. При проверке п. 1 надо воспользоваться определением проекции, свойствами ограничения ([11, подраздел 1.3, предложение 1.3.2; 19]): $\pi_1^2(s \mid X) = \pi_1^2 s \cap X = R \cap X$, где $s \in t$, а $t \in T(R)$. Кроме того, надо учесть, что любая строка схемы $R \cap X$ является ограничением по множеству X некоторой строки схемы R. ▫

Рассмотрим п. 2. Пусть $R \subseteq X$. Используя свойство ограничения из [11, подраздел 1.3, предложение 1.3.2; 19], имеем $\pi_X(t) = \{s \mid X \mid s \in t\} = \{s \mid s \in t\} = t$.

Пусть теперь $t \neq t_{\varnothing}$. Предположим, что $\pi_X(t) = t$ и установим включение $R \subseteq X$. Выберем строку $s_0 \in t$. Тогда $s_0 | X \in \pi_X(t)$, т.е. $s_0 | X = s'$ для некоторой строки $s' \in t$, следуя допущению. Согласно свойствам ограничения [11, подраздел 1.3, предложение 1.3.2; 19] имеем: $\pi_1^2(s_0 | X) = \pi_1^2 s_0 \cap X = R \cap X = \pi_1^2 s' = R$, откуда $R \cap X = R$, т.е. $R \subseteq X$. ▫

Достаточность в п. 3 вытекает из представления проекции через полный образ ([1, предложение 1, п. 1]) и сохранения пустого множества $\varnothing$ полным образом ([11, подраздел 1.3, предложение 1.3.1;19]). Для доказательства необходимости надо учесть те же факты и, кроме того, тотальность ограничения вида $\uparrow X : S \to S$. ▫

Рассмотрим п. 4. Пусть $\pi_X(t) = t_{\varepsilon}$. Из п. 3 вытекает, что $t \neq t_{\varnothing}$ (проверяется от противного), покажем, что $X \cap R = \varnothing$. Зафиксируем строку $s_0 \in t$. Тогда $s_0 | X \in \pi_X(t) = \{\varepsilon\}$, т.е. $s_0 | X = \varepsilon$. Возьмем проекции от обеих частей последнего равенства, тогда по свойствам ограничения [11, подраздел 1.3, предложение 1.3.2; 19] имеем $\pi_1^2 s_0 \cap X = R \cap X = \varnothing$. Обратная импликация устанавливается аналогично. ▫

Включения пп. 5-6 вытекают из представления проекции ([1, предложение 1, п. 1]), верхней оценки полного образа пересечения ([11, подраздел 1.3, предложение 1.3.1; 19]) и нижней оценки полного образа разности ([11, подраздел 1.3, предложение 1.3.1; 19]).

Для доказательства равенства з п. 5 (достаточного условия дистрибутивности проекции относительно пересечений) надо воспользоваться представлением проекции ([1, предложение 1, п. 1]), достаточным условием дистрибутивности полного образа относительно пересечений ([11, подраздел 1.3, предложение 1.3.1; 19]) и следующим очевидным свойством ключа: X – ключ таблицы $t \Leftrightarrow \uparrow X | t : t \to S$ – инъекция (т.е. операция $\uparrow X$ инъективна на таблице t). ▫

Рассмотрим п. 7. На основе свойств полного образа и ограничения ([11, подраздел 1.3, предложения 1.3.1, 1.3.2; 19]) получаем цепочку равенств: $\pi_X(\pi_Y(t)) = \uparrow X[\uparrow Y[t]] = \uparrow X \circ \uparrow Y[t] = (\uparrow X \cap Y)[t] = \pi_{X \cap Y}(t)$. Отсюда и из доказанного п. 2 вытекают равенства $\pi_X(t) = \pi_X(\pi_R(t)) = \pi_{X \cap R}(t)$. ▫□

Предложение 3 (свойства активного домена и насыщения). Пусть *R* – схема таблицы *t*. Выполняются следующие утверждения:

1) $D_{A,t} = \varnothing \Leftrightarrow t = t_{\varnothing} \vee A \notin R$ (критерий пустоты активного домена); $C(t) = t_{\varnothing} \Leftrightarrow t = t_{\varnothing}$ (критерий пустоты насыщения, сохранение пустой таблицы $t_{\varnothing}$ насыщением); $C(t) = t_{\varepsilon} \Leftrightarrow t = t_{\varepsilon}$ (сохранение специальной таблицы t_{ε} насыщением);

2) $t \subseteq C(t)$ (возрастание насыщения по теоретико-множественному включению). □

Доказательство. Проверим п. 1 и начнем с критерия пустоты активного домена. Для этого воспользуемся представлением активного домен (2.1), видом области определенности $dom\, I_A$ (2.2) и критерием пустоты полного образа ([11, подраздел 1.3, предложение 1.3.1; 19]), имеем цепочку эквивалентностей: $D_{A,t} = \varnothing \Leftrightarrow I_A[t] = \varnothing \Leftrightarrow t \bigcap domI_A = \varnothing \Leftrightarrow t \bigcap \bigcup_{\widetilde{R}:A\in\widetilde{R}} S(\widetilde{R}) = \varnothing$.

Поскольку по предположению таблица t имеет схему R, то $t \subseteq S(R)$ и остается учесть следующую кусочную схему, корректность которой проверяется непосредственно:

$$t \bigcap \bigcup_{\widetilde{R}:A\in\widetilde{R}} S(\widetilde{R}) = \begin{cases} t, если\, A \in R, \\ \varnothing, если\, A \notin R. \end{cases}$$

Проверим критерий пустоты насыщения. Достаточность вытекает из представления насыщения [1,предложение 1, п. 3], критерия пустоты обобщенного прямого произведения [11, подраздел 1.3, предложение 1.3.5; 20] и доказанного критерия пустоты активного домена. ▫

Необходимость, т.е. импликацию $C(t) = t_\varnothing \Rightarrow t = t_\varnothing$, докажем от противного.

Итак, пусть $t \neq t_\varnothing$. Очевидно, что $t \neq t_\varepsilon$ (действительно, это проверяется от противного с учетом фактов $C(t_\varepsilon) = t_\varepsilon, t_\varepsilon \neq t_\varnothing$). Значит, схема R таблицы не пуста. Согласно представлению насыщения имеем равенство $C(t) = \prod_{A\in R} D_{A,t}$; поскольку таблица t непуста, то в силу критерия пустоты активного домена все множества $D_{A,t}$, $A \in R$, непусты. Но отсюда по критерию пустоты обобщенного прямого произведения вытекает, что и произведение $\prod_{A\in R} D_{A,t}$ непусто; пришли к противоречию. ▫

Остается проверить эквивалентность $C(t) = t_\varepsilon \Leftrightarrow t = t_\varepsilon$. Достаточность вытекает прямо из определения насыщения; установим необходимость. Пусть $C(t) = t_\varepsilon$, т.е. $C(t)$ имеет пустую схему. Поскольку по определению насыщения эта операция схему не меняет, то и таблица t имеет пустую схему. Но существуют всего две таблицы пустой схемы – $t_\varnothing$ и t_ε, причем случай $t = t_\varnothing$ невозможен (поскольку тогда бы выполнялось равенство $C(t) = t_\varnothing$ согласно доказанному критерию пустоты насыщения). ▫

Рассмотрим п. 2. Случай таблиц пустой схемы $t \in T(\varnothing)$ проверяется непосредственно; пусть далее $t \notin T(\varnothing)$, т.е. $R \neq \varnothing$. Пусть $s \in t$, тогда по свойствам обобщенного прямого произведения [11, подраздел 1.3, предложение 1.3.5; 20] имеем принадлежность $s \in \prod_{A\in R} \{s(A)\}$; поскольку для всех $A \in R$ выполняется включение $\{s(A)\} \subseteq D_{A,t}$, то вследствие монотонности оператора $\prod$ ([11, подраздел 1.3, предложение 1.3.5; 20]) имеем принадлежность $s \in \prod_{A\in R} D_{A,t} = C(t)$. ▫□

2.2. Полугруппа и нижняя полурешетка таблиц по соединению

Ниже приведены основные свойства бинарной операции соединения.

Предложение 4 (свойства соединения). Пусть таблицы t_1, t_2 имеют схемы R_1, R_2 соответственно. Тогда выполняются следующие соотношения:

1) $t_1 \otimes t_2 \in T(R_1 \cup R_2)$;

2) $t_1 \otimes t_2 = t_2 \otimes t_1$ (коммутативность соединения);

3) $t_1 \otimes (t_2 \otimes t_3) = (t_1 \otimes t_2) \otimes t_3$ (ассоциативность соединения);

4) $t \otimes t_\varnothing = t_\varnothing \otimes t = t_\varnothing$ (сохранение пустой таблицы $t_\varnothing$);

5) $t \otimes t_\varepsilon = t_\varepsilon \otimes t = t$ (специальная таблица t_ε – правая и левая единица);

6) $R_2 \subseteq R_1 \Rightarrow t_1 \otimes t_2 \subseteq t_1$;

7) $t \otimes t = t$ (идемпотентность соединения). □

Доказательство. Принадлежность п. 1 вытекает непосредственно из определения соединения. Содержательная интерпретация следующая: схема соединения таблиц получается теоретико-множественным объединением схем таблиц-аргументов. ▫

Пп. 2, 3 вытекают из коммутативности и ассоциативности операции $\bar{\cup}$ [11, подраздел 2.7, лемма 2.7.2], утверждении о наследовании коммутативности и ассоциативности операции при распространении с элементов на множества элементов (с помощью полного образа) [19], а также из представления соединения [1,предложение 1, п. 2].▫

Доказательство п. 4. вытекает из представления операции соединения и сохранения пустого множества декартовым произведением и полным образом [11, подраздел 1.3, предложение 1.3.1; 19]. ▫

Равенства п. 5 вытекают из определения операции соединения и свойств отношения совместности ([11, подраздел 2.7, лемма 2.7.2]; напомним, что специальная строка ε представляет собой пустое множество): $t \otimes t_\varepsilon = \bar{\cup}[t \times \{\varepsilon\}] = \bar{\cup}[\{< s, \varepsilon > | s \in t\}] = \{s \cup \varepsilon \,|\, s \in t\} = \{s \,|\, s \in t\} = t$.▫

Рассмотрим п. 6. Пусть $s \in t_1 \otimes t_2$, тогда $s = s_1 \cup s_2$ для подходящих строк $s_1 \in t_1$, $s_2 \in t_2$, причем $s_1 \approx s_2$. Так как $R_2 \subseteq R_1$, то по свойствам ограничения [11, подраздел 1.3, предложение 1.3.2; 19] имеем $s_2 = s_2|R_2 \subseteq s_2|R_1$. Согласно свойств отношения совместности [11, подраздел 2.7, лемма 2.7.1] выполняется неравенство $s_2|R_1 \subseteq s_1$. Отсюда вытекает включение $s_2 \subseteq s_1$, т.е. $s = s_1 \cup s_2 = s_1 \in t_1$. Таким образом, $t_1 \otimes t_2 \subseteq t_1$. ▫

Переходим к последнему п. 7. Из предыдущего пункта вытекает включение $t \otimes t \subseteq t$, поэтому остается проверить обратное включение. Пусть $s \in t$, тогда $s = s \cup s$ (идемпотентность объединения функций), причем $s \approx s$ (рефлексивность отношения совместности). Отсюда по определению соединения вытекает включение $s \in t \otimes t$ ▫□

Для введения частичных порядков на носителе табличных алгебр установим дискретность таблиц относительно теоретико-множественного включения на множестве строк, т.е. дискретность частично упорядоченного множества (ч. у. м.) вида $\langle t, \subseteq \rangle$; дискретность понимается стандартно (см., например, [24]: $\forall s_1 \forall s_2 (s_1, s_2 \in t \Rightarrow (s_1 \subseteq s_2 \Rightarrow s_1 = s_2))$. Этот факт непосредственно вытекает из свойств ограничения [11, подраздел 1.3, предложение 1.3.2; 19], здесь существенно, что строки одной таблицы имеют одинаковые схемы.

Введем на множестве всех таблиц T бинарное отношение $\lhd$:

$$t_1 \lhd t_2 \overset{def}{\Leftrightarrow} \forall s_1 (s_1 \in t_1 \Rightarrow \exists s_2 (s_2 \in t_2 \,\&\, s_2 \subseteq s_1)). \qquad (2.3)$$

Это отношение допускает содержательную интерпретацию: если $t_1 \lhd t_2$, то информация, представленная в таблице t_1, продолжает часть информации, представленной в таблице t_2.

Предложение 5 (устройство множества таблиц). $< T, \lhd >$ является частично упорядоченным множеством с наименьшим элементом $t_\varnothing$ и наибольшим элементом t_ε. □

Доказательство. Очевидно, что бинарное отношение $\lhd$, заданное в (2.3), является отношение коинициальности, индуцированным теоретико-множественным отношением включения строк. Вследствие дискретности таблиц и упорядочения семейства дискретных множеств отношением коинициальности [22], получаем, что $< T, \lhd >$ является ч. у. м. с наименьшим элементом $t_\varnothing$. ▫

То, что t_ε – наибольший элемент, вытекает из определения отношения $\lhd$ и того очевидного факта, что пустая строка ε – наименьший элемент ч. у. м. $< S, \subseteq >$. С содержательной точки зрения, особая таблица t_ε вообще не содержит информации, поэтому и является наибольшим элементом (отметим, что пустая таблица тоже не содержит информации).▫ □

Далее через $\prod$ обозначается точная нижняя грань (инфимум) множества.

Лемма 1 (связь между частичным порядком $\lhd$ и соединением). Выполняются следующие утверждения:

1) $t_1 \lhd t_1' \wedge t_2 \lhd t_2' \Rightarrow t_1 \otimes t_2 \lhd t_1' \otimes t_2'$ (монотонность соединения относительно порядка $\lhd$ по совокупности аргументов);

2) $t_1 \otimes t_2 = \prod\limits_{\lhd} \{t_1, t_2\}$ (представление соединения через порядок $\lhd$);

3) $t_1 \lhd t_2 \Leftrightarrow t_1 = t_1 \otimes t_2$ (представление порядка $\lhd$ через соединение).□

Доказательство. Начнем с п. 1. Пусть $t_1 \lhd t_1'$ и $t_2 \lhd t_2'$, покажем, что тогда $t_1 \otimes t_2 \lhd t_1' \otimes t_2'$. Случай $t_1 \otimes t_2 = t_\varnothing$ тривиальный, поскольку пустая таблица $t_\varnothing$ есть наименьший элемент ч. у. м. $< T, \lhd >$.

Поэтому пусть далее $t_1 \otimes t_2 \neq t_\varnothing$, рассмотрим произвольную строку $s \in t_1 \otimes t_2$. По определению соединения существуют строки $s_1 \in t_1$, $s_2 \in t_2$, такие, что $s = s_1 \bigcup s_2$, причем $s_1 \approx s_2$. Так как $t_1 \lhd t_1'$ и $t_2 \lhd t_2'$ по предположению, то существуют строки $s_1' \in t_1'$ и $s_2' \in t_2'$, такие, что $s_1' \subseteq s_1$, $s_2' \subseteq s_2$. Отсюда вытекает, что $s_1' \bigcup s_2' \subseteq s_1 \bigcup s_2$, т.е. $s_1' \bigcup s_2'$ является функцией (как подмножество функции); по [11, подраздел 2.7, лемма 2.7.1] отсюда следует, что $s_1' \approx s_2'$. Таким образом, по определению соединения имеем $s_1' \cup s_2' \in t_1' \otimes t_2'$. Остается учесть уже указанное включение $s_1' \cup s_2' \subseteq s_1 \cup s_2$.

Итак, для произвольной строки $s \in t_1 \otimes t_2$ нашлась строка из соединения $t_1' \otimes t_2'$ (а именно $s_1' \cup s_2'$), меньшая s (по теоретико-множественному включению). Поэтому по определению отношения коинициальности имеем требуемое неравенство $t_1 \otimes t_2 \lhd t_1' \otimes t_2'$. ▫

Рассмотрим п. 2. Из определения отношения коинициальности вытекает, что $t_1 \otimes t_2 \lhd t_1$ и $t_1 \otimes t_2 \lhd t_2$; таким образом, $t_1 \otimes t_2 \in \{t_1, t_2\}^{\nabla}$ (в левой части последней принадлежности записан нижний конус, следуя обозначениям [24]) и остается показать, что $t_1 \otimes t_2$ является наибольшим элементом указанного нижнего конуса. Действительно, пусть $t \in \{t_1, t_2\}^{\nabla}$, т.е. $t \lhd t_1$ и $t \lhd t_2$. Из предыдущего доказанного пункта (о монотонности соединения) вытекает неравенство $t \otimes t \lhd t_1 \otimes t_2$; остается учесть идемпотентность соединения (предложение 4, п. 7). ▫

П. 3 следует непосредственно из предыдущего п. 2. ▫ □

Из предложения 4 вытекает, что группоид $< T, \otimes >$ является идемпотентной коммутативной полугруппой. Используя хорошо известную связь между идемпотентными коммутативными полугруппами и полурешетками (см., например, [24; 25]), полугруппу $< T, \otimes >$ можно преобразовать в, например, нижнюю полурешетку, вводя порядок стандартно:

$$t_1 \lhd_{\otimes} t_2 \overset{def}{\Leftrightarrow} t_1 = t_1 \otimes t_2 . \tag{2.4}$$

Следствие 1. Отношение коинициальности на множестве таблиц T совпадает с порядком нижней полурешетки (коммутативной идемпотентной) полугруппы $< T, \otimes >$. □

Доказательство вытекает из определения (2.4) и п. 3 леммы 1. □

3. Обобщенная табличная алгебра

При построении таблиц в работе [1] (первая часть цикла) на них накладывались следующие ограничения: во-первых, строки имели конечные схемы («горизонтальная» конечность); во-вторых, таблицы были конечными множествами строк («вертикальная» конечность) и, в-третьих, строки одной таблицы имели одну схему.

Отметим тот факт, что указанные ограничения, как правило, не использовались в приведенных доказательствах (исключением является подраздел 2.2, где использовалась односхемной строк таблиц, что существенно для обеспечения дискретности таблиц по теоретико-множественному включению).

Снимая эти три ограничения, получим обобщенную табличную алгебру $< \widetilde{T}, \widetilde{\Omega}_{P,\Xi} >$, где $\widetilde{\Omega}_{P,\Xi} \overset{def}{=} \{\cup, \cap, \setminus, \sigma_p, \pi_X, \otimes, \div_{R_2}^{R_1}, Rt_{\xi}, \sim\}_{X, R_1, R_2 \subseteq \boldsymbol{A}}^{p \in P, \xi \in \Xi}$, (чтобы не усложнять обозначения, за операциями обобщенной табличной алгебры сохраняем те же обозначения, что и для табличной алгебры).

В этой алгебре *схемой* называется произвольное множество атрибутов (в частности, бесконечное) $R \subseteq \boldsymbol{A}$, а *таблицей* называется произвольное множество строк (в частности, схемы строк одной таблицы могут буть различными, а количество строк таблицы может быть бесконечным).

Все сигнатурные операции обобщенной табличной алгебры определяются по аналогии с одноименными операциями табличной алгебры.

Самый сложный случай – это случай операции (обобщенного) соединения (обобщенных) таблиц, но представление соединения, указанное в п. 2 предложения 1 из [1], позволяет корректно ввести соответствующее определение.

Что касается деления, то его можно ввести формально как производную операцию, основываясь на хорошо известном представлении (см., например, [11, подраздел 2.13, предложение 2.13.1, п. 10; 26]):

$$t_1 \div_{R_2}^{R_1} t_2 \overset{def}{=} \pi_{R'}(t_1) \setminus_{R'} \pi_{R'}(\pi_{R'}(t_1) \otimes t_2 \setminus_{R_1} t_1), \text{ где } R' = R_1 \setminus R_2 .$$

Правда остается открытым вопрос о области определенности деления; положим, что она равна $T(R_1) \times T(R_2)$.

Кроме того, для введения (обобщенного) переименования надо модифицировать определение множества таблиц ξ -допустимых схем T_ξ : это множество состоит из таблиц, схемы (вообще говоря, разные) всех строк которых должны быть ξ -допустимыми:

$$T_\xi \overset{def}{=} \{t \,|\, \forall s (s \in t \Rightarrow \xi[\pi_1^2(s)] \cap (\pi_1^2(s) \setminus dom\xi) = \varnothing)\} .$$

Рассматривая две введенные в работе алгебры таблиц, имеем очевидную связь между ними.

Теорема 1. Табличная алгебра является собственной подалгеброй обобщенной табличной алгебры. □

Доказательство следует непосредственно из построения указанных алгебр. □

Теорема 2. Все утверждения предложений 1-4 выполняются в обобщенной табличной алгебре. □

Доказательство полностью повторяет соответствующие доказательства для табличной алгебры (поскольку в доказательствах не оиспользовались требования конечности таблиц и односхемности строк одной таблицы). □

4. Основные результаты

1. Показана монотонность операций проекции, соединения, насыщения, переименования относительно теоретико-множественного включения (предложение 1, пп. 1-4).

2. Показана дистрибутивность относительно объединений операций проекции, соединения, переименования (предложение 1, пп. 5-7); приведено естественное достаточное условие в терминах первичного ключа дистрибутивности проекции относительно пересечений (предложение 2, п. 5).

3. Установлено сохранение специальных таблиц $t_{\varnothing}, t_{\varepsilon}$ проекцией (предложение 2, пп. 3, 4); дана верхняя оценка проекции пересечений (предложение 2, п. 5) и нижняя оценка проекции разности (предложение 2, п. 6); рассмотрен случай композиции проекций, показана идемпотентность проекции (предложение 2, п. 7).

4. Установлено сохранение специальных таблиц $t_{\varnothing}, t_{\varepsilon}$ насыщением (предложение 3, п. 1); показано, что насыщение является возрастающим оператором по теоретико-множественному включению (предложение 3, п. 2).

5. Установлена коммутативность и ассоциативность соединения (предложение 4, пп. 2, 3); показано, что соединение сохраняет пустую таблицу $t_{\varnothing}$ (предложение 4, п. 4), а специальная таблица t_{ε} является единицей (правой и левой) по соединению; установлена идемпотентность соединения (предложение 4, п. 7).

6. На множестве всех таблиц по логической схеме отношения коинициальности введено бинарное отношение $\lhd$ и показано, что $< T, \lhd >$ является частично упорядоченным множеством с наименьшим элементом $t_{\varnothing}$ и наибольшим элементом t_{ε} (предложение 5); показано, что порядок $\lhd$ совпадает с порядком нижней полурешетки, соответствующей коммутативной идемпотентной полугруппе $< T, \otimes >$ (следствие 1).

7. Снятием ограничений финитности схем таблиц и самих таблиц, а также требования односхемности строк таблиц построена обобщенная табличная алгебра, которая содержит табличную алгебру в качестве подалгебры (теорема 1).

8. Показано, что все утверждения предложений 1-4 выполняются и для обобщенно табличной алгебры.

5. Заключение

В работе построен содержательный фрагмент теории табличных алгебр. Основная особенность используемой техники заключается, во-первых, в установлении естественных представлений сигнатурных операций этих алгебр в терминах теоретико-множественных конструкций полного образа, ограничения, обобщенного прямого соединения, коинициальности, отношения совместности (все это было сделано в работе [1], являющейся первой часть данного цикла) и, во-вторых, в переносе свойств указанных конструкций на табличный случай.

Это демонстрирует мощность используемого аппарата и, по мнению автора, указанный теоретико-множественный аппарат может успешно применяться и в других областях.

Применение указанных представлений дало возможность не только исследовать операции табличных алгебр, но и построить обобщенную табличную алгебру снятием требований конечности схем, таблиц и ограничения односхемности строк таблиц. Установленный результат позволит подойти к построению моделей данных для NoSQL баз данных [27-29].

Отметим, что вопрос о введении естественного порядка на носителе обобщенной табличной алгебры пока остается открытым.

Отметим также, что многие аспекты реальных табличных систем управления базами данных (например, наличие дубликатов строк в таблицах, агрегатные функции, внешние соединения) не нашли своего отражения в табличных алгебрах, поэтому сама модель требует расширения. Это будет предметом рассмотрения в следующих работах.

ЛИТЕРАТУРА

1. Кахута Н.Д. Математические основания современных реляционных систем управления базами данных. Часть 1: теоретико-множественные представления основных табличных операций // Вестник Харковского национального университета. Серия "Математическое моделирование. Информационные технологии. Автоматизированные системы управления".– 2014. – Вып. ??. – С. ??-??.
2. Брона Ю.Й. Основні співвідношення в табличних алгебрах // Вісник Київського університету. Сер.: фіз.-мат. науки. – 1997. – Вип. 3. – С. 143-148.
3. Брона Ю.Й. Оптимізація обчислення запитів у реляційних базах даних // Питання оптимізації обчислень: міжнародна конференція, 6-8 жовтня 1997 р., Київ, ІК ім. В.М. Глушкова НАНУ: праці. – 1997. – С. 45-49.
4. Буй Д.Б., Брона Ю.Й. Операторы замыкания в теории реляционных баз данных // Тезисы докладов XI Международной конференции по проблемам теоретической кибернетики. Под ред. С.В. Яблонского. – Ульяновск: Изд-во СВНЦ. – 1996. – С. 29-30.
5. Буй Д.Б., Брона Ю.Й. Теоретико-множинні конструкції в теорії реляційних баз даних // Вісник Київського університету. Сер.: фіз.-мат. науки. – 1996. – Вип. 1. – С. 216-224.
6. Буй Д.Б. Теорія програмних алгебр композиційного типу та її застосування: дисертація доктора фізико-математичних наук: 01.05.03 – математичне та програмне забезпечення обчислювальних машин і систем / Буй Дмитро Борисович. – Київ, 2002. – 365 с.
7. Редько В.Н., Брона Ю.Й., Буй Д.Б. Информационный аспект Case-технологий: основные соотношения в табличных алгебрах // Проблемы программирования. – 1997. – Вып. 1. – С. 5-11.
8. Редько В.Н., Буй Д.Б. К основаниям теории реляционных моделей баз данных // Кибернетика и системный анализ. – 1996. – № 4. – С. 3-12.
9. Редько В.Н. Реляционные алгебры: операции деления и переименования / В.Н. Редько, Ю.И. Брона, Д.Б. Буй // Кибернетика и системный анализ. – 1997. – № 5. – С. 3-15.
10. Редько В.Н., Брона Ю.Й., Буй Д.Б. Реляционные алгебры: операции проекции и соединения // Кибернетика и системный анализ. – 1997. – № 4. – С. 89-100.
11. 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. – 198 с.
12. Codd E.F. A Relational Model of Data for Large Shared Data Banks // Communications of the ACM. – 1970. – Vol. 13, № 6. – P. 377-387.

13. Codd E.F. A Data Base Sublanguage Founded on the Relational Calculus // ACM-SIGFIDET Workshop on Data Description, Access and Control: international conference, November 11-12, 1971, San Diego, California: proceedings. – 1971. – P. 35-68.
14. Codd E.F. Normalized Data Base Structure: A Brief Tutorial // ACM-SIGFIDET Workshop on Data Description, Access and Control: international conference, November 11-12, 1971, San Diego, California: proceedings. – 1971. – P. 1-17.
15. Codd E.F. Further Normalization of Data Base Relational Model // Data Base Systems. – 1972. – P. 33-64.
16. Codd E.F. Relational Completeness of Data Base Sublanguages // Data Base Systems. – 1972. – P. 65-93.
17. Codd E.F. Relational Database: A Practical Foundation for Productivity // Communications of the ACM. – 1982. – Vol. 25, № 2. – P. 109-117.
18. Codd E.F. The Relational Model for Database Management [2-nd edition]. – Pearson:Addison-Wesley, 1990. – 538 p.
19. Буй Д.Б., Кахута Н.Д. Властивості теоретико-множинних конструкцій повного образу та обмеження // Вісник Київського університету. Сер.: фіз.-мат. науки. – 2005. – Вип. 2. – С. 232-240.
20. Кахута Н.Д. Відношення сумісності, узагальнене з'єднання та узагальнений прямий добуток // Вісник Київського університету. Сер.: фіз.-мат. науки. – 2007. – Вип. 4. – С. 167-173.
21. Кахута Н.Д. Критерії ін'єктивності бінарних відношень // Вісник Київського університету. Сер.: фіз.-мат. науки. – 2007. – Вип. 3. – С. 141-146.
22. Буй Д.Б., Кахута Н.Д. Властивості відношення конфінальності та устрій множини часткових функцій // Вісник Київського університету. Сер.: фіз.-мат. науки. – 2006. – Вип. 2. – С. 125-135.
23. Elmasri R., Navathe S. Fundamentals of database systems: [4-th edition]. – Pearson: Addison-Wesley, 2004. – 1030 p.
24. Скорняков Л.А. Элементы теории структур. – Москва: Наука, 1982. – 158 с.
25. Мальцев А.И. Алгебраические системы. – Москва: Наука, 1970. – 392 с.
26. Ульман Дж. Основы систем баз данных. – М.: Фин. и стат., 1983. – 334 с.
27. Strozzi C. NoSQL: a relational database management system. – [Электронный ресурс]. – Режим доступа: http://www.strozzi.it/cgi-bin/CSA/tw7/I/en.
28. Pokorny J. NoSQL databases: a step to database scalability in web environment [Текст] // Proc. of the 13th International Conference on Information Integration and Web-Based Applications and Services. – 2011. – P. 278-283.
29. Strauch C. NoSQL databases. – [Электронный ресурс]. – Режим доступа: http://www.christof-strauch/de/nosqldbs.pdf.

Надійшла у першій редакції 09.07.2014, в останній - 17.07.2014.