УДК 004.655

Академик НАН Украины **В. Н. Редько, Д. Б. Буй, А. С. Сенченко**

Некоторые равенства в табличных алгебрах

Найдены необходимые и достаточные условия, при которых одиннадцать включений, выполняемых в табличных алгебрах, превращаются в равенства. Эти условия выражаются в терминах активных доменов таблиц и являются естественными.

Процесс информатизации общества имеет объективный характер. Ядром для подавляющего большинства современных информационных систем являются базы данных. В настоящее время наиболее распространенными остаются реляционные базы данных, математическая модель которых была впервые предложена Э. Коддом в 1970 г. [1]. С математической точки зрения реляционная база данных является конечным множеством конечных отношений различной размерности (арности) между заранее определенными множествами элементарных данных.

Табличные алгебры, введенные В. Н. Редько и Д. Б. Буем, построены на основе реляционных алгебр Э. Кодда и существенно их развивают. Они составляют теоретический фундамент языков запросов современных табличных баз данных. Элементы носителя табличной алгебры уточняют реляционные структуры данных, а сигнатурные операции построены на базе основных табличных манипуляций в реляционных алгебрах и SQL-подобных языках.

В работе [2] установлено значительное количество различных свойств операций табличных алгебр, большинство из которых для общего случая выполняются в виде включений. В настоящей работе приведены критерии перехода одиннадцати таких включений в равенства. Эти равенства представляют интерес для теории табличных алгебр по той причине, что только на основе равенств можно осуществлять эквивалентные преобразования выражений. Эти преобразования необходимы для решения актуальной задачи оптимизации запросов [3, 4].

Основные определения. Зафиксируем некоторое непустое множество атрибутов $A = \{A_1, \ldots, A_n\}$. Произвольное конечное подмножество множества A назовем схемой, причем схема может быть пустым множеством. Строкой s схемы R называется множество пар $s = \{(A'_1, d_1), \ldots, (A'_k, d_k)\}$, проекция которого по первой компоненте равна R, причем атрибуты $A'_1, \ldots, A'_k$ попарно различны, т. е. строка является функциональным бинарным отношением. Таблицей схемы R называется конечное множество строк схемы R. Далее в работе рассматриваем таблицы схемы R с количеством атрибутов k. На множестве всех таких таблиц введены такие операции:

1) объединение $\bigcup\limits_R$ двух таблиц схемы R — таблица, состоящая из тех и только тех строк, которые принадлежат хотя бы одной из исходных таблиц;

2) пересечение $\bigcap\limits_R$ двух таблиц схемы R — таблица, состоящая из тех и только тех строк, которые принадлежат одновременно обеим исходным таблицам;

3) разность $T_1 \underset{R}{-} T_2$ двух таблиц схемы R — таблица, состоящая из тех и только тех строк, которые принадлежат таблице T_1 и не принадлежат таблице T_2.

Для введения операции насыщения необходимо одно вспомогательное понятие. Активным доменом атрибута A относительно таблицы T называется множество $D_{A,T} = \{d \mid \exists s \in T \wedge (A, d) \in s\}$, состоящее, говоря содержательно, из всевозможных значений атрибута A в таблице T. Насыщением $C(T)$ называется таблица $\prod\limits_{A \in R} D_{A,T}$, где R — схема таблицы T, а $\prod$ — оператор прямого (декартового) произведения, отвечающий индексированию $A \mapsto D_{A,T}$, $A \in R$ [5]. Активным дополнением таблицы T называется таблица $\widetilde{T} = C(T) \underset{R}{-} T$.

Введем определение операции проекции. Проекцией по множеству атрибутов $X \subseteq R$ называется унарная параметрическая операция π_X, значением которой является таблица, состоящая из ограничений по X всех строк исходной таблицы: $\pi_X(T) = \{s \mid x \mid s \in T\}$. Здесь ограничение понимается стандартно: $s \mid x = s \bigcap X \times pr_2 s$, где $pr_2 s$ — проекция строки s по второй компоненте. Пусть $X = \{X_1, \ldots, X_p\}$. Далее в работе через $O = \{O_1, \ldots, O_{k-p}\}$ обозначим множество атрибутов $R - X$, не участвующих в проекции.

Для введения операции соединения (в некоторых источниках, например в [6], эта операция называется эквисоединением) необходимо одно вспомогательное понятие. Бинарные отношения ρ и τ называются совместными (обозначается $\rho \approx \tau$), если $\rho \mid x = \tau \mid x$, где $X = pr_1 \rho \bigcap pr_1 \tau$ [2]. Соединением называется бинарная операция $\otimes$, значением которой является таблица, состоящая из всевозможных объединений совместных строк исходных таблиц, т. е. $T_1 \otimes T_2 = \left\{s_1 \bigcup s_2 \mid s_1 \in T_1 \wedge s_2 \in T_2 \wedge s_1 \approx s_2\right\}$. Пусть T_1 — таблица схемы R_1, T_2 — таблица схемы R_2. Полусоединением [7] таблицы T_1 по таблице T_2 называется таблица $T_1 \triangleright T_2 = \{s_1 \in T_1 \mid \exists s_2 \in T_2 \wedge s_1 \approx s_2\}$, состоящая, говоря содержательно, из всех строк T_1, участвующих в соединении $T_1 \otimes T_2$.

Кроме этих операций на множестве всех таблиц введены операции селекции, деления таблиц и переименования атрибутов; эти операции не будут использованы в настоящей работе, поэтому их определения не приводим. Табличной алгеброй называют частичную алгебру с носителем — множеством всех таблиц произвольной схемы и приведенными выше девятью операциями (насыщение и полусоединение рассматриваются как вспомогательные операции). В табличной алгебре выделяют две особые таблицы: таблицу $T_\varepsilon = \{\varepsilon\}$, где ε — пустая строка, при этом схема таблицы T_ε является пустым множеством, и таблицу $T_\varnothing = \varnothing$ — пустое множество строк произвольной (в том числе и непустой) схемы.

Основные результаты. В работе [2] в подразделах о насыщении, активном дополнении, проекции и соединении сформулирован и доказан ряд свойств этих операции. В настоящей работе найдены необходимые и достаточные условия (в виде одиннадцати теорем), при которых включения превращаются в равенства для таблиц, не являющихся особыми; для особых таблиц эти равенства тоже выполняются, но в этом случае могут не выполняться критерии.

Теорема 1 (дистрибутивность насыщения относительно объединения). *Пусть T_1, T_2 — таблицы схемы R. Равенство $C\big(T_1 \underset{R}{\bigcup} T_2\big) = C(T_1) \underset{R}{\bigcup} C(T_2)$ выполняется тогда и только тогда, когда выполняется хотя бы одно из двух условий:*

а) существует не более одного атрибута, для которого активные домены относительно таблиц T_1 и T_2 различаются;

б) активный домен каждого атрибута $A \in R$ относительно одной таблицы является подмножеством активного домена этого же атрибута относительно другой таблицы, т. е. выполняется утверждение $\forall A (A \in R \Rightarrow (D_{A,T_1} \subseteq D_{A,T_2} \vee D_{A,T_2} \subseteq D_{A,T_1}))$.

Теорема 2 (дистрибутивность насыщения относительно пересечения). *Пусть T_1, T_2 — таблицы схемы R. При $C(T_1)\bigcap\limits_R C(T_2) \neq T_\varnothing$ равенство $C\Big(T_1\bigcap\limits_R T_2\Big) = C(T_1)\bigcap\limits_R C(T_2)$ выполняется тогда и только тогда, когда для каждого атрибута $A \in R$ выполняется равенство $D_{A,T_1\bigcap\limits_R T_2} = D_{A,T_1} \bigcap D_{A,T_2}$.*

Теорема 3 (закон двойного отрицания). *Равносильны утверждения:*

а) $C(\widetilde{T}) = C(T)$;

б) $\widetilde{\widetilde{T}} = T$;

в) для каждого значения x активного домена каждого атрибута A_q существуют такие значения $d_1, \ldots, d_{q-1}$, $d_{q+1}, \ldots, d_k$ активных доменов атрибутов $A_1, \ldots, A_{q-1}$, $A_{q+1}, \ldots$, A_k соответственно, что для строки $s = \{(A_1, d_1), \ldots, (A_{q-1}, d_{q-1}), (A_q, x), (A_{q+1}, d_{q+1}), \ldots$, $(A_k, d_k)\}$ выполняется $s \notin T$.

Теорема 4 (связь разности и активного дополнения). *Равенство $T_1\bigcap\limits_R \widetilde{T_2} = T_1 \underset{R}{-} T_2$ при $T_1 \underset{R}{-} T_2 \neq T_\varnothing$ выполняется тогда и только тогда, когда для каждого атрибута $A \in R$ выполняется включение $D_{A,T_1} \subseteq D_{A,T_2}$.*

Теорема 5 (первый закон де Моргана). *Равенство $\widetilde{T_1}\bigcap\limits_R \widetilde{T_2} = \widetilde{\Big(T_1\bigcup\limits_R T_2\Big)}$ при $\widetilde{\Big(T_1\bigcup\limits_R T_2\Big)} \neq$ $\neq T_\varnothing$ выполняется тогда и только тогда, когда выполняется одно из четырех взаимоисключающих условий:*

1) $\forall\, A(A \in R \Rightarrow D_{A,T_1} = D_{A,T_2})$;

2) $\forall\, A(A \in R \Rightarrow D_{A,T_2} \subseteq D_{A,T_1})$ и для всех индексов q, значений $x \in D_{A_q,T_1} - D_{A_q,T_2}$ и всех значений $d_1, \ldots, d_{q-1}$, $d_{q+1}, \ldots, d_k$, принадлежащих соответственно активным доменам $D_{A_1,T_1}, \ldots, D_{A_{q-1},T_1}$, $D_{A_{q+1},T_1} \ldots, D_{A_k,T_1}$, строка $\{(A_1, d_1), \ldots, (A_{q-1}, d_{q-1}), (A_q, x)$, $(A_{q+1}, d_{q+1}), \ldots, (A_k, d_k)\} \in T_1$;

3) $\forall\, A(A \in R \Rightarrow D_{A,T_1} \subseteq D_{A,T_2})$ и для всех индексов q, значений $x \in D_{A_q,T_2} - D_{A_q,T_1}$ и всех значений $d_1, \ldots, d_{q-1}$, $d_{q+1}, \ldots, d_k$, принадлежащих соответственно активным доменам $D_{A_1,T_2}, \ldots, D_{A_{q-1},T_2}$, $D_{A_{q+1},T_2} \ldots, D_{A_k,T_2}$, строка $\{(A_1, d_1), \ldots, (A_{q-1}, d_{q-1}), (A_q, x)$, $(A_{q+1}, d_{q+1}), \ldots, (A_k, d_k)\} \in T_2$;

4) существует такой атрибут A_q, для которого существуют значения $x \in D_{A_q,T_1} -$ $- D_{A_q,T_2}$, $y \in D_{A_q,T_2} - D_{A_q,T_1}$, причем $D_{A_q,T_1} \bigcap D_{A_q,T_2} \neq \varnothing$; кроме того, для всех $i \neq q$ выполняются равенства $D_{A_i,T_1} = D_{A_i,T_2}$; наконец, для всех $z_1 \in D_{A_q,T_1} - D_{A_q,T_2}$, всех $z_2 \in D_{A_q,T_2} -$ $- D_{A_q,T_1}$ и всех $d_1, \ldots, d_{q-1}$, $d_{q+1}, \ldots, d_k$, принадлежащих соответственно активным доменам $D_{A_1,T_1}, \ldots, D_{A_{q-1},T_1}$, $D_{A_{q+1},T_1} \ldots, D_{A_k,T_1}$, строка $\{(A_1, d_1), \ldots, (A_{q-1}, d_{q-1})$, (A_q, z_1), (A_{q+1}, d_{q+1}), …, $(A_k, d_k)\} \in T_1$, а строка $\{(A_1, d_1), \ldots, (A_{q-1}, d_{q-1}), (A_q, z_2), (A_{q+1}, d_{q+1}), \ldots$, $(A_k, d_k)\} \in T_2$.

Теорема 6 (второй закон де Моргана). *Равенство $\widetilde{\Big(T_1\bigcap\limits_R T_2\Big)} = \widetilde{T_1}\bigcup\limits_R \widetilde{T_2}$ при $\widetilde{\Big(T_1\bigcap\limits_R T_2\Big)} \neq$ $\neq T_\varnothing$ выполняется тогда и только тогда, когда выполняется одно из двух взаимоисключающих условий:*

1) для каждого атрибута $A \in R$ выполняются равенства $D_{A,T_1} = D_{A,T_1\bigcap\limits_R T_2}$ и $D_{A,T_2} =$ $= D_{A,T_1\bigcap\limits_R T_2}$;

2) для каждого атрибута $A_q \in R$ и каждого такого значения x, что $x \in D_{A_q,T_1} -$ $- D_{A_q,T_1\bigcap\limits_R T_2}$ и всех значений $d_1, \ldots, d_{q-1}$, $d_{q+1}, \ldots, d_k$, которые принадлежат соответст-

венно активным доменам $D_{A_1,T_1},\dots,D_{A_{q-1},T_1}$, $D_{A_{q+1},T_1},\dots,D_{A_k,T_1}$, *строка* $\{(A_1,d_1),\ \dots,$ $(A_{q-1},d_{q-1}),(A_q,x),(A_{q+1},d_{q+1}),\dots,(A_k,d_k)\}\in T_1$. *Для каждого атрибута* $A_w\in R$ *и каждого такого значения* y, *что* $y\in D_{A_w,T_2}-D_{A_w,T_1\bigcap\limits_R T_2}$ *и всех значений* $d_1,\dots,d_{w-1}$, d_{w+1}, $\dots$, d_k, *которые принадлежат соответственно активным доменам* $D_{A_1,T_2},\dots,D_{A_{w-1},T_2}$, $D_{A_{w+1},T_2},\dots,D_{A_k,T_2}$, *строка* $\{(A_1,d_1),\dots,(A_{w-1},d_{w-1}),(A_w,y),(A_{w+1},d_{w+1}),\dots,(A_k,d_k)\}\in$ $\in T_2$.

Теорема 7 (о перестановочности проекции и активного дополнения). *При* $T\neq T_\varnothing$ *равенство* $(\widetilde{\pi_X(T)})=\pi_X(\widetilde{T})$ *выполняется тогда и только тогда, когда для каждой строки* $s=\{(X_1,x_1),\dots,(X_p,x_p)\}\in\pi_X(T)$ *и любых значений* $o_1\in D_{O_1,T},\dots,o_{k-p}\in D_{O_{k-p},T}$, *строка* $s'=\{(X_1,x_1),\dots,(X_p,x_p),(O_1,o_1),\dots,(O_{k-p},o_{k-p})\}\in T$.

Теорема 8 (дистрибутивность проекции относительно пересечения). *При* $\bigcap\limits_i T_i\neq T_\varnothing$ *равенство* $\pi_X\Big(\bigcap\limits_i T_i\Big)=\bigcap\limits_i\pi_X(T_i)$ *выполняется тогда и только тогда, когда для каждой строки* $s=\{(X_1,x_1),\dots,(X_p,x_p)\}\in\bigcap\limits_i\pi_X(T)$ *существуют такие значения* $o_1\in D_{O_1,\bigcap\limits_i T_i},\dots,o_{k-p}\in D_{O_{k-p},\bigcap\limits_i T_i}$, *что строка* $s'=$ $=\{(X_1,x_1),\dots,(X_p,x_p),(O_1,o_1),\dots,(O_{k-p},o_{k-p})\}\in\bigcap\limits_i T_i$.

Теорема 9 (дистрибутивность проекции относительно разности). *При* $\pi_X(T_1\underset{R}{-}T_2)\neq$ $\neq T_\varnothing$ *равенство* $\pi_X(T_1)\underset{X}{-}\pi_X(T_2)=\pi_X(T_1\underset{R}{-}T_2)$ *выполняется тогда и только тогда, когда для каждой строки* $s=\{(X_1,x_1),\dots,(X_p,x_p)\}\in\pi_X(T_1)\bigcap\limits_X\pi_X(T_2)$ *и всех значений* $o_1\in D_{O_1,T_1\bigcup\limits_R T_2},\dots,o_{k-p}\in D_{O_{k-p},T_1\bigcup\limits_R T_2}$, *из принадлежности* $s'=\{(X_1,x_1),\dots,(X_p,x_p),$ $(O_1,o_1),b\dots,(O_{k-p},o_{k-p})\}\in T_1$ *следует принадлежность* $s'\in T_2$.

Теорема 10 (дистрибутивность насыщения относительно соединения). *Пусть* $R_1=$ $=\{A_1,\dots,A_m,G_{m+1},\dots,G_p\}$ *— схема таблицы* T_1, $R_2=\{A_1,\dots,A_m,H_{m+1},\dots,H_q\}$ *— схема таблицы* T_2, $R'=R_1\bigcap R_2$, $T_1'=T_1\underset{R_1}{-}T_1\rhd T_2$ *и* $T_2'=T_2\underset{R_2}{-}T_2\rhd T_1$. *При* $T_1\otimes T_2\neq$ $\neq T_\varnothing$ *равенство* $C(T_1\otimes T_2)=C(T_1)\otimes C(T_2)$ *выполняется тогда и только тогда, когда одновременно выполняются два условия:*

а) для всех $i\in\{m+1,\dots,p\}$ *выполняется включение* $D_{G_i,T_1'}\subseteq D_{G_i,T_1\rhd T_2}$; *для всех* $i\in\{m+1,\dots,q\}$ *выполняется включение* $D_{H_i,T_2'}\subseteq D_{H_i,T_2\rhd T_1}$;

б) для всех $i\in\{1,\dots,m\}$ *выполняются следующие две импликации:* $(x\in(D_{A_i,T_1'}-$ $-D_{A_i,T_1\rhd T_2}))\Rightarrow x\notin D_{A_i,T_2}$ *и* $(x\in(D_{A_i,T_2'}-D_{A_i,T_2\rhd T_1}))\Rightarrow x\notin D_{A_i,T_1}$.

Теорема 11 (дистрибутивность активного дополнения относительно соединения). *Равенство* $\widetilde{T_1}\otimes\widetilde{T_2}=(\widetilde{T_1\otimes T_2})$ *выполняется тогда и только тогда, когда выполняется хотя бы одно из двух условий:*

а) $T_1=T_2$;

б) $\widetilde{T_1}\otimes\widetilde{T_2}=T_\varnothing$ *и* $(\widetilde{T_1\otimes T_2})=T_\varnothing$.

Таким образом, в работе исследована взаимосвязь между операциями пересечения, объединения, разности, насыщения, активного дополнения, проекции и соединения в табличных алгебрах. Найдены необходимые и достаточные условия, при которых включения, установленные в [2], превращаются в равенства: критерии дистрибутивности насыщения относительно объединения (теорема 1), пересечения (теорема 2) и соединения (теорема 10), выполнимости аналога закона двойного отрицания (теорема 3) и законов де Моргана (теоремы 5 и 6), взаимосвязи разности и активного дополнения (теорема 4), перестановочности проекции и активного дополнения (теорема 7), дистрибутивности проекции относительно пересечения (теорема 8) и разности (теорема 9) и дистрибутивности активного дополнения относительно соединения (теорема 11). Результаты работы представляют теоретический и практический интерес. На основании равенств можно вводить аналоги определяющих соотношений, являющихся эффективным средством задания и анализа различных дискретных структур, а также осуществлять эквивалентные преобразования выражений, необходимые для их оптимизации, в том числе и для оптимизации запросов в реляционных базах данных.

1. *Codd E. F.* A Relational model of data for large shared data banks // Communications of the ACM. – 1970. – **13**, No 6. – P. 377–387.
2. *Редько В. Н., Брона Ю. И., Буй Д. Б., Поляков С. А.* Реляційні бази даних: табличні алгебри та SQL-подібні мови. – Київ: ВД “Академперіодика”, 2001. – 198 с.
3. *Knuth D. E.* The art of computer programming. Vol. 4, Fascicle 0. Introduction to combinatorial algorithms and Boolean functions. – Upper Saddle River: Addison-Wesley Professional, 2008. – 240 p.
4. *Мендкович Н. А., Кузнецов С. Д.* Обзор развития методов лексической оптимизации запросов // Тр. ИСП РАН. – 2012. – **23**. – С. 195–214.
5. *Куратовский К.* Топология. Т. 1. – Москва: Мир, 1966. – 594 с.
6. *Мейер Д.* Теория реляционных баз данных. – Москва: Мир, 1987. – 608 с.
7. *Конноли Т., Бегг К.* Базы данных. Проектирование, реализация и сопровождение. Теория и практика. – Москва: ИД “Вильямс”, 2003. – 1440 с.

Киевский национальный университет им. Тараса Шевченко

Поступило в редакцию 05.12.2013

Академік НАН України **В. Н. Редько, Д. Б. Буй, О. С. Сенченко**

Деякі рівності в табличних алгебрах

Знайдено необхідні та достатні умови, при яких одинадцять включень, які виконуються в табличних алгебрах, перетворюються у рівності. Ці умови виражаються в термінах активних доменів таблиць та є природними.

Academician of the NAS of Ukraine **V. N. Red’ko, D. B. Buy, A. S. Senchenko**

Some equalities in table algebras

The necessary and sufficient conditions, due to which eleven inclusions realized in table algebras become equalities, are found. These conditions are expressed in terms of active domains of the tables and are natural.