УДК 004.655

©2013. **А. С. Сенченко**

О ДИСТРИБУТИВНОСТИ В ТАБЛИЧНЫХ АЛГЕБРАХ ОПЕРАЦИИ НАСЫЩЕНИЯ ОТНОСИТЕЛЬНО ОПЕРАЦИЙ ОБЪЕДИНЕНИЯ И ПЕРЕСЕЧЕНИЯ

В работе найдены необходимые и достаточные условия, при которых в табличных алгебрах операция насыщения дистрибутивна относительно операций объединения и пересечения таблиц. Приведены примеры, иллюстрирующие данные критерии.

Ключевые слова: *табличная алгебра, насыщение, база данных.*

1. Введение. В настоящее время системы управления базами данных широко используются практически во всех сферах деятельности человека. Наиболее распространённой является реляционная модель данных, впервые предложенная Э. Коддом в 1970 году. С математической точки зрения реляционная база данных является конечным набором конечных отношений различной размерности между заранее определёнными множествами элементарных данных.

Табличные алгебры, введённые В.Н. Редько и Д.Б. Буем, построены на основе реляционных алгебр Э. Кодда и существенно их уточняют. 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 языке SQL.

В ставшей уже классической монографии по табличным алгебрам [1] найдено и доказано много различных свойств операций табличных алгебр. В настоящей работе найдены и доказаны необходимые и достаточные условия, при которых некоторые включения, доказанные в [1], превращаются в равенства.

2. Основные определения. Зафиксируем некоторое непустое множество $A = \{A_1, \ldots, A_n\}$, элементы которого называются атрибутами. Произвольное конечное подмножество $R = \{A'_1, \ldots,\ A'_k\} \subseteq A$ назовем схемой, причем схема может являться 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. 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 $\bigcup$ двух таблиц – таблица, состоящая из тех и только тех строк, которые принадлежат хоть одной из исходных таблиц;

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

Автор благодарит Дмитрия Борисовича Буя за постановку задачи и полезные замечания.

3) разность $T_1 - T_2$ двух таблиц – таблица, состоящая из тех и только тех строк, которые принадлежат таблице T_1 и не принадлежат таблице T_2.

Для введения операции насыщения нам необходимо дать определение понятия, используемого в работе в дальнейшем. Активным доменом атрибута A относительно таблицы T называется множество $D_{A,T} = \{d | \exists s \in T \wedge (A, d) \in s\}$, состоящее из всевозможных значений атрибута A в таблице T. Насыщением $C(T)$ называется таблица $\prod\limits_{A \in R} D_{A,T}$, где $\prod$ – оператор прямого (декартового) произведения всех атрибутов схемы T. Другими словами, мы можем понимать насыщение как аналог декартового произведения активных доменов всех атрибутов таблицы в применении к именным множествам. Активным дополнением таблицы T называется таблица $\tilde{T} = C(T) - T$.

Кроме этих операций на множестве всех таблиц введены операции проекции, селекции, соединения (в некоторых источниках, например в [2], эта операция называется эквисоединением), деления таблиц и операция 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 рассматривается как вспомогательная операция). В табличной алгебре выделяют две пустые таблицы: таблицу T_ε, схема которой является пустым множеством и таблицу $T_\emptyset$ – пустое множество строк произвольной (в том числе и непустой) схемы.

3. Основные результаты. В монографии [1] в подразделе о свойствах насыщения и активного дополнения сформулирован и доказан ряд свойств этих операций, большая часть которых являются включениями. Автором были найдены необходимые и достаточные условия (в виде двух теорем), при которых эти включения превращаются в равенства для непустых таблиц, для пустых таблиц эти равенства тоже выполняются, но в этом случае могут не выполняться критерии. Кроме формулировки и доказательства к каждой теореме будут приведены примеры, в которых будет показано, что выполнение/невыполнение критериев приводит к выполнению/невыполнению равенства.

Теорема 1 (о дистрибутивности насыщения по объединению). *Для непустых таблиц T_1 и T_2 равенство $C(T_1 \bigcup T_2) = C(T_1) \bigcup C(T_2)$ выполняется тогда и только тогда, когда выполняется хоть одно из двух условий:*

а) хотя бы для $k-1$ атрибута значения их активных доменов относительно таблиц T_1 и T_2 попарно совпадают, то есть, существует не более одного атрибута, для которого значения активного домена относительно таблиц T_1 и T_2 различается;

б) значение активного домена каждого атрибута относительно одной таблицы является подмножеством значения активного домена соответствующего атрибута относительно другой таблицы, то есть выполняются включения $\forall i\ D_{A_i,T_1} \subseteq D_{A_i,T_2}$ или $\forall i\ D_{A_i,T_2} \subseteq D_{A_i,T_1}$.

Доказательство. Необходимость. Пусть выполняется равенство $C(T_1 \bigcup T_2) = C(T_1) \bigcup C(T_2)$. Допустим, что условие (а) не выполняется, то есть значение активного домена минимум двух атрибутов попарно различны. Пусть $D_{A_q,T_1} \neq D_{A_q,T_2}$ и $D_{A_w,T_1} \neq D_{A_w,T_2}$. Покажем, что в этом случае обязательно должно выполняться условие (б), то есть должны одновременно выполняться включения $D_{A_q,T_1} \subset D_{A_q,T_2}$ и $D_{A_w,T_1} \subset D_{A_w,T_2}$ (или $D_{A_q,T_1} \supset D_{A_q,T_2}$ и $D_{A_q,T_1} \supset D_{A_q,T_2}$).

Из $D_{A_q,T_1} \neq D_{A_q,T_2}$ следует, что существует хоть один элемент x одного активного домена, который не принадлежит другому, то есть $x \in D_{A_q,T_1} - D_{A_q,T_2} \vee x \in D_{A_q,T_2} - D_{A_q,T_1}$. Пусть $x \in D_{A_q,T_1} - D_{A_q,T_2}$. Докажем, что в таком случае выполняются включения $D_{A_q,T_2} \subset D_{A_q,T_1}$ и $D_{A_w,T_2} \subset D_{A_w,T_1}$. Допустим, что не выполняется второе включение, то есть $D_{A_w,T_2} \not\subset D_{A_w,T_1}$. Следовательно, существует такой $y \in D_{A_w,T_2}$, что $y \notin D_{A_w,T_1}$. Включения $x \in D_{A_q,T_1}$ и $y \in D_{A_w,T_2}$ по определениям объединения таблиц и активного домена влекут включения $x \in D_{A_q,T_1 \cup T_2}$ и $y \in D_{A_w,T_1 \cup T_2}$, поэтому по определению активного домена существуют такие строки $s', s'' \in T_1 \bigcup T_2$, что $(A_q, x) \in s'$ и $(A_w, y) \in s''$. Тогда по определению насыщения в $C(T_1 \bigcup T_2)$ входит строка $s = \{(A_1, d_1), \ldots, (A_{q-1}, d_{q-1}), (A_q, x), (A_{q+1}, d_{q+1}), \ldots, (A_{w-1}, d_{w-1}), (A_w, y), (A_{w+1}, d_{q+1}), \ldots, (A_k, d_k)\}$, где $d_1, \ldots, d_{q-1}, d_{q+1}, \ldots, d_{w-1}, d_{w+1}, \ldots, d_k$ – некоторые элементы активных доменов соответствующих атрибутов $A_1, \ldots, A_{q-1}, A_{q+1}, \ldots, A_{w-1}, A_{w+1}, \ldots, A_k$ относительно таблицы $T_1 \bigcup T_2$. Поскольку $x \notin D_{A_q,T_2}$, то $s \notin C(T_2)$, а поскольку $y \notin D_{A_w,T_1}$, то $s \notin C(T_1)$, следовательно $s \notin C(T_1) \bigcup C(T_2)$, и значит $C(T_1 \bigcup T_2) \neq C(T_1) \bigcup C(T_2)$, что неверно по допущению. Поэтому включение $D_{A_w,T_2} \subset D_{A_w,T_1}$ выполняется. Далее, исходя из доказанности включения $D_{A_w,T_2} \subset D_{A_w,T_1}$, полностью аналогично доказывается и выполнимость включения $D_{A_q,T_2} \subset D_{A_q,T_1}$. Затем точно таким же способом можно доказать выполнимость включений $D_{A_i,T_2} \subseteq D_{A_i,T_1}$ и по всем остальным атрибутам A_i (заметим, что по условию (б) строгость этих включений не требуется).

Случай, когда $x \in D_{A_q,T_2} - D_{A_q,T_1}$, влечет выполнение включений $D_{A_q,T_1} \subset D_{A_q,T_2}$ и $D_{A_w,T_1} \subset D_{A_w,T_2}$, а следовательно, и включений $D_{A_i,T_1} \subseteq D_{A_i,T_2}$ для всех $i \neq q \wedge i \neq w$ доказывается аналогично путем замены индексов таблицы T_1 на T_2 и наоборот. Таким образом мы показали, что при выполнении равенства $C(T_1 \bigcup T_2) = C(T_1) \bigcup C(T_2)$ и не выполнении условия (а) обязательно выполняется условие (б), что доказывает необходимость теоремы.

Достаточность. Сначала докажем вспомогательное утверждение.

Лемма 1. *Пусть выполняется принадлежность $z \in D_{A,T_1 \cup T_2}$. Тогда:*
а) равенство $D_{A,T_1} = D_{A,T_2}$ влечет $z \in D_{A,T_1} \wedge z \in D_{A,T_2}$;
б) включение $D_{A,T_1} \subseteq D_{A,T_2}$ влечет $z \in D_{A,T_2}$.

Доказательство леммы 1. Пусть $z \in D_{A,T_1 \cup T_2}$ и $D_{A,T_1} = D_{A,T_2}$. Принадлежность $z \in D_{A,T_1 \cup T_2}$ возможна только в том случае, когда существует такая строка $s \in T_1 \bigcup T_2$, что $(A, z) \in s$. По определению объединения таблиц при этом выполняется хоть одна из принадлежностей $s \in T_1$ или $s \in T_2$. Тогда $s \in T_1$ влечет $z \in D_{A,T_1}$ и из равенства $D_{A,T_1} = D_{A,T_2}$ следует принадлежность $z \in D_{A,T_2}$. Тот факт, что $s \in T_2$ влечет $z \in D_{A,T_1}$ доказывается аналогично; случай (а) доказан.

Пусть теперь $z \in D_{A,T_1 \cup T_2}$ и $D_{A,T_1} \subseteq D_{A,T_2}$. Так же как и в пункте (а) леммы в этом случае существует такая строка $s \in T_1 \bigcup T_2$, что $(A, z) \in s$, следовательно выполняется $s \in T_1$ или $s \in T_2$. Тогда $s \in T_1$ влечет $z \in D_{A,T_1}$, и из включения

$D_{A,T_1} \subseteq D_{A,T_2}$ следует принадлежность $z \in D_{A,T_2}$, а $s \in T_2$ влечет принадлежность $z \in D_{A,T_2}$ по определению активного домена. □

Докажем теперь достаточность условия (а). Пусть у таблиц T_1 и T_2 значения активных доменов по $k-1$ атрибуту попарно совпадают: для всех $i \neq q$ выполняются равенства $D_{A_i,T_1} = D_{A_i,T_2}$. Докажем, что в этом случае выполняется равенство $C(T_1 \bigcup T_2) = C(T_1) \bigcup C(T_2)$.

В монографии [1] (лемма 2.2.1, пункт 7) доказано включение $C(T_1) \bigcup C(T_2) \subseteq C(T_1 \bigcup T_2)$. Для доказательства искомого равенства нужно доказать включение $C(T_1 \bigcup T_2) \subseteq C(T_1) \bigcup C(T_2)$. От противного, допустим, что включение не выполняется, то есть существует строка $s = \{(A_1, d_1), \ldots, (A_{q-1}, d_{q-1}), (A_q, x), (A_{q+1}, d_{q+1}), \ldots, (A_k, d_k)\}$, которая принадлежит $C(T_1 \bigcup T_2)$ и не принадлежит $C(T_1) \bigcup C(T_2)$. По определениям активного домена и насыщения выполняются принадлежности $d_1 \in D_{A_1,T_1\cup T_2}, \ldots, d_{q-1} \in D_{A_{q-1},T_1\cup T_2}, x \in D_{A_q,T_1\cup T_2}, d_{q+1} \in D_{A_{q+1},T_1\cup T_2}, \ldots, d_k \in D_{A_k,T_1\cup T_2}$. С учётом условий $D_{A_i,T_1} = D_{A_i,T_2}$ для всех $i \neq q$ и пункта (а) леммы 1 получаем, что $d_1 \in D_{A_1,T_1}$ и $d_1 \in D_{A_1,T_2}, \ldots, d_{q-1} \in D_{A_{q-1},T_1}$ и $d_{q-1} \in D_{A_{q-1},T_2}, d_{q+1} \in D_{A_{q+1},T_1}$ и $d_{q+1} \in D_{A_{q+1},T_2}, \ldots, d_k \in D_{A_k,T_1}$ и $d_k \in D_{A_k,T_2}$. Факт принадлежности $x \in D_{A_q,T_1\cup T_2}$ влечет существование в таблице $T_1 \bigcup T_2$ такой строки s' , что $(A_q, x) \in s'$. По определению объединения таблиц в этом случае выполняется хоть одна из принадлежностей $s' \in T_1$ или $s' \in T_2$, то есть выполняется хотя бы одно из условий: $x \in D_{A_q,T_1}$ или $x \in D_{A_q,T_2}$. Если выполняется $x \in D_{A_q,T_1}$, то $s' \in C(T_1)$, а если выполняется $x \in D_{A_q,T_2}$, то $s' \in C(T_1)$. В любом случае выполняется $s' \in C(T_1) \bigcup C(T_2)$, что противоречит допущению. Достаточность условия (а) доказана.

Докажем теперь достаточность условия (б). Пусть выполняются включения $\forall i \; D_{A_i,T_1} \subseteq D_{A_i,T_2}$. От противного, допустим, что $C(T_1 \bigcup T_2) \neq C(T_1) \bigcup C(T_2)$. Как уже было показано в доказательстве достаточности условия (а), это неравенство возможно только в том случае, когда существует такая строка $s = \{(A_1, d_1), \ldots, (A_k, d_k)\}$, что $s \in C(T_1 \bigcup T_2)$ и $s \notin C(T_1) \bigcup C(T_2)$. Тот факт, что $s \in C(T_1 \bigcup T_2)$ влечет выполнение принадлежностей $d_i \in D_{A_i,T_1\cup T_2}$ для всех i . Из условия $\forall i \; D_{A_i,T_1} \subseteq D_{A_i,T_2}$ ввиду пункта (б) леммы 1 следует, что $\forall i \; d_i \in D_{A_i,T_2}$, поэтому по определению насыщения $s \in C(T_2)$, и значит, $s \in C(T_1) \bigcup C(T_2)$. Таким образом, допущение $s \notin C(T_1) \bigcup C(T_2)$ неверно; достаточность условия (б) доказана. □

Проиллюстрируем критерии теоремы 1 на следующих примерах

Пример 1.1.

$$\text{Пусть } T_1 = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 2 & 3 \\ 2 & 2 & 3 \\ 2 & 3 & 3 \end{array} \text{ и } T_2 = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 2 & 2 & 1 \\ 2 & 3 & 1 \\ 3 & 3 & 1 \end{array}.$$

Значения активных доменов $D_{A,T_1} = \{1, 2\}, D_{B,T_1} = \{2, 3\}, D_{C,T_1} = \{3\}$ и $D_{A,T_2} = \{2, 3\}, D_{B,T_1} = \{2, 3\}, D_{C,T_1} = \{1\}$ не удовлетворяют ни одному из условий, равенство не должно выполняться. Действительно,

$$C(T_1 \bigcup T_2) = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 2 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 1 \\ 1 & 3 & 3 \\ 2 & 2 & 1 \\ 2 & 2 & 3 \\ 2 & 3 & 1 \\ 2 & 3 & 3 \\ 3 & 2 & 1 \\ 3 & 2 & 3 \\ 3 & 3 & 1 \\ 3 & 3 & 3 \end{array}, \; C(T_1) = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 2 & 3 \\ 1 & 3 & 3 \\ 2 & 2 & 3 \\ 2 & 3 & 3 \end{array}, \; C(T_2) = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 2 & 2 & 1 \\ 2 & 3 & 1 \\ 3 & 2 & 1 \\ 3 & 3 & 1 \end{array}, \text{ то есть}$$

$C(T_1 \bigcup T_2) \neq C(T_1) \bigcup C(T_2)$.

Пример 1.2.

$$\text{Пусть } T_1 = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 2 & 3 \\ 2 & 2 & 3 \\ 2 & 3 & 3 \end{array} \text{ и } T_2 = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 2 & 4 \\ 2 & 2 & 4 \\ 2 & 3 & 4 \end{array}.$$

Значения активных доменов $D_{A,T_1} = \{1,2\}, D_{B,T_1} = \{2,3\}, D_{C,T_1} = \{3\}$ и $D_{A,T_2} = \{1,2\}, D_{B,T_1} = \{2,3\}, D_{C,T_1} = \{4\}$ удовлетворяют условию (а), равенство должно выполняться. Действительно,

$$C(T_1 \bigcup T_2) = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 2 & 3 \\ 1 & 2 & 4 \\ 1 & 3 & 3 \\ 1 & 3 & 4 \\ 2 & 2 & 3 \\ 2 & 2 & 4 \\ 2 & 3 & 3 \\ 2 & 3 & 4 \end{array}, \; C(T_1) = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 2 & 3 \\ 1 & 3 & 3 \\ 2 & 2 & 3 \\ 2 & 3 & 3 \end{array}, \; C(T_2) = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 2 & 4 \\ 1 & 3 & 4 \\ 2 & 2 & 4 \\ 2 & 3 & 4 \end{array}, \text{ то есть}$$

$C(T_1 \bigcup T_2) = C(T_1) \bigcup C(T_2)$.

Пример 1.3.

$$\text{Пусть } T_1 = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 2 & 3 \\ 2 & 2 & 1 \\ 2 & 3 & 3 \end{array} \text{ и } T_2 = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 2 & 2 & 1 \\ 2 & 3 & 1 \end{array}.$$

Значения активных доменов $D_{A,T_1} = \{1,2\}, D_{B,T_1} = \{2,3\}, D_{C,T_1} = \{1,3\}$ и $D_{A,T_2} = \{2\}, D_{B,T_1} = \{2,3\}, D_{C,T_1} = \{1\}$ удовлетворяют условию (б), равенство должно выполняться. Действительно,

$$C(T_1 \bigcup T_2) = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 2 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 1 \\ 1 & 3 & 3 \\ 2 & 2 & 1 \\ 2 & 2 & 3 \\ 2 & 3 & 1 \\ 2 & 3 & 3 \end{array}, \ C(T_1) = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 2 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 1 \\ 1 & 3 & 3 \\ 2 & 2 & 1 \\ 2 & 2 & 3 \\ 2 & 3 & 1 \\ 2 & 3 & 3 \end{array}, \ C(T_2) = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 2 & 2 & 1 \\ 2 & 3 & 1 \end{array}, \text{ то есть}$$

$C(T_1 \bigcup T_2) = C(T_1) \bigcup C(T_2)$.

Теорема 2 (о дистрибутивности насыщения по пересечению). *При $C(T_1) \bigcap C(T_2) \neq T_\emptyset$ эквивалентны утверждения:*

1) выполняется равенство $C(T_1 \bigcap T_2) = C(T_1) \bigcap C(T_2)$;

2) выполняются равенства $\forall i \ D_{A_i, T_1 \bigcap T_2} = D_{A_i,T_1} \bigcap D_{A_i,T_2}$;

3) для любого индекса i и каждого элемента x из множества $D_{A_i,T_1} \bigcap D_{A_i,T_2}$ существует такая строка $s \in T_1 \bigcap T_2$, что $(A_i, x) \in s$.

Доказательство. Покажем сначала эквивалентность утверждений (1) и (2). Пусть выполняется равенство (1). От противного, допустим, что не выполняется равенство (2), то есть $D_{A_q, T_1 \bigcap T_2} \neq D_{A_q,T_1} \bigcap D_{A_q,T_2}$ для некоторого индекса q. При этом возможны два случая:

а) $\exists x \in_{A_q, T_1 \bigcap T_2}$ и $x \notin D_{A_q,T_1} \bigcap D_{A_q,T_2}$. Тогда по определению активного домена, существует такая строка $s \in T_1 \bigcap T_2$, что $(A_q, x) \in s$. Включение $s \in T_1 \bigcap T_2$ влечёт $s \in T_1$ и $s \in T_2$, следовательно $x \in D_{A_q,T_1}$ и $x \in D_{A_q,T_2}$, то есть $x \in D_{A_q,T_1} \bigcap D_{A_q,T_2}$. Получившееся противоречие доказывает невозможность этого случая.

б) $\exists y \in D_{A_q,T_1} \bigcap D_{A_q,T_2}$ и $y \notin D_{A_q, T_1 \bigcap T_2}$. Тогда $y \in D_{A_q,T_1}$ и $y \in D_{A_q,T_2}$. По условию таблица $C(T_1) \bigcap C(T_2)$ непустая, следовательно, в ней существует хоть одна строка. Пусть $s = \{(A_1, d_1), \ldots, (A_k, d_k)\} \in C(T_1) \bigcap C(T_2)$. Тогда по определению активного домена $d_1 \in D_{A_1,T_1}$ и $d_1 \in D_{A_1,T_2}, \ldots, d_k \in D_{A_k,T_1}$ и $d_k \in D_{A_k,T_2}$. По определению насыщения получаем, что $s' = \{(A_1, d_1), \ldots, (A_{q-1}, d_{q-1}), (A_q, y), (A_{q+1}, d_{q+1}), \ldots, (A_k, d_k)\} \in C(T_1)$ и $s' \in C(T_2)$, то есть $s' \in C(T_1) \bigcap C(T_2)$. Из равенства (1) следует, что $s' \in C(T_1 \bigcap T_2)$, поэтому $y \in D_{A_q, T_1 \bigcap T_2}$, что доказывает невозможность этого случая; импликация $\big(1\big) \Rightarrow \big(2\big)$ доказана.

Пусть теперь выполняется равенство (2). От противного, допустим, что не выполняется равенство (1). В монографии [1] доказано включение $C(T_1 \bigcap T_2) \subseteq C(T_1) \bigcap C(T_2)$, поэтому равенство (1) может не выполняться только в случае, когда существует такая строка $s = \{(A_1, d_1), \ldots, (A_k, d_k)\} \in C(T_1) \bigcap C(T_2)$, что $s \notin C(T_1 \bigcap T_2)$. Рассмотрим этот случай. Из $s \in C(T_1) \bigcap C(T_2)$ следует, что $s \in C(T_1)$ и $s \in C(T_2)$. Из равенства (2) и определения активного домена получаем, что $d_1 \in D_{A_1, T_1 \bigcap T_2}, \ldots, d_k \in D_{A_k, T_1 \bigcap T_2}$. Тогда по определению насыщения выполняется $\{(A_1, d_1), \ldots, (A_k, d_k)\} \in C(T_1 \bigcap T_2)$, что противоречит допущению, импликация $\big(2\big) \Rightarrow \big(1\big)$ доказана, поэтому утверждения (1) и (2) равносильны.

Докажем теперь, что эквивалентны утверждения (1) и (3). Пусть выполняется равенство (1). От противного, допустим, что не выполняется утверждение (3), то есть $\exists x \in D_{A_q,T_1} \bigcap D_{A_q,T_2} | \ \forall s \in T_1 \bigcap T_2 \ (A_q, x) \notin s$ для некоторого индекса q. По доказанному выше равенство (1) эквивалентно равенству (2). Из равенства (2) следует, что $x \in D_{A_q, T_1 \bigcap T_2}$, значит по определению активного домена для некоторой строки $s' \in T_1 \bigcap T_2$ выполняется принадлежность $(A_q, x) \in s'$, что противоречит допущению. Импликация $(1) \Rightarrow (3)$ доказана.

Пусть теперь выполняется утверждение (3). От противного, допустим, что не выполняется равенство (1), что может быть только в том случае, когда $\exists s = \{(A_1, d_1), \ldots, (A_k, d_k)\} \in C(T_1) \bigcap C(T_2)$ и $s \notin C(T_1 \bigcap T_2)$. Рассмотрим этот случай. $s \notin C(T_1 \bigcap T_2)$ влечет $d_q \notin D_{A_q, T_1 \bigcap T_2}$ для некоторого индекса q. Поскольку $s \in C(T_1) \bigcap C(T_2)$, и, значит $s \in C(T_1)$ и $s \in C(T_2)$, то по определениям активного домена и насыщения выполняются включения $d_q \in D_{A_q,T_1}$ и $d_q \in D_{A_q,T_2}$, поэтому $d_q \in D_{A_q,T_1} \bigcap D_{A_q,T_2}$. По условию (3) в этом случае должна существовать такая строка $s' \in T_1 \bigcap T_2$, что $(A_q, d_q) \in s'$. По определению активного домена $d_q \in D_{A_q, T_1 \bigcap T_2}$, противоречие доказывает неверность допущения, импликация $(3) \Rightarrow (1)$ доказана. $\square$

Проиллюстрируем критерий теоремы 2 на примерах.

Пример 2.1.

$$\text{Пусть } T_1 = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 2 & 1 \\ 1 & 3 & 3 \\ 2 & 3 & 1 \\ 4 & 2 & 1 \\ 4 & 2 & 3 \end{array} \text{ и } T_2 = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 1 & 1 \\ 1 & 2 & 1 \\ 3 & 1 & 3 \\ 3 & 2 & 4 \\ 4 & 2 & 3 \end{array}.$$

Значения активных доменов $D_{A,T_1} = \{1,2,4\}, D_{B,T_1} = \{2,3\}, D_{C,T_1} = \{1,3\}$ и $D_{A,T_2} = \{1,3,4\}, D_{B,T_1} = \{1,2\}, D_{C,T_1} = \{1,3,4\}$; общие значения активных доменов $(A,1), (A,4), (B,2), (C,1), (C,3)$. Общие строки T_1 и T_2 $\{(A,1), (B,2), (C,1)\}$ и $\{(A,4), (B,2), (C,3)\}$ покрывают все общие значения доменов, выполняется утверждение (3), равенство (1) должно выполняться. Действительно,

$$C(T_1 \bigcap T_2) = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 2 & 1 \\ 1 & 2 & 3 \\ 4 & 2 & 1 \\ 4 & 2 & 3 \end{array}, \ C(T_1) \bigcap C(T_2) = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 2 & 1 \\ 1 & 2 & 3 \\ 4 & 2 & 1 \\ 4 & 2 & 3 \end{array}, \text{ то есть } C(T_1 \bigcap T_2) = C(T_1) \bigcap C(T_2).$$

Пример 2.2.

$$\text{Пусть } T_1 = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 1 & 3 \\ 1 & 2 & 2 \\ 1 & 4 & 2 \\ 4 & 1 & 3 \\ 4 & 4 & 2 \end{array} \text{ и } T_2 = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 1 & 3 \\ 1 & 2 & 1 \\ 2 & 1 & 3 \\ 2 & 2 & 1 \\ 4 & 1 & 3 \end{array}.$$

Значения активных доменов $D_{A,T_1} = \{1,4\}, D_{B,T_1} = \{1,2,4\}, D_{C,T_1} = \{2,3\}$ и $D_{A,T_2} = \{1,2,4\}, D_{B,T_1} = \{1,2\}, D_{C,T_1} = \{1,3\}$; общие значения активных доменов $(A,1), (A,4), (B,1), (B,2), (C,3)$. Общие строки $\{(A,1), (B,1), (C,3)\}$ и $\{(A,4),$

$(B,1),(C,3)\}$ не покрывают все общие значения доменов (не покрыто значение $(B,2)$), утверждение (3) не выполняется, равенство (1) не должно выполняться. Действительно,

$$C(T_1 \bigcap T_2) = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 1 & 3 \\ 4 & 1 & 3 \end{array}, \quad C(T_1) \bigcap C(T_2) = \begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ 1 & 1 & 3 \\ 1 & 2 & 3 \\ 4 & 1 & 3 \\ 4 & 2 & 3 \end{array}, \text{ то есть } C(T_1 \bigcap T_2) \neq$$

$C(T_1) \bigcap C(T_2)$.

4. Выводы. В работе исследованы свойства операций насыщения и активного дополнения табличных алгебр. Найдены критерии, при которых некоторые включения из [1] превращаются в равенства. Результаты работы могут быть использованы в теории обобщенных табличных алгебр и, на наш взгляд, для оптимизации запросов в реляционных базах данных.

1. 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.
2. *Мейер Д.* Теория реляционных баз данных: [пер. с англ.] / Д. Мейер. – Москва: Мир, 1987. – 608 с.

A. S. Senchenko

On a distributivity of a saturation to join and intersection in table algebra.

In this paper there is found the necessary and sufficient conditions due to which a saturation operation is distributive to operations join and intersection. This conditions are illustrated in examples.

Keywords: *table algebra, saturation, database.*

О. С. Сенченко

Про дистрибутивність в табличних алгебрах операції насичення відносно операцій об'єднання та перетину.

В роботі знайдено необхідні та достатні умови, за яких у табличних алгебрах операція насичення є дистрибутивною відносно операцій об'єднання та перетину таблиць. Наведені приклади, що ілюструють знайдені критерії.

Ключові слова: *таблична алгебра, насичення, бази даних.*

Донбасский государственный педагогический ун-т
`senchenko@pisem.net`

Получено 05.04.13