УДК 004.655
DOI: 10.37069/1683-4720-2018-32-13

©2018. **А.С. Сенченко**

ВЗАИМОСВЯЗИ МЕЖДУ ПЕРЕСЕЧЕНИЕМ, ОБЪЕДИНЕНИЕМ И ДРУГИМИ СИГНАТУРНЫМИ ОПЕРАЦИЯМИ В ТАБЛИЧНЫХ АЛГЕБРАХ

В настоящее время реляционные базы данных, математическая модель которых была предложена Э. Коддом, остаются наиболее распространенными. Табличные алгебры построены на основе реляционных алгебр Э. Кодда, существенно их развивают и составляют теоретический фундамент языков запросов современных табличных баз данных. В реляционных и табличных алгебрах актуальной является задача эквивалентного преобразования выражений с целью их минимизации или приведения к стандартному виду; она является одним из этапов оптимизации запросов, ее решение может значительно уменьшить время обработки информации в реляционных системах управления базами данных. Для решения этой задачи используются взаимосвязи между табличными операциями. На сегодня установлено значительное количество таких взаимосвязей, большинство из которых для общего случая выполняются в виде включений. Автором были найдены критерии перехода некоторых таких включений в равенства. В настоящей работе рассмотрены взаимосвязи пересечения и объединения таблиц с другими сигнатурными операциями табличных алгебр: разностью, селекцией, проекцией, насыщением, активным дополнением, соединением, переименованием атрибутов.

MSC: 68P15.

Ключевые слова: *пересечение, объединение, базы данных, табличные алгебры.*

1. Введение.

В настоящее время системы управления базами данных широко используются во многих сферах деятельности человека. Наиболее распространенными остаются реляционные (табличные) базы данных, основанные на реляционных моделях Э. Кодда [1]. Реляционные алгебры построены на базе теории множеств и отношений и являются основанием логики работы реляционных баз данных. Табличные алгебры, введенные В.Н. Редько и Д.Б. Буем [2], построены на основе реляционных алгебр Кодда, существенно их развивают и дополняют. 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.

Одной из актуальных задач в реляционных и табличных алгебрах является задача эквивалентного преобразования выражений с целью их упрощения / минимизации или приведения к стандартному виду; она является одним из этапов оптимизации запросов [3], ее решение может значительно уменьшить время обработки информации в реляционных системах управления базами данных. Для решения этой задачи используются взаимосвязи между основными табличными

операциями. Можно выделить пять видов взаимосвязей между табличными операциями: а) выражение одной операции через другие;
б) перестановочность (для двух унарных операций);
в) дистрибутивность (для двух бинарных или для бинарной и унарной операций);
г) аналоги известных свойств из теории множеств (для пересечения, объединения, разности таблиц).
д) коммутативность, ассоциативность, идемпотентность (для бинарных операций).

На сегодня [4] установлено значительное количество таких взаимосвязей, большинство из которых для общего случая выполняются в виде включений. Автором были найдены критерии перехода некоторых таких включений в равенства. Эти критерии выражаются в терминах активных доменов таблиц и являются естественными. В настоящей работе рассмотрены взаимосвязи пересечения и объединения таблиц с сигнатурными операциями табличных алгебр: разностью, селекцией, проекцией, насыщением, активным дополнением, соединением, переименованием атрибутов.

2. Основные определения.

Далее рассмотрим основные определения из теории табличных алгебр, а также укажем некоторые отличия между табличными и реляционными алгебрами Кодда.

Зафиксируем некоторое непустое множество $A = \{A_1, \dots, A_n\}$, элементы которого называются атрибутами. Произвольное конечное подмножество $R = \{A'_1, \dots, A'_k\} \subseteq A$ назовем схемой, причем схема может являться пустым множеством. Строкой (кортежом в реляционных алгебрах Кодда) s схемы R ($s(R)$) называется множество пар $s = \{(A'_1, d_1), \dots, (A'_k, d_k)\}$, проекция которого по первой компоненте равна R. Таблицей (отношением в реляционных алгебрах Кодда) схемы R ($T(R)$) называется конечное множество строк схемы R. Выделяют две особые таблицы: таблицу $T_\varepsilon = \{\varepsilon\}$, где ε – пустая строка, при этом схема таблицы T_ε является пустым множеством, и таблицу $T_\emptyset = \emptyset$ – пустое множество строк произвольной (в том числе и непустой) схемы.

Так же, как и в теории множеств, вводится понятие подтаблицы: таблица $T_1(R)$ является подтаблицей таблицы $T_2(R)$ (обозначается $T_1 \subseteq T_2$), если выполняется импликация: $\forall s(R) s \in T_1 \Rightarrow s \in T_2$. Кроме того, для любой таблицы $T(R)$ выполняется $T_\emptyset \subseteq T(R)$; если одновременно выполняются $T_1 \subseteq T_2$ и $T_2 \subseteq T_1$, то $T_1 = T_2$. Для удобства и уменьшения размера выражений, содержащих табличные операции, в основном мы не будем указывать явно схемы таблиц.

Далее в работе рассматриваем таблицы схемы R. На множестве всех таких таблиц введены такие параметрические операции:
1) пересечение $\bigcap\limits_R$ двух таблиц схемы R – таблица, состоящая из тех и только тех строк, которые принадлежат одновременно обеим исходным таблицам;
2) объединение $\bigcup\limits_R$ двух таблиц схемы R – таблица, состоящая из тех и только тех строк, которые принадлежат хотя бы одной из исходных таблиц;
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 $D_{A,T} = \{d | \exists s \in T \wedge (A, d) \in s\}$, 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 $A \mapsto D_{A,T}$, $A \in R$ [4]. Таблица T называется насыщенной, если $T = C(T)$. Активным дополнением таблицы T называется таблица $\tilde{T} = C(T) \underset{R}{-} T$.

Введем определение операции селекции. Селекцией по предикату $P : S \to \{true, false\}$, где S – множество всех строк, называется унарная параметрическая операция σ_P , которая таблице сопоставляет ее подтаблицу, содержащую строки, на которых предикат P принимает истинное значения.

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: $\pi_X(T) = \{s \mid x \ \ | \ s \in T\}$. Здесь ограничение понимается стандартно: $s \mid x = s \bigcap X \times pr_2 s$, где $pr_2 s$ – проекция строки s по второй компоненте.

Для введения операции соединения необходимо одно вспомогательное понятие. Бинарные отношения ρ и τ называются совместными (обозначается $\rho \approx \tau$), если $\rho \mid x = \tau \mid x$, где $X = pr_1 \rho \bigcap pr_1 \tau$ [4]. Соединением называется бинарная операция $\otimes$, значением которой является таблица, состоящая из всевозможных объединений совместных строк исходных таблиц, т.е. $T_1 \otimes T_2 = \{s_1 \bigcup s_2 \mid s_1 \in T_1 \wedge s_2 \in T_2 \wedge s_1 \approx s_2\}$. Пусть T_1 – таблица схемы R_1, T_2 – таблица схемы R_2.

Введем определение операции переименования атрибутов. Переименованием называется унарная, в общем случае частичная операция RT_ξ , где ξ – инъективное отображение на множестве атрибутов. Эта операция осуществляет только переименование атрибутов таблиц в соответствии с отображением-параметром ξ. Содержательно говоря, переименование таблицы сводится к переименованию первых компонент пар – элементов строк.

Табличной алгеброй называют частичную алгебру с носителем — множеством всех таблиц произвольной схемы и приведёнными выше операциями (насыщение – вспомогательная операция), а также операцией деления, которая в этой статье не используется. Заметим, что в реляционных алгебрах Кодда [1] в качестве сигнатурных выделяют восемь операций: пересечение, объединение, разность, проекцию, селекцию, соединение, деление, а также, декартово произведение (в табличных алгебрах эта операция не является сигнатурной) таблиц; операции активного дополнения, переименования, а также, агрегации и многочисленные модификации операции соединения рассматриваются в различных пополнениях реляционных алгебр Кодда [3].

Рассмотрим взаимосвязи между пересечением, объединением и селекцией. В

[4] доказано выполненit равенств $\sigma_P(\bigcap_i T_i) = \bigcap_i \sigma_P(T_i)$ и $\sigma_P(\bigcup_i T_i) = \bigcup_i \sigma_P(T_i)$, то есть селекция дистрибутивна относительно пересечения и объединения таблиц.

Рассмотрим взаимосвязи между пересечением, объединением и соединением. В [4] доказано, что при совпадении схем таблиц T_i (только в этом случае определены $\bigcap_i T_i$ и $\bigcup_i T_i$) выполненяются равенства $T \otimes (\bigcap_i T_i) = \bigcap_i (T \otimes T_i)$ и $T \otimes (\bigcup_i T_i) = \bigcup_i (T \otimes T_i)$, то есть соединение дистрибутивно относительно пересечения и объединения таблиц.

Рассмотрим взаимосвязи между пересечением, объединением и переименовании атрибутов. В [4] доказано, что при корректном задании переименования атрибутов выполняются равенства $RT_\xi(\bigcap_i T_i) = \bigcap_i RT_\xi(T_i)$ и $RT_\xi(\bigcup_i T_i) = \bigcup_i RT_\xi(T_i)$, то есть переименование дистрибутивно относительно пересечения и объединения таблиц.

Взаимосвязи между пересечением, объединением и другими сигнатурными операциями табличных алгебр будут рассмотрены ниже.

3. Взаимосвязи между пересечением и табличными операциями.

Далее, если обратное не оговорено явно, для компактности изложения полагаем, что все упомянутые таблицы имеют непустую схему R с количеством атрибутов $k > 0$.

Известно, что пересечение может быть выражено через разность таблиц: $T_1 \bigcap_R T_2 = T_1 \underset{R}{-} (T_1 \underset{R}{-} T_2)$. Доказательство тривиально.

Операция пересечения коммутативна, ассоциативна и идемпотентна, то есть выполняются равенства $T_1 \bigcap_R T_2 = T_2 \bigcap_R T_1$, $T_1 \bigcap_R (T_2 \bigcap_R T_3) = (T_1 \bigcap_R T_2) \bigcap_R T_3$ и $T \bigcap_R T = T$. Доказательство тривиально.

Операции пересечения и объединения дистрибутивны между собой (двумя вариантами) то есть выполняются равенства $T_1 \bigcup_R (T_2 \bigcap_R T_3) = (T_1 \bigcup_R T_2) \bigcap_R (T_1 \bigcup_R T_3)$ и $T_1 \bigcap_R (T_2 \bigcup_R T_3) = (T_1 \bigcap_R T_2) \bigcup_R (T_1 \bigcap_R T_3)$. Доказательство тривиально.

Рассмотрим взаимосвязи между пересечением и разностью таблиц. Справедливы такие утверждения.

Лемма 1. *Выполняется равенство* $T_1 \bigcap_R (T_2 \underset{R}{-} T_3) = (T_1 \bigcap_R T_2) \underset{R}{-} (T_1 \bigcap_R T_3)$.

Доказательство. Пусть $T_1 \bigcap_R (T_2 \underset{R}{-} T_3) \neq T_\emptyset$ и $s \in T_1 \bigcap_R (T_2 \underset{R}{-} T_3)$. Тогда, по определению пересечения и разности, $s \in T_1$, $s \in T_2$ и $s \notin T_3$. Из $s \in T_1$ и $s \in T_2$ следует $s \in T_1 \bigcap_R T_2$, а из $s \in T_1$ и $s \notin T_3$ следует $s \notin T_1 \bigcap_R T_3$, поэтому $s \in (T_1 \bigcap_R T_2) \underset{R}{-} (T_1 \bigcap_R T_3)$. Таким образом, $T_1 \bigcap_R (T_2 \underset{R}{-} T_3) \subseteq (T_1 \bigcap_R T_2) \underset{R}{-} (T_1 \bigcap_R T_3)$.

Пусть теперь $(T_1 \bigcap_R T_2) \underset{R}{-} (T_1 \bigcap_R T_3) \neq T_\emptyset$ и $s \in (T_1 \bigcap_R T_2) \underset{R}{-} (T_1 \bigcap_R T_3)$. Тогда, по определению разности, $s \in T_1 \bigcap_R T_2$ и $s \notin T_1 \bigcap_R T_3$. Из $s \in T_1 \bigcap_R T_2$ следует $s \in T_1$ и $s \in T_2$, а из $s \in T_1$ и $s \notin T_1 \bigcap_R T_3$ следует $s \notin T_3$, поэтому $s \in T_2 \underset{R}{-} T_3$, поэтому

$s \in T_1 \bigcap_R (T_2 \underset{R}{-} T_3)$. Таким образом, $(T_1 \bigcap_R T_2) \underset{R}{-} (T_1 \bigcap_R T_3) \subseteq T_1 \bigcap_R (T_2 \underset{R}{-} T_3)$. $\square$

Лемма 2. *Выполняется включение* $T_1 \underset{R}{-} (T_2 \bigcap_R T_3) \supseteq (T_1 \underset{R}{-} T_2) \bigcap_R (T_1 \underset{R}{-} T_3)$.

Доказательство. Пусть $(T_1 \underset{R}{-} T_2) \bigcap_R (T_1 \underset{R}{-} T_3) \neq T_\emptyset$ и $s \in (T_1 \underset{R}{-} T_2) \bigcap_R (T_1 \underset{R}{-} T_3)$. Тогда, по определению пересечения, $s \in T_1 \underset{R}{-} T_2$ и $s \in T_1 \underset{R}{-} T_3$, поэтому $s \in T_1$, $s \notin T_2$ и $s \notin T_3$. Из $s \notin T_2$ и $s \notin T_3$ следует $s \notin T_2 \bigcap_R T_3$. По определению разности таблиц $s \in T_1 \underset{R}{-} (T_2 \bigcap_R T_3)$. $\square$

Найдем критерий, при котором данное включение превращается в равенство.

Лемма 3. *Равенство* $T_1 \underset{R}{-} (T_2 \bigcap_R T_3) = (T_1 \underset{R}{-} T_2) \bigcap_R (T_1 \underset{R}{-} T_3)$ *выполняется тогда и только тогда, когда* $T_1 \bigcap_R T_2 = T_1 \bigcap_R T_3$.

Доказательство. Необходимость. Пусть $T_1 = T_\emptyset$. Тогда $T_1 \bigcap_R T_2 = T_\emptyset$, $T_1 \bigcap_R T_3 = T_\emptyset$, $T_1 \underset{R}{-} (T_2 \bigcap_R T_3) = T_\emptyset$, то есть равенство выполняется.

Пусть теперь $T_1 \neq T_\emptyset$. Если в этом случае $T_1 \underset{R}{-} (T_2 \bigcap_R T_3) = T_\emptyset$, то по определению разности для любой строки $s \in T_1$ выполняется $s \in (T_2 \bigcap_R T_3)$, поэтому $s \in T_2$ и $s \in T_3$, следовательно $T_1 \underset{R}{-} T_2 = T_\emptyset$ и $T_1 \underset{R}{-} T_3 = T_\emptyset$, то есть равенство выполняется.

Пусть теперь $T_1 \underset{R}{-} (T_2 \bigcap_R T_3) \neq T_\emptyset$ и $s \in T_1 \underset{R}{-} (T_2 \bigcap_R T_3)$. Тогда, по определению разности, $s \in T_1$ и $s \notin T_2 \bigcap_R T_3$. Из $T_1 \bigcap_R T_2 = T_1 \bigcap_R T_3$ и $s \notin T_2 \bigcap_R T_3$ следует, что $s \notin T_2$ и $s \notin T_3$, поэтому $s \in T_1 \underset{R}{-} T_2$ и $s \in T_1 \underset{R}{-} T_3$, откуда следует $s \in (T_1 \underset{R}{-} T_2) \bigcap_R (T_1 \underset{R}{-} T_3)$, что доказывает необходимость.

От противного докажем достаточность утверждения. Пусть $T_1 \bigcap_R T_2 \neq T_1 \bigcap_R T_3$. Тогда $T_1 \neq T_\emptyset$. Для определенности положим, что существует такая строка s, что $s \in T_1 \bigcap_R T_2$ и $s \notin T_1 \bigcap_R T_3$. Тогда $s \in T_1$, $s \in T_2$ и $s \notin T_3$, следовательно, $s \notin T_2 \bigcap_R T_3$, то есть $s \in T_1 \underset{R}{-} (T_2 \bigcap_R T_3)$. Из $s \in T_1$ и $s \in T_2$ вытекает $s \notin (T_1 \underset{R}{-} T_2)$, поэтому $s \notin (T_1 \underset{R}{-} T_2) \bigcap_R (T_1 \underset{R}{-} T_3)$, то есть равенство не выполняется. Аналогично доказывается невыполнимость равенства при существовании такой строки s, что $s \notin T_1 \bigcap_R T_2$ и $s \in T_1 \bigcap_R T_3$; это доказывает достаточность утверждения. $\square$

Рассмотрим взаимосвязи между пересечением и насыщением таблиц. В [4] доказано включение $C(T_1 \bigcap_R T_2) \subseteq C(T_1) \bigcap_R C(T_2)$. В [5] найдены и доказаны критерии перехода этого включения в равенство для случая $C(T_1) \bigcap_R C(T_2) \neq T_\emptyset$. Из $C(T_1 \bigcap_R T_2) \subseteq C(T_1) \bigcap_R C(T_2)$, непустоты схемы таблиц и $C(T_1) \bigcap_R C(T_2) = T_\emptyset$ следует, что $C(T_1 \bigcap_R T_2) = T_\emptyset$, то есть для случая $C(T_1) \bigcap_R C(T_2) = T_\emptyset$ равенство $C(T_1 \bigcap_R T_2) = C(T_1) \bigcap_R C(T_2)$ выполняется всегда. Сформулируем полученный критерий.

Теорема 1. *Равенство* $C(T_1 \bigcap_R T_2) = C(T_1) \bigcap_R C(T_2)$ *выполняется при* $C(T_1) \bigcap_R C(T_2) = T_\emptyset$. *В противном случае это равенство выполняется при выполнении хотя бы одного из условий:*

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

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

Рассмотрим взаимосвязи между пересечением и активным дополнением. Для данных операций рассмотрены аналоги известных законов де Моргана (в данном случае можно говорить о взаимосвязях между пересечением, объединением и активным дополнением).

В [4] доказано включение $\tilde{T}_1 \bigcap_R \tilde{T}_2 \subseteq (T_1 \tilde{\bigcup_R} T_2)$. В [6] найдены и доказаны критерии перехода этого включения в равенство для случая $(T_1 \tilde{\bigcup_R} T_2) \neq T_\emptyset$. Из $\tilde{T}_1 \bigcap_R \tilde{T}_2 \subseteq (T_1 \tilde{\bigcup_R} T_2)$, непустоты схемы таблиц и $(T_1 \tilde{\bigcup_R} T_2) = T_\emptyset$ следует $\tilde{T}_1 \bigcap_R \tilde{T}_2 = T_\emptyset$, то есть при $(T_1 \tilde{\bigcup_R} T_2) = T_\emptyset$ это включение становится равенством. Сформулируем полученный критерий.

Теорема 2 (первый закон де Моргана). *Равенство* $\tilde{T}_1 \bigcap_R \tilde{T}_2 = (T_1 \tilde{\bigcup_R} T_2)$ *выполняется при* $(T_1 \tilde{\bigcup_R} T_2) = T_\emptyset$. *В противном случае это равенство выполняется при выполнении хотя бы одного из четырёх 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* $D_{A_q,T_1} \bigcap D_{A_q,T_2} \neq \emptyset$; *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* $\{(A_1, d_1), \ldots, (A_{q-1}, d_{q-1}), (A_q, z_1), (A_{q+1}, d_{q+1}), \ldots, (A_k, d_k)\} \in T_1$, *а строка* $\{(A_1, d_1), \ldots, (A_{q-1}, d_{q-1}), (A_q, z_2), (A_{q+1}, d_{q+1}), \ldots, (A_k, d_k)\} \in T_2$.

В [4] доказано включение $(T_1 \tilde{\bigcap_R} T_2) \subseteq \tilde{T}_1 \bigcup_R \tilde{T}_2$. В [6] найдены и доказаны критерии перехода этого включения в равенство для случая $(T_1 \tilde{\bigcap_R} T_2) \neq T_\emptyset$. Покажем, что это включение также превращается в равенство при $(T_1 \tilde{\bigcap_R} T_2) = T_\emptyset$ и выполнении любого из указанных в [6] условий.

Лемма 4. *Равенство $(T_1 \tilde{\bigcap_R} T_2) = \tilde{T}_1 \bigcup_R \tilde{T}_2$ при $(T_1 \tilde{\bigcap_R} T_2) = T_\emptyset$ выполняется тогда и только тогда, когда выполняется хотя бы одно из двух взаимоисключающих условий:*

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

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

Доказательство. Пусть $(T_1 \tilde{\bigcap_R} T_2) = T_\emptyset$ и выполняется условие (1). В [4] доказано, что $(T_1 \tilde{\bigcap_R} T_2) = T_\emptyset$ влечет $T_1 \bigcap_R T_2 = C(T_1 \bigcap_R T_2)$.

Допустим, что $T_1 \bigcap_R T_2 = T_\emptyset$. Тогда $T_1 = T_\emptyset$ и $T_2 = T_\emptyset$, так как существование строк в любой из таблиц T_1 или T_2 противоречит равенствам $D_{A,T_1} = D_{A,T_1 \bigcap_R T_2}$ и $D_{A,T_2} = D_{A,T_1 \bigcap_R T_2}$ соответственно. В этом случае $\tilde{T}_1 \bigcup_R \tilde{T}_2 = T_\emptyset$, то есть исходное включение превращается в равенство.

Пусть теперь $T_1 \bigcap_R T_2 \neq T_\emptyset$. Покажем, что в этом случае должны выполняться равенство $T_1 = T_2$. Допустим $T_1 \neq T_2$. Положим для определенности, что существует такая строка s, что $s \in T_1$ и $s \notin T_2$. Выполнение равенств условия (1) эквивалентно равенству активных доменов таблиц T_1, T_2 и $T_1 \bigcap_R T_2$, следовательно, $C(T_1) = C(T_2) = C(T_1 \bigcap_R T_2)$. Тогда $s \in T_1$ влечет $s \in C(T_1)$, поэтому $s \in C(T_1 \bigcap_R T_2)$. Из $s \notin T_2$ вытекает $s \notin T_1 \bigcap_R T_2$, то есть $s \in (T_1 \tilde{\bigcap_R} T_2)$, что противоречит предположению $(T_1 \tilde{\bigcap_R} T_2) = T_\emptyset$, поэтому $T_1 = T_2$. Тогда, заменив в исходном выражении таблицу T_2 таблицей T_1, получим верное равенство $\tilde{T}_1 = \tilde{T}_1$. Таким образом, мы показали, что при $(T_1 \tilde{\bigcap_R} T_2) = T_\emptyset$ и выполнении условия (1) исходное выражение превращается в равенство. Аналогично доказывается выполнимость исходного равенства при существовании такой строки s, что $s \notin T_1$ и $s \in T_2$.

Покажем теперь, что при $(T_1 \tilde{\bigcap_R} T_2) = T_\emptyset$, невыполнении условия (1) и выполнении условия (2) исходное выражение тоже превращается в равенство, то есть $\tilde{T}_1 \bigcup_R \tilde{T}_2 = T_\emptyset$.

От противного, допустим, что $\tilde{T}_1 \bigcup_R \tilde{T}_2 \neq T_\emptyset$. Для определенности положим $\tilde{T}_1 \neq T_\emptyset$ и $s = \{(A_1, d_1), \ldots, (A_k, d_k)\} \in \tilde{T}_1$; тогда $s \notin T_1$. Покажем, что в этом случае должны выполняться принадлежности $d_1 \in D_{A_1, T_1 \bigcap_R T_2}, \ldots, d_k \in D_{A_k, T_1 \bigcap_R T_2}$. Допустим, что это неверно, то есть $d_q \in D_{A_q, T_1 \bigcap_R T_2}$ для некоторого индекса q. Тогда по определению активного дополнения $s \notin T_1$ и выполняются принадлежности $d_1 \in D_{A_1, T_1}, \ldots, d_k \in D_{A_k, T_1}$, то есть $d_q \in D_{A_q, T_1} - D_{A_q, T_1 \bigcap_R T_2}$. По условию (2) в этом случае должна выполняться принадлежность $s \in T_1$, противоречие доказывает неверность допущения. Следовательно, таблица $T_1 \bigcap_R T_2$ непуста, а поскольку $(T_1 \tilde{\bigcap_R} T_2) = T_\emptyset$, то таблица $T_1 \bigcap_R T_2$ является насыщенной. Из этого и из выполнения принадлежностей $d_1 \in D_{A_1, T_1 \bigcap_R T_2}, \ldots, d_k \in D_{A_k, T_1 \bigcap_R T_2}$ следует, что $\{(A_1, d_1), \ldots, (A_k, d_k)\} \in T_1 \bigcap_R T_2$, поэтому $s \notin T_1$, что противоречит первоначальному допущению. Поэтому случай $\tilde{T}_1 \neq T_\emptyset$ невозможен. Точно таким же образом доказывается невозможность случая $\tilde{T}_2 \neq T_\emptyset$, поэтому $\tilde{T}_1 \bigcup_R \tilde{T}_2 = T_\emptyset$, то есть исходное равенство выполняется. $\square$

Таким образом, была доказана

Теорема 3 (второй закон де Моргана). $(T_1 \tilde{\bigcap_R} T_2) = \tilde{T}_1 \bigcup_R \tilde{T}_2$ *выполняется тогда и только тогда, когда выполняется хотя бы одно из двух взаимоисключающих условий из леммы 4.*

Рассмотрим взаимосвязи между пересечением и проекцией.

В [4] доказано включение $\pi_X(\bigcap_i T_i) \subseteq \bigcap_i \pi_X(T_i)$. В [7] найдены и доказаны критерии перехода этого включения в равенство для случая $\bigcap_i T_i \neq T_\emptyset$. Покажем, что это включение также превращается в равенство при $\bigcap_i T_i = T_\emptyset$ и выполнении указанного в [7] условия.

Теорема 4 (дистрибутивность проекции относительно пересечения). *Равенство* $\pi_X(\bigcap_i T_i) = \bigcap_i \pi_X(T_i)$ *при* $\bigcap_i T_i = T_\emptyset$ *выполняется тогда и только тогда, когда для каждой строки* $s = \{(X_1, x_1), \ldots, (X_p, x_p)\} \in \bigcap_i \pi_X(T)$ *существуют такие значения* $o_1 \in D_{O_1, \bigcap_i T_i}, \ldots, o_{k-p} \in D_{O_{k-p}, \bigcap_i T_i}$, *что строка* $s' = \{(X_1, x_1), \ldots, (X_p, x_p), (O_1, o_1), \ldots, (O_{k-p}, o_{k-p})\} \in \bigcap_i T_i$.

Доказательство. Пусть $\bigcap_i T_i \neq T_\emptyset$. В [7] доказано, что равенство $\pi_X(\bigcap_i T_i) = \bigcap_i \pi_X(T_i)$ выполняется тогда и только тогда, когда выполняется указанное условие.

Пусть теперь $\bigcap_i T_i = T_\emptyset$ От противного, допустим, что равенство не выполняется. Поскольку $\pi_X(T_\emptyset) = T_\emptyset$, то в таблице $\bigcap_i \pi_X(T)$ существует некоторая строка $s = \{(X_1, x_1), \ldots, (X_p, x_p)\}$. По условию леммы в этом случае должны существовать такие значения $o_1 \in D_{O_1, \bigcap_i T_i}, \ldots, o_{k-p} \in D_{O_{k-p}, \bigcap_i T_i}$, что строка $s' = \{(X_1, x_1), \ldots, (X_p, x_p), (O_1, o_1), \ldots, (O_{k-p}, o_{k-p})\} \in \bigcap_i T_i$, следовательно, $s' \in \bigcap_i T_i$, что противоречит предположению $\bigcap_i T_i = T_\emptyset$. $\Box$

4. Взаимосвязи между объединением и табличными операциями.

Операция объединения коммутативна, ассоциативна и идемпотентна, то есть выполняются равенства $T_1 \bigcup_R T_2 = T_2 \bigcup_R T_1$, $T_1 \bigcup_R (T_2 \bigcup_R T_3) = (T_1 \bigcup_R T_2) \bigcup_R T_3$ и $T \bigcup_R T = T$. Доказательство тривиально.

Рассмотрим взаимосвязи между объединением и разностью таблиц.

Лемма 5. *Выполняется включение* $T_1 \bigcup_R (T_2 \underset{R}{-} T_3) \supseteq (T_1 \bigcup_R T_2) \underset{R}{-} (T_1 \bigcup_R T_3)$.

Доказательство. Пусть $(T_1 \bigcup_R T_2) \underset{R}{-} (T_1 \bigcup_R T_3) \neq T_\emptyset$ и $s \in (T_1 \bigcup_R T_2) \underset{R}{-} (T_1 \bigcup_R T_3)$. Если при этом $s \in T_1$, то по определению объединения $s \in T_1 \bigcup_R (T_2 \underset{R}{-} T_3)$. Если же $s \notin T_1$, то по определению разности $s \in T_2$ и $s \notin T_1 \bigcup_R T_3$, следовательно, $s \notin T_3$. Тогда $s \in T_2 \underset{R}{-} T_3$, поэтому $s \in T_1 \bigcup_R (T_2 \underset{R}{-} T_3)$. $\Box$

Найдем критерий, при котором данное включение превращается в равенство.

Лемма 6. *Равенство* $T_1 \bigcup_R (T_2 \underset{R}{-} T_3) = (T_1 \bigcup_R T_2) \underset{R}{-} (T_1 \bigcup_R T_3)$ *выполняется тогда и только тогда, когда* $T_1 = T_\emptyset$.

Доказательство. Пусть $T_1 = T_\emptyset$. Тогда $T_1 \bigcup_R (T_2 \underset{R}{-} T_3) = (T_2 \underset{R}{-} T_3)$ и $(T_1 \bigcup_R T_2) \underset{R}{-} (T_1 \bigcup_R T_3) = (T_2 \underset{R}{-} T_3)$, то есть равенство выполняется.

Пусть теперь $T_1 \neq T_\emptyset$ и $s \in T_1$. Тогда $s \in T_1 \bigcup_R (T_2 \underset{R}{-} T_3)$. С другой стороны, $s \in T_1 \bigcup_R T_2$ и $s \in T_1 \bigcup_R T_3$, поэтому $s \notin (T_1 \bigcup_R T_2) \underset{R}{-} (T_1 \bigcup_R T_3)$, то есть равенство не выполняется. $\Box$

Лемма 7. *Выполняется включение* $T_1 \underset{R}{-} (T_2 \bigcup_R T_3) \subseteq (T_1 \underset{R}{-} T_2) \bigcup_R (T_1 \underset{R}{-} T_3)$.

Доказательство. Пусть $T_1 \underset{R}{-} (T_2 \bigcup_R T_3) \neq T_\emptyset$ и $s \in T_1 \underset{R}{-} (T_2 \bigcup_R T_3)$. Тогда, по определению разности, $s \in T_1$ и $s \notin T_2 \bigcup_R T_3$, поэтому $s \in T_1$, $s \notin T_2$ и $s \notin T_3$. Из $s \in T_1$ и $s \notin T_2$ следует $s \in T_1 \underset{R}{-} T_2$, поэтому $s \in (T_1 \underset{R}{-} T_2) \bigcup_R (T_1 \underset{R}{-} T_3)$. $\Box$

Найдем критерий, при котором данное включение превращается в равенство.

Лемма 8. *Равенство* $T_1 \underset{R}{-} (T_2 \underset{R}{\bigcup} T_3) = (T_1 \underset{R}{-} T_2) \underset{R}{\bigcup} (T_1 \underset{R}{-} T_3)$ *выполняется тогда и только тогда, когда* $T_1 \underset{R}{\bigcap} T_2 = T_1 \underset{R}{\bigcap} T_3$.

Доказательство. Необходимость. Пусть $T_1 = T_\emptyset$. Тогда $T_1 \underset{R}{-} (T_2 \underset{R}{\bigcup} T_3) = T_\emptyset$ и $(T_1 \underset{R}{-} T_2) \underset{R}{\bigcup} (T_1 \underset{R}{-} T_3) = T_\emptyset$, то есть равенство выполняется. Пусть теперь $T_1 \neq T_\emptyset$ и $(T_1 \underset{R}{-} T_2) \underset{R}{\bigcup} (T_1 \underset{R}{-} T_3) \neq T_\emptyset$. Тогда существует $s \in (T_1 \underset{R}{-} T_2) \underset{R}{\bigcup} (T_1 \underset{R}{-} T_3)$, следовательно выполняется хотя бы одна из принадлежностей $s \in T_1 \underset{R}{-} T_2$ или $s \in T_1 \underset{R}{-} T_3$, поэтому $s \in T_1 \wedge s \notin T_2$ или $s \in T_1 \wedge s \notin T_3$. От противного, допустим, что $s \notin T_1 \underset{R}{-} (T_2 \underset{R}{\bigcup} T_3)$. Поскольку выполняется $s \in T_1$, это возможно только в том случае, когда $s \in T_2 \underset{R}{\bigcup} T_3$, что невозможно ввиду выполнения равенства $T_1 \underset{R}{\bigcap} T_2 = T_1 \underset{R}{\bigcap} T_3$ и хотя бы одного из условий $s \notin T_2$ или $s \notin T_2$; это доказывает необходимость.

От противного докажем достаточность утверждения. Пусть $T_1 \underset{R}{\bigcap} T_2 \neq T_1 \underset{R}{\bigcap} T_3$. Покажем, что в этом случае равенство не выполняется. $T_1 \underset{R}{\bigcap} T_2 \neq T_1 \underset{R}{\bigcap} T_3$ влечет $T_1 \neq T_\emptyset$. Для определенности положим, что существует такая строка s, что $s \in T_1 \underset{R}{\bigcap} T_2$ и $s \notin T_1 \underset{R}{\bigcap} T_3$. Тогда $s \in T_1$, $s \in T_2$ и $s \notin T_3$, следовательно, $s \in T_2 \underset{R}{\bigcup} T_3$, то есть $s \notin T_1 \underset{R}{-} (T_2 \underset{R}{\bigcup} T_3)$; при этом $s \in T_1 \underset{R}{\bigcup} T_3$, следовательно, $s \in (T_1 \underset{R}{-} T_2) \underset{R}{\bigcup} (T_1 \underset{R}{-} T_3)$. Аналогично доказывается невыполнимость равенства при существовании такой строки s, что $s \notin T_1 \underset{R}{\bigcap} T_2$ и $s \in T_1 \underset{R}{\bigcap} T_3$. $\square$

Рассмотрим взаимосвязи между объединениями и насыщением таблиц. В [4] доказано включение $C(T_1 \underset{R}{\bigcup} T_2) \supseteq C(T_1) \underset{R}{\bigcup} C(T_2)$. В [5] найдены и доказаны критерии перехода этого включения в равенство для случая $T_1 \neq T_\emptyset$ и $T_2 \neq T_\emptyset$. Несложно видеть что при $T_1 = T_\emptyset$ указанное включение превращается в равенство $C(T_2) = C(T_2)$, а при $T_2 = T_\emptyset$ – в равенство $C(T_1) = C(T_1)$. Сформулируем полученный критерий.

Теорема 5. *Если* $T_1 = T_\emptyset$ *или* $T_2 = T_\emptyset$*, то равенство* $C(T_1 \bigcup T_2) = C(T_1) \bigcup C(T_2)$ *выполняется. В противном случае это равенство выполняется при выполнении хотя бы одного из условий:*

1) существует не более одного атрибута, для которого значения активного домена относительно таблиц T_1 *и* T_2 *различается;*

2) выполняется хотя бы одно включение $\forall i\ D_{A_i,T_1} \subseteq D_{A_i,T_2}$ *или* $\forall i\ D_{A_i,T_2} \subseteq D_{A_i,T_1}$.

Рассмотрим взаимосвязи между объединением и проекцией. В [4] доказано выполнения равенства $\pi_X(\bigcup\limits_i T_i) = \bigcup\limits_i \pi_X(T_i)$, то есть проекция дистрибутивна относительно объединения таблиц.

5. Выводы.

В работе рассмотрены взаимосвязи пересечения и объединения таблиц с другими сигнатурными операциями табличных алгебр: разностью, селекцией, проекцией, насыщением, активным дополнением, соединением, переименованием атрибутов. Найдены необходимые и достаточные условия, при которых некоторые взаимосвязи, которые для общего случая являются включениями, переходят в равенства. Эти результаты могут быть использованы для преобразования выражений, содержащих табличные операции с целью их минимизации или перехода к стандартному виду. В дальнейшем планируется рассмотреть взаимосвязи между другими табличными операциями, а также найти вероятности того, что найденные в работе критерии будут выполняться для произвольных таблиц и таблиц специального вида.

Цитированная литература

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., Буй Д.Б.* К основаниям теории реляционных моделей баз данных // Кибернетика и системный анализ. – 1996. – Т. 32, № 4. – С. 3–12.
3. *Гарсиа-Молина Г., Ульман Дж., Уидом Дж.* Системы баз данных. – Москва: «Вильямс», 2004. – 1088 с.
4. *Редько В.Н., Брона Ю.Й., Буй Д.Б., Поляков С.А.* Реляційні бази даних: табличні алгебри та SQL-подібні мови. – Київ: «Академперіодика», 2001. – 198 с.
5. *Сенченко А.С.* О дистрибутивности в табличных алгебрах операции насыщения относительно операций объединения и пересечения // Труды ИПММ НАН Украины. – 2013. – Т. 26. – С. 197–204.
6. *Сенченко О.С.* Закони де Моргана у табличних алгебрах // Вісник Київського нац. ун-ту ім. Т. Шевченка. Серія: фіз.-мат. науки. Спецвипуск. – 2013. – С. 158–163.
7. *Сенченко А.С.* Свойства операции проекции в табличных алгебрах // Труды ИПММ НАН Украины. – 2014. – Т. 28. – С. 137–144.

References

1. Codd, E.F. (1996). A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM, 13*(6), 377-387.
2. *Red'ko V.N., Bui D.B.* (1996). Foundations of the theory of relational database models. Cybernetics and Systems Analysis, Vol.**32**, Issue 4, pp. 471–478.
3. Red'ko, V.N., Bui, D.B. (1996). Foundations of the theory of relational database models. *Cybernetics and Systems Analysis, 32*(4), 471-478.
4. Garcia-Molina, H., Ullman, J.M., Widom, J. (2008). *Database Systems: The Complete Book* (2nd Edition). Pearson.
5. Red'ko, V.N., Brona, Yu. Y., Bui, D.B., Polyakov, S.A. (2001). *Relational databases: table algebras and SQL-like languages.* Kyiv: Academperiodica (in Ukrainian).
6. Senchenko, A.S. (2013). On a distributivity of a saturation to join and intersection in table algebra. *Trudy IPMM NAN Ukrainy, 26*, 197-204 (in Russian).
7. Senchenko, O.S. (2013). De Morgan's laws in table algebra. *Visnyk Kievskgo nacional'nogo un-ta im. T. Shevchenko. Seriya: fiz.-mat. nauki. Spec. issue*, 158-163 (in Ukrainian).
8. Senchenko, A.S. (2014). The properties of projection operation in table algebra. *Trudy IPMM NAN Ukrainy, 28*, 137-144 (in Russian).

A.S. Senchenko

The interrelations between intercestion, union and other signature operations in table algebra.

Currently, databases are widely used in almost all areas of human activity. For all variety of different types of databases the most common are relational (table) databases, mathematical model of which was proposed by E. Codd. From mathematical point of view, a relational database is a finite set of finite relations between different predefined sets of basic data. Table algebra introduced by V.N. Red'ko and D.B. Buy is based on Codd's relational algebra and significantly improves it. It formed the theoretical foundation of modern database query language. Elements of the carrier of table algebra specify relational data structures, and signature operations are based on the basic table manipulations in relational algebra and SQL-like languages. One of the most actual tasks in relational and table algebras is the problem of equivalent transformation of expressions in order to minimize or reduce them to a standard form; it is one of the stages of query optimization, and can also significantly reduce the processing time of information in relational database management systems. For the decision of this problem the interrelations between the basic table operations are used. In the present, a significant number of such interrelations have been established, most of which for the general case are performed as inclusions. The author has found criteria for the transition of some such inclusions into equalities. These criteria are expressed in terms of the active domains of the tables and are natural. In this paper, the interrelations of the intersection and the union of tables with other signature operations of table algebras: difference, selection, projection, saturation, active complement, join, renaming of attributes are considered.

Keywords: *intersection, union, databases, table algebra.*

О.С. Сенченко

Взаємозв'язки між перетином, об'єднанням та іншими сигнатурними операціями в табличних алгебрах.

В наш час реляційні бази даних, математична модель яких була запропонована Е. Коддом, залишаються найбільш розповсюдженими. Табличні алгебри побудовані на основі реляційних алгебр Е. Кодда, суттєво їх розвивають та складають теоретичний фундамент мов запитів сучасних табличних баз даних. У реляційних та табличних алгебрах актуальною є задача еквівалентного перетворення виразів із метою їх мінімізації або приведення до стандартного вигляду; вона є одним із етапів оптимізації запитів, її розв'язання може суттєво зменшити час обробки інформації в реляційних системах управління базами даних. Для розв'язування цієї задачі використовуються взаємозв'язки між основними табличными операціями. На сьогодні встановлено значну кількість таких взаємозв'язків, більшість із яких для загального випадку виконуються в вигляді включень. Автором були знайдені критерії переходу деяких таких включень у рівності. У роботі розглянуто взаємозв'язки перетину та об'єднання таблиць з іншими сигнатурними операціями табличних алгебр: різницею, селекцією, проекцією, насиченням, активним доповненням, з'єднанням, перейменуванням атрибутів.

Ключові слова: *перетин, об'еднання, база даних, табличні алгебри.*

Институт прикладной математики и механики НАН Украины, Славянск
senchenko.a76@gmail.com

Получено 18.12.18