МАТЕМАТИЧЕСКОЕ И АРХИТЕКТУРНОЕ ОБЕСПЕЧЕНИЕ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛЕНИЙ
(Вычислительные системы)
1985 год Выпуск 109

УДК 681.324

ОБ ОДНОМ МЕТОДЕ ОБРАБОТКИ МАССИВОВ

Е.В.Суворов

В работе рассматривается один из возможных подходов к выполнению информационно-логических операций различной сложности,основанный на предварительном упорядочении единого массива, состоящего из всех массивов, участвующих в операции.

Упорядочение отдельных массивов с целью ускорения выполнения некоторых информационно-логических операций широко используется [1]. Один из способов выполнения основных теоретико-множественных операций описан в [2]. Особенность предлагаемого ниже метода заключается в том, что каждый из массивов, участвующих в выполняемой операции, предварительно снабжается своим двоичным кодом меток, а реализация заданной операции сводится к упорядочению единого массива (вместе с метками), после чего выполняется последовательность простых логических и специальных операций над векторами меток, приводящая к необходимому результату.

Вначале вводятся необходимые определения. Затем рассматривается выполнение различных информационно-логических операций над двумя массивами. Вводятся определения специальных операций над двоичными векторами. Дается обобщение предлагаемого метода,обеспечивающее выполнение операций над произвольным количеством массивов. Далее рассматриваются способы выполнения операций α-алгебры.

Пусть B - произвольный упорядоченный по возрастанию массив, содержащий n элементов b_i $(i = \overline{1,n})$. Разделительным вектором φ упорядоченного массива B назовем двоичный вектор размерности n, содержащий s единиц $(s \le n)$ i-й разряд которого $\varphi^i = 0$, если $b_i = b_{i+1}$, и $\varphi^i = 1$, если $b_i \ne b_{i+1}$ $(i = \overline{1,n-1})$, а n-й разряд $\varphi^n = 1$. Таким образом, упорядоченный массив B разделяется вектором φ на s подмассивов $B^1,\dots,B^r,\dots,B^s$ так, что все элементы в r-м подмассиве B^r одинаковы $(r=\overline{1,s})$, при этом позиции единиц вектора φ соответствуют позициям последних элементов s подмассивов, во всех остальных его позициях содержатся нули (рис.I). Значения элементов верхних подмассивов меньше значений элементов нижних подмассивов.

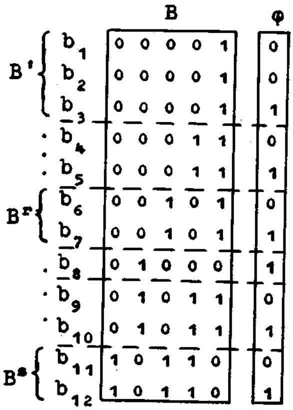

Рис.I

Пусть $A_i(W_i,V_i)$ – произвольный массив, состоящий из n_i m_i-разрядных кортежей, где домены[ж)] W_i и V_i могут быть составными, а m_{W_i} и m_{V_i} – разрядности элементов доменов W_i и V_i соответственно $(m_i = m_{W_i} + m_{V_i})$. Массив $A(W,\underline{V}) = (A_1(W_1,V_1),\dots,A_i(W_i,V_i),\dots,A_k(W_k,V_k))$ размерности $n=\sum_{i=1}^{k} n_i$, состоящий из k массивов $A_i(W_i,V_i)$ $(i=\overline{1,k})$, назовем **единым массивом** $A(W,\underline{V})$ **по домену** V, если домены V_i массивов $(i=\overline{1,k})$ совместимы по объединению. При этом разрядности m_{W_i} $(i=\overline{1,k})$ могут не совпадать. На рис.2 показан массив $A(W,\underline{V})$ для $k=4$.

Если $A(W,\underline{V})$ – единый массив, то **вектором меток** l_i массива A_i назовем двоичный вектор размерности n, содержащий единицы в тех позициях, которым соответствуют позиции кортежей массива A_i, и нули – во всех остальных. Совокупность $L = (l_1,\dots,l_i,\dots,l_k)$ векторов, соответствующих массивам $A_1,\dots,A_i,\dots,A_k$, можно рассматривать как массив меток единого массива $A(W,\underline{V})$. Вектор l_i представляет собой i-й столбец массива L (рис.2). Тогда если в p-й строке единого массива $A(W,\underline{V})$ $(p=\overline{1,n})$ содержится кортеж A_i-го массива, то $l_{ip}=1$, а $l_{jp}=0$ для $j\neq i$ $(j=\overline{1,k})$. Очевидно, что $\bigvee_{i=1}^{k} l_i = \underline{1}$, $l_i \wedge l_j = \underline{0}$ для $i\neq j$

ж) Здесь используются понятия "кортеж" и "домен", принятые в теории реляционных баз данных [3].

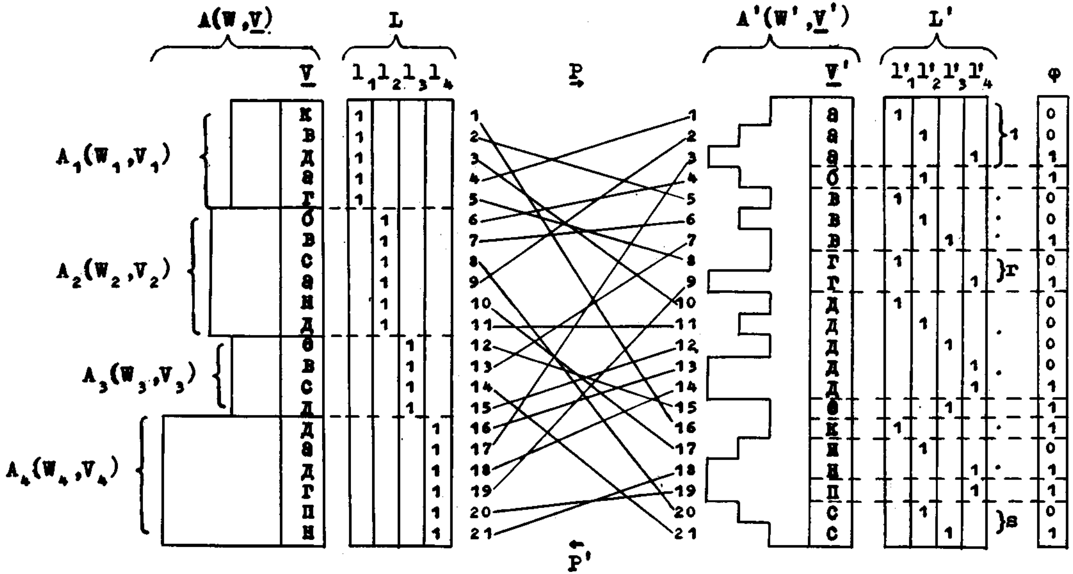

Рис.2. Упорядочение единого массива $A(W,\underline{V})$ и формирование векторов меток L' и вектора φ.

($i = \overline{1,k}$; $j = \overline{1,k}$), где $\underline{1}(\underline{0})$ – двоичный вектор размерности n, со – держащий только единицы (нули).

Операцией упорядочения $\mathcal{U}[A(W,\underline{V}),L] \to A'(W',\underline{V}'),L',\varphi$ назовем процедуру, которая преобразует исходный единый массив $A(W,\underline{V})$ (вместе с соответствующими строками массива меток L) так, чтобы элементы домена V располагались в порядке возрастания. При этом формируются единый упорядоченный по домену $\underline{V}$ массив $A'(W',\underline{V}')$, перегруппированный по строкам массив меток L', и разделительный вектор φ размерности n (рис.2).

Пусть $A = a_1,\ldots,a_i,\ldots,a_n$ – произвольный массив, содержащий n элементов a_i и $P = p_1,\ldots,p_i,\ldots,p_n$ – численный вектор той же размерности, компонентами p_i которого являются n первых чисел натурального ряда. Тогда операцией перестановки $P(A) = C$ называется процедура, которая по заданным A и P строит массив $C = c_1,\ldots,c_i,\ldots,c_n$ так, что i-я компонента a_i массива A переставляется в позицию p_i массива C, т.е. $c_{p_i} = a_i$ ($i = \overline{1,n}$). Если при этом элементы массива C располагаются в порядке возрастания, то вектор P называется упорядочивающим вектором массива A. Операцией обратной перестановки $P'(A) = C$ называется процедура, которая по заданным A и P строит массив C так, что i-я компонента c_i массива C равна a_j, где $j = p_i$ ($i = \overline{1,n}$; $j = \overline{1,n}$). Очевидно, что если $P(A) = A'$, то $P'(A') = A$.

Операцией упорядочения $\mathcal{B}[A(W,\underline{V})] \to P,\varphi$ назовем процедуру, которая на основе упорядочения по возрастанию элементов домена $\underline{V}$ исходного единого массива $A(W,\underline{V})$ формирует упорядочивающий вектор P для домена $\underline{V}$ и разделительный вектор φ. Размерности P и φ равны n. Для рассматриваемого примера массива $A(W,\underline{V})$ на рис.2 показан вектор φ, а также граф соединений, отображающий перестановку, соответствующую упорядочивающему вектору P. Очевидно, что если P – упорядочивающий вектор для домена $\underline{V}$, то $P(A(W,\underline{V})) = A'(W',\underline{V}')$ и $P(L) = L'$.

Если во всех массивах A_i единого массива $A(W,\underline{V})$ домены W_i ($i = \overline{1,k}$) отсутствуют, то операцию упорядочения $\mathcal{U}$ будем записывать как $\mathcal{U}[A,L] \to A',L',\varphi$, а операцию упорядочения $\mathcal{B}$ – как $\mathcal{B}[A] \to P,\varphi$. Если отсутствует массив меток L, то операцию упорядочения $\mathcal{U}$ будем записывать как $\mathcal{U}[A] \to A',\varphi$.

Далее рассмотрим выполнение информационно-логических операций над массивами с использованием операций упорядочения $\mathcal{U}$ и $\mathcal{B}$ и операций P и P'. При этом по мере необходимости будут вводиться определения. Без нарушения общности можно считать, что в единых массивах A домены W отсутствуют.

I. Исключение из массива дублирующих элементов. Пусть A_1 – массив размерности n_1, в котором имеются одинаковые элементы. Необходимо получить массив, в котором каждый из различных элементов содержится по одному разу. Применяя к A_1 операцию упорядочения $\mathcal{U}[A_1] \rightarrow A_1', \varphi$, получаем вектор φ, содержащий s единиц, который разделяет упорядоченный массив A_1' на s подмассивов. Все элементы r-го подмассива $(r=\overline{1,s})$ одинаковы и, так как позиции единиц вектора φ соответствуют позициям последних элементов s подмассивов, φ отмечает единицами те элементы из A_1', которые составляют массив, не содержащий дублирующих элементов.

Очевидно, что операции $\mathcal{B}[A_1] \rightarrow P,\varphi$; $P'(\varphi) = z$ формируют двоичный вектор z, отмечающий единицами те элементы исходного массива A_1, которые также составляют массив, не содержащий дублирующих элементов.

2. Объединение двух массивов. Пусть A_1 и A_2 – совместимые по объединению массивы размерностей n_1 и n_2 соответственно. Необходимо получить массив $A_3 = A_1 \cup A_2$. Если $A = (A_1, A_2)$ – единый массив размерности $n = n_1 + n_2$, то, применяя к A операцию $\mathcal{U}[A] \rightarrow A', \varphi$, получаем вектор φ размерности n, отмечающий единицами те элементы упорядоченного массива A', которые составляют массив $A_3 = A_1 \cup A_2$, не содержащий дублирующих элементов. В то же время операции $\mathcal{B}[A] \rightarrow P,\varphi$; $P'(\varphi) = z$ формируют вектор z размерности n, отмечающий единицами те элементы исходных массивов A_1 и A_2, которые составляют массив $A_3 = A_1 \cup A_2$, не содержащий дублирующих элементов.

3. Выделение элементов одного массива, входящих (или не входящих) в другой. Пусть A_1 и A_2 – совместимые по объединению массивы размерностей n_1 и n_2 соответственно. Нужно выделить элементы массива A_1, входящие в A_2. Если $A = (A_1, A_2)$ – единый массив размерности $n = n_1 + n_2$ и l_1 – вектор меток (размерности n) массива A_1, то, применяя к A операцию $\mathcal{U}[A, l_1] \rightarrow A', l', \varphi$, получаем вектор φ, содержащий s единиц, который разделяет упорядоченный массив

A' на s подмассивов, а вектор $1_1'$ – на s сегментов, при этом каждому r-му подмассиву соответствует свой r-й сегмент вектора $1_1'(r = \overline{1,s})$. Единицы вектора $1_1'$ отмечают элементы массива A_1 в A' (рис. 3).

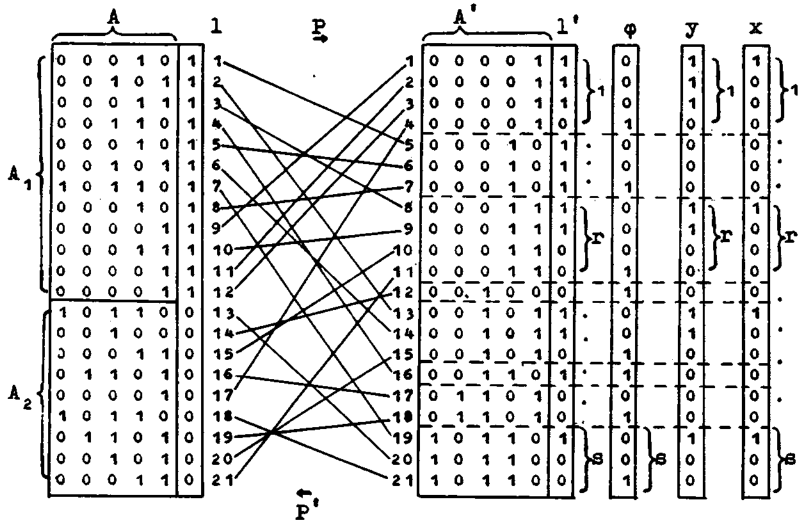

Рис. 3

В е к т о р о м   в х о ж д е н и я y назовем двоичный вектор размерности n, отмечающий единицами те элементы массива A_1 из A', которые входят в A_2. По векторам φ и $1_1'$ можно сформировать вектор вхождения, используя следующие предложения.

а) Если r-й сегмент $1_1'$ содержит нули и единицы, то r-й подмассив содержит элементы из A_2 (отмеченные нулями) и элементы из A_1 (отмеченные единицами). Так как все элементы r-го подмассива одинаковы, то элементы из A', отмеченные единицами $1_1'$ в r-м сегменте, являются элементами из A_1 и входят в A_2. В этом случае содержимое r-го сегмента вектора вхождения y совпадает с содержимым r-го сегмента $1_1'$ (рис.3).

б) Если r-й сегмент $1_1'$ содержит только нули (единицы), то r-й подмассив содержит только одинаковые элементы из $A_2(A_1)$. В этих случаях r-й сегмент вектора вхождения y должен содержать только нули.

Итак, для формирования вектора вхождения по векторам φ и $1_1'$ справедливо следующее правило: если r-й сегмент $1_1'$ содержит хотя бы один нуль, то содержимое этого сегмента есть содержимое соответствующего r-го сегмента вектора вхождения, если же r-й сегмент $1_1'$ содержит только единицы, то соответствующий сегмент вектора вхождения - полностью нулевой $(r = \overline{1,s})$.

Операцией $\hat{\alpha}(\varphi,1) = y$ назовем процедуру, которая по разделительному вектору φ и двоичному вектору 1 строит вектор y в соответствии с приведенным правилом. Размерности $\varphi, 1$ и y одинаковы.

Таким образом, операции $\mathcal{U}[A,1_1] \to A',1_1',\varphi$; $\hat{\alpha}(\varphi,1_1')=y$ формируют вектор y, отмечающий единицами те элементы в A', которые являются элементами A_1 и входят в A_2 (рис.3).

Чтобы определить, какие элементы из A_1 не входят в A_2, достаточно выполнить далее операцию покомпонентной конъюнкции $1_1'\cdot\bar{y} = d$, где $\bar{y}$ - инверсия вектора y. Тогда вектор d отмечает те элементы массива A_1 из A', которые не входят в A_2.

Очевидно, что вектор $y_1 = \hat{\alpha}(\varphi,\overline{1_1'})$, где $\overline{1_1'}$ - инверсия $1_1'$, отмечает единицами те элементы массива A_2 из A', которые входят в A_1. Выполняя далее операцию $\overline{1_1'}\cdot\bar{y}_1 = d_1$, получаем вектор d_1, отмечающий единицами те элементы A_2 из A', которые не входят в A_1.

В то же время операции $\mathcal{B}[A] \to P,\varphi$; $P(1_1) = 1_1'$; $\hat{\alpha}(\varphi, 1_1') = y$; $P'(y) = z$ формируют вектор z, отмечающий единицами элементы исходного массива A_1, входящие в A_2. Если здесь после операции $\hat{\alpha}$ выполнить операции $1_1'\cdot\bar{y} = d$; $P'(d) = z_1$, то получим вектор z_1, отмечающий единицами элементы исходного массива A_1, не входящие в A_2.

Последовательность операций $\mathcal{B}[A] \to P,\varphi$: $P(1_1) = 1_1'$; $\hat{\alpha}(\varphi,1_1') = y_1$; $P'(y_1)=z_2$ формирует вектор z_2, отмечающий единицами элементы исходного массива A_2, входящие в A_1. Если здесь после операции $\hat{\alpha}$ выполнить операции $\overline{1_1'}\cdot\bar{y}_1 = d_1$; $P'(d_1) = z_3$, то получим вектор z_3, отмечающий единицами элементы исходного массива A_2, не входящие в A_1.

4. **Определение полного вхождения одного массива в другой.** Из приведенного выше правила видно: если массив A_1 входит в A_2, то среди всех s сегментов $1_1'$ не существует ни одного сегмента, который содер-

жит только единицы, т.е. в этом случае любой r-й сегмент l_1' должен содержать хотя бы один нуль ($r = \overline{1,s}$) (рис.3).

Операцией $\hat{\beta}(\varphi,l)=\beta$ назовем процедуру, которая по разделительному вектору φ и двоичному вектору l формирует значение $\beta = 1$, если каждый из s сегментов вектора l содержит хотя бы один нуль, и $\beta = 0$ – в противном случае.

Таким образом, если $A=(A_1,A_2)$ – единый массив и l_1 – вектор меток массива A_1, то операции $\mathcal{U}[A,l_1] \to A',l_1',\varphi$; $\hat{\beta}(\varphi,l_1') = \beta$ (или $\mathcal{B}[A] \to P,\varphi$; $P(l_1) = l_1'$; $\hat{\beta}(\varphi,l_1')=\beta$) формируют значение $\beta = 1$, если A_1 входит в A_2, и $\beta = 0$ – в противном случае. Очевидно, что если A_2 входит в A_1, то $\hat{\beta}(\varphi, \overline{l_1'}) = 1$, в противном случае $\hat{\beta}(\varphi, \overline{l_1'}) = 0$. В примере на рис.3 $\beta = 0$, так как 2-й и 6-й сегменты l_1' содержат только единицы.

5. Пересечение двух массивов. Пересечение двух массивов A_1 и A_2 есть массив $A_3 = A_1 \cap A_2$, состоящий из элементов A_1, входящих в A_2 (или наоборот), не содержащий дублирующих элементов. Если вектор вхождения y, полученный в результате операции $\hat{\alpha}(\varphi, l_1') = y$, в r-м сегменте содержит более одной единицы, то элементы из A_1, отмеченные единицами r-го сегмента y, одинаковы (рис.3). Для исключения дублирующих элементов в этом сегменте оставляется одна единица, например, ближайшая к $(r-1)$-му сегменту.

Операцией $\gamma(\varphi,l) = x$ назовем процедуру, которая по разделительному вектору φ и двоичному вектору l строит вектор x по правилу: если r-й сегмент l содержит одну едницу или более, то в r-м сегменте x единственная единица содержится в той же позиции, что и единица r-го сегмента l, ближайшая к $(r-1)$-му сегменту, если же r-й сегмент l содержит только нули, то r-й сегмент x полностью нулевой ($r = \overline{1,s}$). Размерности φ, l и x одинаковы.

Таким образом, операции $\mathcal{U}[A,l] \to A', l_1', \varphi$; $\hat{\alpha}(\varphi,l_1') = y$; $\gamma(\varphi,y) = x$ формируют вектор x (рис.3), отмечающий единицами элементы A_1 из A', которые входят в A_2 и составляют массив $A_3 = A_1 \cap A_2$, не содержащий дублирующих элементов. В то же время последовательность операций $\mathcal{B}[A] \to P,\varphi$; $P(l_1) = l_1'$; $\hat{\alpha}(\varphi,l_1') = y$; $\gamma(\varphi, y) = x$; $P'(x) = z$ формирует вектор z, отмечающий единицами элементы исходного массива A_1, составляющие массив $A_3 = A_1 \cap A_2$.

6. Разность двух массивов. Если $A = (A_1,A_2)$ – единый массив и l_1 – вектор меток массива A_1, то

операции $\mathcal{U}[A,l_1] \to A', l_1', \varphi$; $\hat{\alpha}(\varphi, l_1') = y$; $l_1' \cdot \bar{y} = d$; $\gamma(\varphi, d) = x$ (где d отмечает единицами элементы A_1 из A', не входящие в A_2) формируют вектор x, отмечающий единицами элементы из A', составляющие массив $A_3 = A_1 \setminus A_2$. Операции $\mathcal{B}[A] \to P, \varphi$; $P(l_1) = l_1'$; $\hat{\alpha}(\varphi, l_1') = y$; $l_1' \cdot \bar{y} = d$; $\gamma(\varphi, d) = x$; $P'(x) = z$ формируют вектор z, отмечающий единицами элементы исходного массива A_1, составляющие массив $A_3 = A_1 \setminus A_2$.

До сих пор мы имели дело с информационно-логическими операциями над двумя массивами. Далее будут рассмотрены основные операции над единым массивом, состоящим из k массивов ($k > 2$).

7. Выделение элементов массива A_i, входящих в массив A_j ($i,j = \overline{1,k}$; $i \neq j$). Пусть $A = (A_1, \ldots, A_i, \ldots, A_k)$ – единый массив размерности $n = \sum_{i=1}^{k} n_i$, состоящий из k массивов A_i размерности n_i ($n = \overline{1,k}$), а $l = (l_1, \ldots, l_i, \ldots, l_k)$ – массив меток единого массива A (рис.4). Применяя к A операцию $\mathcal{U}[A, L] \to A', L', \varphi$, получаем вектор φ, который содержит s единиц и разделяет упорядоченный массив A' на s подмассивов и каждый вектор l_i' из L' ($i = \overline{1,k}$) – на s сегментов. При этом r-му подмассиву соответствует r-й сегмент ($r = \overline{1,s}$) каждого вектора l_i' ($i = \overline{1,k}$). Единицы вектора l_i' отмечают элементы A_i из A' ($i = \overline{1,k}$).

Введем определения. Расширенным вектором вхождения y назовем двоичный вектор размерности n, отмечающий единицами те элементы A_i из A', которые входят в массив A_j ($i,j = \overline{1,k}$; $i \neq j$). Промежуточным вектором t^{ij} векторов l_i' и l_j' ($i,j = \overline{1,k}$; $i \neq j$) назовем вектор размерности n, p-я компонента которого будет $t_p^{ij} = 0$, если $l_{ip}' = 0$ и $l_{jp}' = 1$; $t_p^{ij} = 1$, если $l_{ip}' = 1$ и $l_{jp}' = 0$; $t_p^{ij} = M$, если $l_{ip}' = 0$ и $l_{jp}' = 0$ ($p = \overline{1,n}$). Вектор t^{ij} разделен вектором φ на s сегментов (рис. 4).

Покажем, что по векторам φ, l_i', l_j' и t^{ij} можно сформировать расширенный вектор вхождения, используя следующие предложения.

а) Если r-й сегмент t^{ij} содержит как нули, так и единицы (независимо от того, есть ли в этом сегменте еще и M), то r-й подмассив, кроме элементов других массивов, отмеченных M (если таковые имеются), содержит элементы из A_i (отмеченные единицами

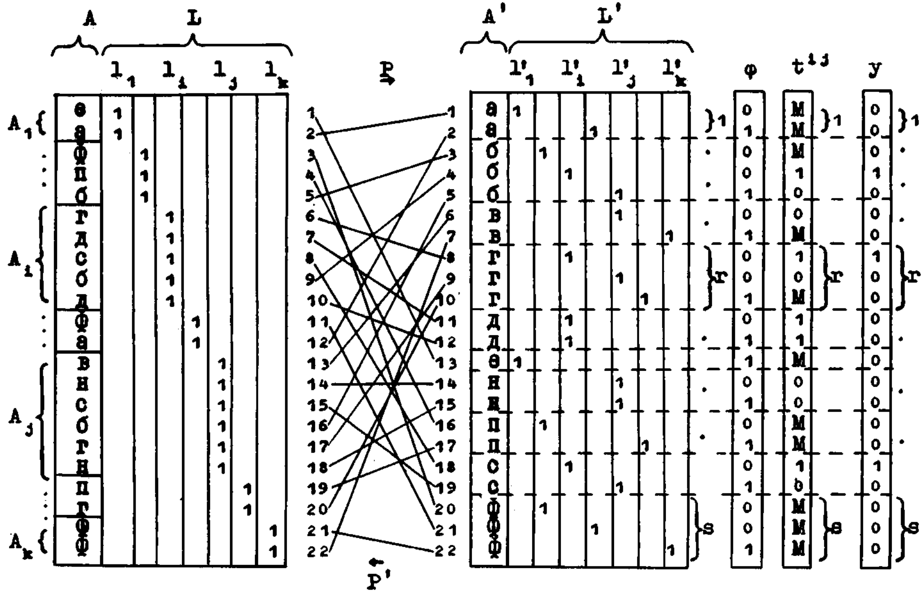

Рис.4. Выделение элементов массива A_i, входящих в A_j.

l_j'). Так как все элементы r-го подмассива одинаковы, то элементы A_i r-го подмассива входят в A_j. В этом случае содержимое r-го сегмента расширенного вектора вхождения y совпадает с содержимым r-го сегмента l_i' (рис.4).

б) Если r-й сегмент t^{ij}, кроме М, когда таковые имеются, содержит только нули (единицы), то r-й подмассив, кроме элементов других массивов, отмеченных М, содержит одинаковые элементы из A_j, отмеченные единицами l_j' (из A_i, отмеченные единицами l_i') и не содержит элементов из A_i (A_j). В этих случаях r-й сегмент расширенного вектора вхождения y содержит только нули.

в) Если r-й сегмент t^{ij} содержит только М, то r-й подмассив содержит элементы (одинаковые) из других массивов, но не содержит элементов из A_i и A_j. В этом случае r-й сегмент расширенного вектора вхождения y содержит только нули.

Итак, для формирования расширенного вектора вхождения y по векторам φ, l_i' и l_j' справедливо следующее правило: по векторам l_i' и l_j' формируется вектор t^{ij}; если r-й сегмент t^{ij} содержит хотя бы один нуль или только М, то содержимое этого сегмента l_i' есть содержимое соответствующего сегмента расширенного вектора вхождения y, если же r-й сегмент t^{ij}, кроме М, содержит только единицы, то r-й сегмент расширенного вектора вхождения y содержит только нули ($r=\overline{1,s}$).

О п е р а ц и е й $\alpha(\varphi,l,\omega) = y$ назовем процедуру, которая по разделительному вектору φ и векторам l и ω строит вектор y в соответствии с приведенным правилом.

Из этого правила видно, что если массив A_i входит в A_j, то среди всех s сегментов t^{ij} не существует ни одного сегмента, который содержал бы только единицы (и, быть может, М) (рис.4).

О п е р а ц и е й $\beta(\varphi, l, \omega) = \beta$ назовем процедуру, которая по разделительному вектору φ и векторам l и ω строит промежуточный вектор $t^{l\omega}$, а затем формирует значение $\beta=1$, если каждый из s сегментов $t^{l\omega}$ содержит хотя бы один нуль (или только М), и $\beta=0$ – в противном случае.

Таким образом, если $A = (A_1,\ldots,A_i,\ldots,A_k)$ и $L = (l_1,\ldots,l_i,\ldots,l_k)$, то операции $\mathcal{U}[A,L] \to A',L',\varphi$; $\alpha(\varphi,l_i',l_j') = y$ формируют расширенный вектор вхождения y (рис.4), отмечающий единицами элементы A_i из A', входящие в A_j ($i,j = \overline{1,k}$; $i \neq j$). В то же время последовательность операций $\mathcal{B}[A] \to P,\varphi$; $P(l_i) = l_i'$;

$P(l_j) = l'_j$; $\alpha(\varphi, l'_i, l'_j) = y$; $P'(y)=z$ формирует вектор z, отмечающий единицами элементы исходного массива A_i, входящие в A_j ($i,j = \overline{1,k}$; $i \neq j$).

8. Пересечение k массивов. Пусть $A = (A_1, \ldots, A_i, \ldots, A_k)$ и $L = (l_1, \ldots, l_i, \ldots, l_k)$. Пересечением k массивов $\bigcap_{i=1}^{k} A_i = A_0$ будем называть массив A_0, состоящий (для определенности) из элементов A_1, входящих во все массивы A_i ($i = \overline{2,k}$), не содержащий дублирующих элементов. Последовательность операций $\mathcal{U}[A, L] \rightarrow A', L', \varphi$; $y_1 = \alpha(\varphi, l'_1, l'_2)$; ...; $y_i = \alpha(\varphi, y_{i-1}, l'_{i+1})$; ...; $y_{k-1} = \alpha(\varphi, y_{k-2}, l'_k)$; $\gamma(\varphi, y_{k-1}) = x$ (где вектор y_i единицами отмечает элементы A_1 из A', входящие в массивы $A_2, A_3, \ldots, A_i, A_{i+1}$) формирует вектор x, отмечающий единицами элементы A_1 из A', составляющие массив $A_0 = \bigcap_{i=1}^{k} A_i$. Аналогично последовательность операций $\mathcal{B}[A] \rightarrow P, \varphi$; $P(l_1) = l'_1$; $P(l_2) = l'_2$; $y_1 = \alpha(\varphi, l'_1, l'_2)$; ...; $P(l_{i+1}) = l'_{i+1}$; $y_i = \alpha(\varphi, y_{i-1}, l'_{i+1})$; ...; $P(l_k) = l'_k$; $y_k = \alpha(\varphi, y_{k-2}, l'_k)$; $y(\varphi, y_{k-1}) = x$; $P_1(x) = z$ формирует вектор z, отмечающий единицами элементы исходного массива A_1, составляющие массив $A_0 = \bigcap_{i=1}^{k} A_i$.

9. Объединение выбранных g массивов. Пусть $A = (A_1, \ldots, A_i, \ldots, A_k)$ и $L = (l_1, \ldots, l_i, \ldots, l_k)$. Нужно выполнить объединение g массивов $A_{i_1}, A_{i_2}, \ldots, A_{i_j}, \ldots, A_{i_g}$ из A ($g \leq k$). Последовательность операций $\mathcal{U}[A\ L] \rightarrow A', L', \varphi$; $l = \bigvee_{j=1}^{g} l'_{i_j}$; $\gamma(\varphi, l) = x$ формирует вектор x, отмечающий единицами элементы из A', составляющие массив $A_0 = \bigcup_{j=1}^{g} A_{i_j}$, не содержащий дублирующих элементов. В то же время последовательность операций $\mathcal{B}[A] \rightarrow P, \varphi$; $l = \bigvee_{j=1}^{g} l_{i_j}$; $P(l) = l'$; $\gamma(\varphi, l') = x$; $P'(x) = z$ формирует вектор z, отмечающий единицами элементы из A, составляющие массив $A_0 = \bigcup_{j=1}^{g} A_{i_j}$.

10. Пусть $A = (A_0, A_1, \ldots, A_i, \ldots, A_k)$ и $L = (l_0, l_1, \ldots, l_i, \ldots, l_k)$ и необходимо определить, какие элементы каждого из мас-

сивов A_i $(i=\overline{1,k})$ входят в ключевой массив A_0. В последовательности операций $\mathcal{U}[A,L] \to A', L', \varphi;\ y_1 = \alpha(\varphi, l_1', l_0');\ldots;y_i = \alpha(\varphi, l_i', l_0');\ldots;y_k = \alpha(\varphi, l_k', l_0')$ вектор y_i отмечает единицами элементы A_i из A', входящие в A_0 $(i=\overline{1,k})$. Аналогично в последовательности операций $\mathcal{B}[A] \to P, \varphi;\ P(L) = L';\ y_1 = \alpha(\varphi, l_1', l_0');\ldots;y_i=\alpha(\varphi, l_i', l_0');\ldots;y_k = \alpha(\varphi, l_k', l_0');\ P'(y_1) = z_1;\ldots;P'(y_i);\ldots;P'(y_k) = z_k$ вектор z_i отмечает единицами элементы исходного массива A_i, входящие в A_0 $(i = \overline{1,k})$.

II. Пусть $A = (A_0, A_1, \ldots, A_i, \ldots, A_k)$ и $L = (l_0, l_1, \ldots, l_i, \ldots \ldots, l_k)$. Необходимо определить, в какие массивы $A_i (i=\overline{1,k})$ ключевой массив A_0 входит полностью. В последовательности операций $\mathcal{U}[A,L] \to A', L', \varphi; \beta_1 = \beta(\varphi, l_0', l_1');\ldots;\beta_i = \beta(\varphi, l_0', l_i');\ldots,\beta_k = \beta(\varphi, l_0', l_k')$ значение $\beta_i = 1$, если A_0 входит в A_i, и $\beta=0$ – в противном случае $(i=\overline{1,k})$. Аналогично в последовательности операций $\mathcal{B}[A] \to P, \varphi;\ P(L) = L';\ \beta_1 = \beta(\varphi, l_0', l_1');\ldots;\beta_i(\varphi, l_0', l_i'),\ldots,\beta_k = \beta(\varphi, l_0', l_k')$ значение $\beta_i = 1$, если A_0 входит в A_i $(i = \overline{1,k})$.

Проиллюстрируем возможности предлагаемого метода на следующем примере. Требуется определить, входит ли массив A_0 в массив $B = (A_1 \cap A_2) \cup (A_1 \cap A_3) \cup (A_3 \cap A_4 \cap A_2)$. Пусть $A = (A_0, A_1, A_2, A_3, A_4)$, $L = (l_0, l_1, l_2, l_3, l_4)$. Тогда последовательность операций $\mathcal{U}[A,L] \to A', L', \varphi;\ y_1 = \alpha(\varphi, l_1', l_2');\ y_2 = \alpha(\varphi, l_1', l_3);\ y_3 = \alpha(\varphi, l_3', l_4');\ y_4 = \alpha(\varphi, y_3, l_2');\ l = y_1 \vee y_2 \vee y_4;\ \beta = \beta(\varphi, l_0', l)$ формирует $\beta = 1$, если A_0 входит в B, и $\beta=0$ – в противном случае. Аналогично последовательность операций $\mathcal{B}[A] \to P, \varphi;\ P(L) = L';\ y_1 = \alpha(\varphi, l_1', l_2');\ y_2 = \alpha(\varphi, l_1', l_3');\ y_3 = \alpha(\varphi, l_3', l_4');\ y_4=\alpha(\varphi, y_3, l_2');\ l = y_1 \vee y_2 \vee y_4;\ \beta = \beta(\varphi, l_0', l)$ формирует $\beta=1$, если A_0 входит в B.

Таким образом, было показано, что решение достаточно сложных задач обработки массивов сводится к упорядочению единого массива, состоящего из массивов-аргументов ($\mathcal{U}$), либо к построению упорядочивающей перестановки для единого массива ($\mathcal{B}$), после чего над полученными двоичными векторами меток выполняются простые логические ($\wedge, \vee, \neg$) и специальные ($P, P', \alpha, \beta, \gamma$) операции. Рассмотренный метод позволяет также реализовать любую функцию обработки k массивов, заданную, например, логической формулой в дизъюнктивно нормальной форме: $B = \cup\cap A_i^*$ (где A_i^* – некоторый i-й массив A_i $(i=\overline{1,k})$ или его дополнение $\overline{A}_i$ для области $X = \bigcup_{i=1}^{k} A_i$).

Поскольку логические и специальные операции могут быть выполнены очень быстро, время решения информационно-логической задачи в основном определяется временем упорядочения единого массива.

Такая последовательность (перегруппировка меток в соответствии с упорядочением - логическая обработка меток) позволяет выполнять широкий круг информационно-логических операций поискового типа над массивами, в том числе операции реляционной алгебры Кодда.

Далее рассмотрим способы выполнения операций реляционной алгебры на основе предлагаемого метода.

В реляционной модели баз данных все данные представляются в виде набора поименованных таблиц (отношений) [4]. Для формирования запросов к реляционной базе Коддом было предложено два вида языков: реляционная алгебра (α-алгебра) и реляционное исчисление [5]. Язык α-алгебры является процедурным языком и представляет собой набор операций, аргументами которых являются массивы [5,6]. Любую информацию, хранимую в реляционной базе, можно извлечь путем выполнения некоторой последовательности операций α-алгебры.

Операции α-алгебры делятся на теоретико-множественные (объединение, пересечение, разность, расширенное декартово произведение) и специальные (проекция, ограничение, деление, соединение). Выполнение объединения, пересечения и разности было рассмотрено выше. Расширенное декартово произведение здесь не рассматривается, так как оно не относится к поисковым операциям. При реализации проекции основная работа - исключение из результирующего массива дублирующих кортежей (рассмотрено выше). Выполнение ограничения сводится к простому покомпонентному сравнению двух массивов.

Итак, рассмотрим здесь выполнение операций деления и соединения на основе предлагаемого метода. Далее будем использовать все обозначения и понятия, которые определены выше.

Введем определение. О п е р а ц и е й $\mu(\varphi, l) = g$ назовем процедуру, которая по некоторому двоичному вектору l и разделительному вектору φ, содержащему s единиц и разделяющему l на s сегментов, строит вектор g такой, что содержимое r-го сегмента $g(r = \overline{1,s})$ совпадает с содержимым r-го сегмента l, если r-й сегмент l - первый от начала сегмент, содержащий хотя бы одну единицу, во всех других сегментах g содержатся только нули.

Д е л е н и е*) (I-й вариант). Пусть $R_1(W_1,V_1)$ и $R_2(V_2)$ – отношения (массивы) размерностей n_1 и n_2 соответственно, у которых домены V_1 и V_2 совместимы по объединению. Нужно получить результирующий массив операции деления $R = R_1[V_1 \div V_2]R_2$. (В примере на рис.5 необходимо определить фамилии тех людей, которые владеют языками АЛГОЛ, КОБОЛ и ФОРТРАН.)

Применяя к $R_1(W_1,V_1)$ операцию упорядочения $\mathcal{U}[R_1(\underline{W}_1,V_1)] \to R_1'(\underline{W}_1',V_1'),\varphi_1$, получаем вектор φ_1 размерности n_1, содержащий s_1 единиц, который разделяет упорядоченный по домену $\underline{W}_1'$ массив $R_1'(W_1', V_1')$ на s_1 подмассивов (рис.5). Элементы домена $\underline{W}_1'$ в r-м подмассиве одинаковы ($r = \overline{1,s_1}$). Построим единый массив $A(W,\underline{V}) = (R_1'(\underline{W}_1',V_1'),R_2(\emptyset, V_2))$ размерности $n = n_1+n_2$ и вектор меток l размерности n, отмечающий единицами кортежи массива R_2. Применяя к $A(W,\underline{V})$ операцию $\mathcal{B}[A(W,\underline{V})] \to P, \varphi_2$, получаем вектор φ_2 размерности n (содержащий s_2 единиц) и упорядочивающий вектор P размерности n. На рис.5 показан граф соединений, отображающий перестановку, соответствующую вектору P, а также вектор l', полученный в результате перестановки $P(l) = L'$.

После этих операций остается проверить полное вхождение массива R_2 в каждый r-й ($r=\overline{1,s_1}$) подмассив элементов домена V_1' массива R_1'. Эта проверка производится путем выполнения s_1 итераций. В r-й итерации выполняются следующие операции над векторами:

$$z_r=z_{r-1}\cdot\overline{g}_{r-1};\ g_r=\mu(\varphi_1,z_r);\ g_r^*=g_r\smile\underline{O}(n_2);\ g_r'=P(g_r^*);\ \beta_r=\beta(\varphi_2,l',g_r');$$

$$f_r = \begin{cases} f_{r-1} \vee g_r, \text{если } \beta_r=1, \\ f_{r-1}, \text{ если } \beta_r=0, \end{cases}$$

где $r=\overline{1,s_1}$; $z_0 = \underline{1}$; $g_0 = \underline{O}$; $f_0 = \underline{O}$ ($\underline{I}$ и $\underline{O}$ – единичный и нулевой векторы соответственно). Здесь вектор $z_r = z_{r-1}\cdot\overline{g}_{r-1}$ в сегментах с номерами меньше чем r содержит только нули, а содержимое остальных его сегментов совпадает с содержимым соответствующих сегментов вектора z_0. Вектор $g_r = \mu(\varphi_1,z_r)$ выделяет единицами все кортежи r-го подмассива массива R_1'. Вектор $g_r^* = g_r\smile\underline{O}(n_2)$ размерности n есть сцепление вектора g_r размерности n_1 с вектором $\underline{O}$ размерности n_2. Переменная $\beta_r= (\beta_2,l',g_r') = 1$, если массив R_2 (эле-

*) Определение деления см. в [6].

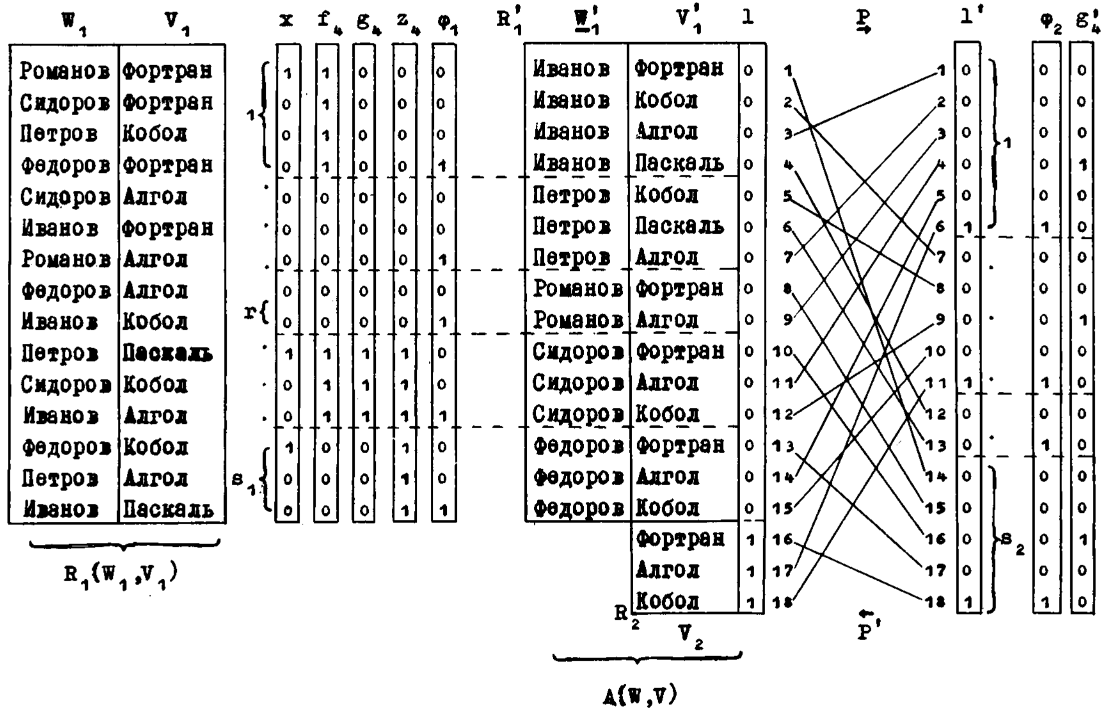

Рис.5. Выполнение операции деления (I-й вариант).

менты которого отмечены единицами 1) полностью входит в r-й подмассив элементов домена V_1' (отмеченных единицами g_r) массива $R_1'(\underline{W}_1', V_1')$, и $\beta_r=0$ – в противном случае. Если $\beta_r=1$, то элементы домена $\underline{W}_1'$ r-го подмассива (отмеченные единицами вектора g_r) массива $R_1'(\underline{W}_1', V_1')$ (все эти элементы одинаковы) входят в результирующий массив R. Вектор f_r единицами отмечает все элементы тех подмассивов домена $\underline{W}_1'$ массива $R_1'(\underline{W}_1', V_1')$, которые уже входят в результирующий массив R после выполнения r-й итерации.

За последней (s_1-й) итерацией следует заключительная операция $\gamma(\varphi_1, f_{s_1}) = x$, формирующая вектор x, отмечающий единицами элементы домена $\underline{W}_1'$ массива R_1', составляющие массив $R=R_1[V_1+V_2]R_2$, не содержащий дублирующих элементов. На рис.5 приведены значения векторов z_4, g_4, g_4' и f_4 после выполнения 4-й итерации (для этой итерации $\beta_4=1$), а также значение результирующего вектора x (отмечающего единицами фамилии людей, которые владеют языками АЛГОЛ, КОБОЛ и ФОРТРАН).

Д е л е н и е (2-й вариант)[ж]. Применяя к $R_1(W_1, V_1)$ операцию $\mathcal{B}[R_1(\underline{W}_1, V_1)] \to P_1, \varphi_1$, получаем вектор φ_1 размерности n_1 (содержащий s_1 единиц) и упорядочивающий вектор P_1 размерности n_1. Построим единый массив $A(W,\underline{V}) = (R_1(W_1,V_1), R_2(\emptyset, V_2))$ размерности $n = n_1+n_2$ и вектор меток 1 размерности n, отмечающий единицами кортежи массива R_2 (рис.6). Применяя к $A(W,\underline{V})$ операцию $\mathcal{B}[A(W,\underline{V})] \to P_2, \varphi_2$, получаем вектор φ_2 размерности n (содержащий s_2 единиц) и упорядочивающий вектор P_2 размерности n. На рис. 6 показаны графы соединений, отображающие перестановки, соответствующие векторам P_1 и P_2, а также вектор $1'$, полученный в результате перестановки $P_2(1) = 1'$.

Затем, так же как в I-м варианте, выполняются s_1 итераций. В r-й итерации реализуются следующие операции над векторами:

$$z_r=z_{r-1}\cdot\bar{g}_{r-1};\ g_r=\mu(\varphi_1, z_r);\ g_r'=P_1'(g_r);\ g_r^*=g_r' \smile \underline{0}(n_2);\ g_r''=P_2(g_r^*);$$

$$\beta_r=\beta(\varphi_2, 1', g_r'');\quad f_r=\begin{cases} f_{r-1} \vee g_r, & \text{если } \beta_r=1, \\ f_{r-1}, & \text{если } \beta_r=0, \end{cases}$$

[ж] Отличие I-го и 2-го вариантов состоит в том, что в последнем кортежи исходных массивов остаются на прежних местах.

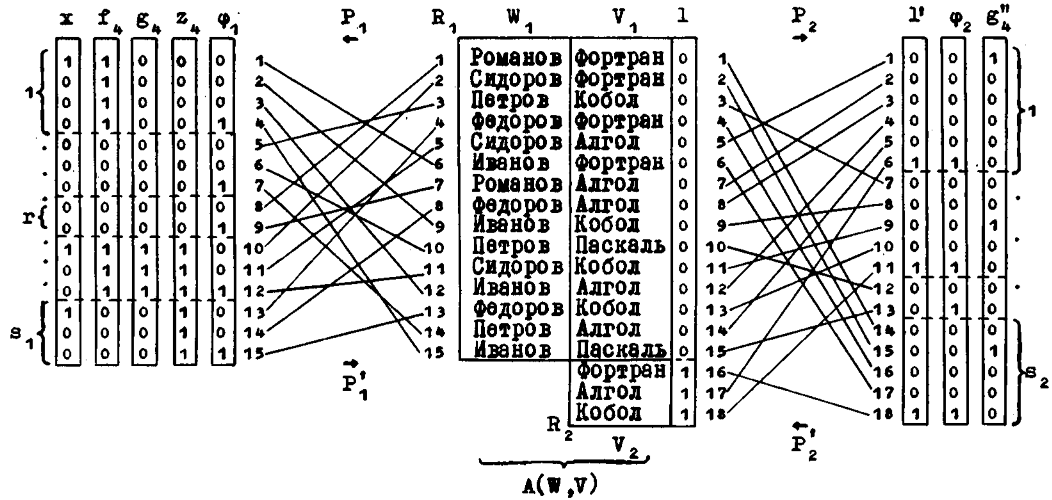

Рис.6. Выполнение операции деления (2-й вариант).

где $r=\overline{1,s_1}$; $z_0 = \underline{1}$; $g_0 = \underline{0}$; $f_0 = \underline{0}$. Здесь вектор $g_r=\mu(\varphi_1,z_r)$ в r-м сегменте содержит только единицы и во всех остальных сегментах – только нули, а вектор $g_r^! = P_1^!(g_r)$ отмечает единицами кортежи исходного массива R_1, которые содержат одинаковые элементы домена W_1 и в совокупности составляют r-й подмассив массива R_1. Переменная $\beta_r = \beta(\varphi_2, l', g_r'')=1$, если массив $R_2(V_2)$ входит в r-й подмассив элементов домена V_1 (отмеченных единицами $g_r^!$) массива $R_1(W_1,V_1)$, и $\beta_r=0$ – в противном случае. Если $\beta_r=1$, то элементы домена W_1 r-го подмассива (отмеченные единицами $g_r^!$) массива R_1 (все эти элементы одинаковы) входят в результирующий массив R.

За последней (s_1-й) итерацией следуют заключительные операции: $x = \gamma(\varphi_1, f_{s_1})$ и $P_1^!(x) = z$. Вектор z отмечает единицами элементы домена W_1 массива $R_1(W_1,V_1)$, составляющие массив $R = R_1[V_1 \div V_2]R_2$, не содержащий дублирующих элементов. На рис.6 приведены значения векторов z_4, g_4, g_4'' и f_4 после выполнения 4-й итерации (для этой итерации $\beta_4 = 1$), а также значение вектора x.

Таким образом, реализация деления (после операций $\mathcal{U}$ или $\mathcal{B}$) сводится к выполнению определенной последовательности операций над двоичными векторами.

Далее рассмотрим выполнение операции соединения[ж)] $R = R_1[V_1 \theta V_2]R_2$ при различных критериях сравнения θ.

Пусть $R_1(W_1, V_1)$ и $R_2(W_2,V_2)$ – массивы размерностей n_1 и n_2 соответственно, у которых домены V_1 и V_2 совместимы по объединению, $A(W,\underline{V}) = (R_1(W_1,V_1), R_2(W_2,V_2))$ – единый массив размерности $n = n_1+n_2$ и l – вектор меток размерности n, отмечающий единицами кортежи массива $R_1(W_1,V_1)$ (рис.7). Нужно получить результирующий массив R операции соединения при $\theta \in \{=, \neq, >, \geq, <, \leq\}$.

С о е д и н е н и е $R = R_1[V_1 = V_2]R_2$. Применяя к $A(W,\underline{V})$ операцию $\mathcal{U}[A(W,\underline{V}), l] \to A'(W',\underline{V}'), l', \varphi$, получаем вектор φ размерности n, содержащий s единиц, который разделяет упорядоченный по домену $\underline{V}$ массив A' на s подмассивов, а вектор l' – на s сегментов. Каждому r-му подмассиву соответствует r-й сегмент l' ($r=\overline{1,s}$) (рис.7). Кортежи r-го подмассива A', отмеченные единицами l' (нулями l'), составляют r-й подмассив R_1^r (R_2^r) массива $R_1(R_2)$.

ж) Определение соединения см. в [6].

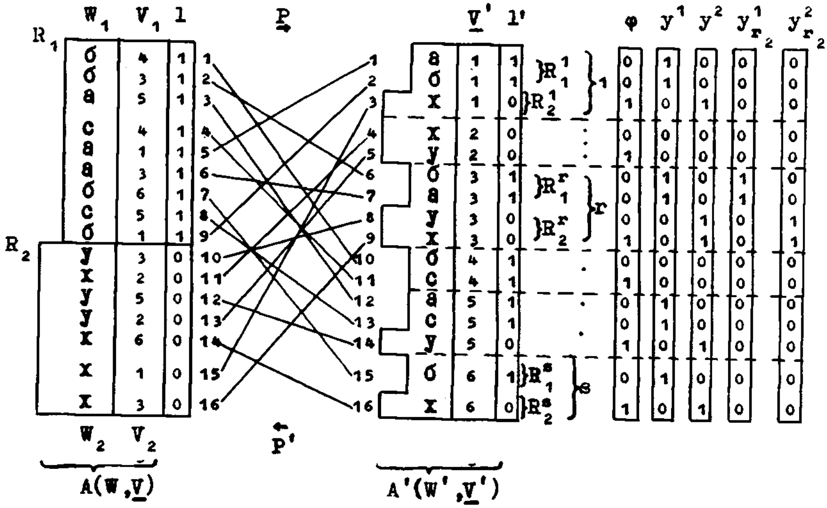

Рис. 7

Если r-й сегмент $1'$ содержит только единицы (нули), то r-й подмассив массива A' содержит только кортежи из R_1(R_2), отмеченные единицами $1'$ (нулями $1'$). В обоих случаях эти кортежи не участвуют в формировании результата соединения (рис.7). Если r-й сегмент $1'$ содержит и единицы и нули, то r-й подмассив массива A' содержит кортежи из R_1 и R_2. В таком случае эти кортежи удовлетворяют условию соединения ($V_1 = V_2$) и декартово произведение r-х подмассивов R_1^r и R_2^r образует r-й подмассив $R^r = R_1^r \otimes R_2^r$ результирующего массива R (где $\otimes$ - операция расширенного декартова произведения). Результат соединения $R_1[V_1=V_2]R_2=R=\bigcup_{r=1}^{s} R^r$.

Ниже приведена последовательность операций для реализации соединения $R_1[V_1 = V_2]R_2$. Сначала выполняются вспомогательные операции $y^1 = \hat{\alpha}(\varphi,1')$ и $y^2 = \hat{\alpha}(\varphi,\bar{1}')$. Вектор y^1 отмечает единицами кортежи массива R_1 из A', а вектор y^2 - кортежи массива R_2 из A', участвующие в соединении. После этого реализация соединения сводится к выполнению k итерации (где k - количество сегментов вектора $1'$, содержащих и нули, и единицы). В i-й итерации выполняются следующие операции:

$$z^1_{r_i}=z^1_{r_{i-1}}\cdot\bar{y}^1_{r_{i-1}};\ y^1_{r_i}=\mu(\varphi,z^1_{r_i});\ z^2_{r_i}=z^2_{r_{i-1}}\cdot\bar{y}^2_{r_{i-1}};\ y^2_{r_i}=\mu(\varphi,z^2_{r_i});$$

$$R^i = R_1^{r_i} \otimes R_2^{r_i},$$

где $i=\overline{1,k}$; $z^1_{r_0}=y^1$; $y^1_{r_0}=\underline{0}$; $z^2_{r_0}=y^2$; $y^2_{r_0}=\underline{0}$. Здесь $r_1,\ldots,r_i,\ldots$ $\ldots,r_k$ – номера сегментов вектора $1'$ (в порядке возрастания), в которых содержатся и единицы и нули. Содержимое r_i-го сегмента вектора y_{r_i} ($y^2_{r_i}$) совпадает с содержимым r_i-го сегмента вектора $y^1(y^2)$, во всех остальных сегментах вектора $y^1_{r_i}$ ($y^2_{r_i}$) содержатся нули. Таким образом, вектор $y^1_{r_i}$ ($y^2_{r_i}$) в i-й итерации выделяет единицами те кортежи $R_1(R_2)$, находящиеся в r_i-м подмассиве A', которые составляют r_i-й подмассив $R_1^{r_i}(R_2^{r_i})$, участвующий в декартовом произведении $R^i = R_1^{r_i} \otimes R_2^{r_i}$ $(i=\overline{1,k})$. Совокупность полученных подмассивов $R^i(i=\overline{1,k})$ составляет массив $R=\bigcup_{i=1}^{k} R^i = R_1[V_1=V_2]R_2$. (На рис.7 приведены значения векторов $y^1_{r_i}$ и $y^2_{r_i}$ для 2-й итерации.)

Соединение $R_1[V_1 = V_2]R_2$ можно выполнить также, используя операцию $\mathcal{B}$. При этом кортежи исходных массивов R_1 и R_2 остаются на прежних местах. После операций $\mathcal{B}[A(W,\underline{V})] \to P,\varphi$; $1' = P(1)$; $y^1=\alpha(\varphi,1')$; $y^2 = \alpha(\varphi,\overline{1'})$ реализация соединения сводится к выполнению k итераций. В i-й итерации выполняются следующие операции:

$$z^1_{r_i}=z^1_{r_{i-1}}\cdot\overline{y}^1_{r_{i-1}};\ y^1_{r_i}=\mu(\varphi,z^1_{r_i});\ z^2_{r_i}=z^2_{r_{i-1}}\cdot\overline{y}^2_{r_{i-1}};\ y^2_{r_i}=\mu(\varphi,z^2_{r_i});$$

$$x^1_{r_i}=P'(y^1_{r_i});\ x^2_{r_i}=P'(y^2_{r_i});\ R^i = R_1^{r_i}\otimes R_2^{r_i},$$

где $i=\overline{1,k}$; $z^1_{r_0}=y^1$; $z^2_{r_0}=y^2$; $y^2_{r_0}=y^1_{r_0}=\underline{0}$. Здесь вектор $x^1_{r_i}= P'(y^1_{r_i})$ ($x^2_{r_i}=P'(y^2_{r_i})$) выделяет единицами те кортежи исходного массива $R_1(R_2)$, которые составляют r_i-й подмассив $R_1^{r_i}(R_2^{r_i})$, участвующий в декартовом произведении $R^i = R_1^{r_i} \otimes R_2^{r_i}$. Результат $R = \bigcup_{i=1}^{k} R^i$.

С о е д и н е н и е $R = R_1[V_1 < V_2]R_2$. Введем определения. О п е р а ц и е й $\tau(\varphi,l) = d$ (о п е р а ц и е й $\tau'(\varphi,l) = d$) назовем процедуру, которая по некоторому двоичному вектору l и разделительному вектору φ, содержащему s единиц и разделяющему вектор l на s сегментов, строит вектор d такой, что: а) если I-й, 2-й,...,(k-I)-й сегменты вектора l содержат только нули, а k-й сегмент, $k = \overline{1,s-1}$ ($k = \overline{1,s}$), содержит хотя бы одну единицу, то содержимое первых k ($k-1$) сегментов вектора d совпадает с содержимым первых $k(k-1)$ сегментов вектора l; б) если, начиная с некоторого g-го сегмента, $g=\overline{2,s}$ ($g = \overline{1,s}$), содержащего хотя бы один нуль, в векторе l сегменты $(g+1)$-й,...,s-й содержат только единицы, то содержимое последних $s-g+1$ ($s-g$) сегментов вектора d совпадает с инверсией содержимого соответствующих сегментов вектора l; в) если при этом $g-k > 1$ ($g-k \geq 0$), то $(k+1)$-й,..., $(g-1)$-й (k-й,$(k+1)$-й,...,g-й) сегменты вектора d содержат только единицы; г) если $k = g$, или $k = g+1$, или $k = s$, или $g=1$ (если $k=g+1$), то все сегменты вектора d содержат только нули.

Из определений операции τ и операции соединения $R_1[V_1<V_2]R_2$ следует, что после применения к $A(W,\underline{V})$ операции $\mathcal{U}[A(W,\underline{V}),l] \to A'(W',\underline{V}'),l',\varphi$ вектор $d = \tau(\varphi,l')$ отмечает единицами те кортежи массивов R_1 и R_2 из A', которые участвуют в соединении, так как удовлетворяют условию соединения ($V_1 < V_2$) (рис.8). При этом вектор $y'=l'\cdot d$ отмечает единицами кортежи массива R_1 из A', а вектор $y^2 = \overline{l'}\cdot d$ – кортежи массива R_2 из A', участвующие в соединении. Пусть $r_1,\ldots,r_j,\ldots,r_p$ ($u_1,\ldots,u_i,\ldots,u_q$) – номера тех сегментов (в порядке возрастания), каждый из которых содержит хотя бы одну единицу вектора $y^1(y^2)$. Кортежи массива R_1, находящиеся в r_j-м подмассиве A', соединяются с кортежами массива R_2, находящимися в тех подмассивах u_i ($i=\overline{1,q}$) массива A', для которых $u_i > r_j$, так как (в силу упорядоченности массива A' по домену $\underline{V}'$) для этих кортежей выполняется условие соединения ($V_1<V_2$). Поэтому декартово произведение кортежей массива R_1 из A', отмеченных единицами r_j-го сегмента y^1, с кортежами массива R_2 из A', отмеченных единицами тех сегментов u_i ($i=\overline{1,q}$) вектора y^2, для которых $u_i > r_j$, составляет j-й подмассив R^j результирующе-

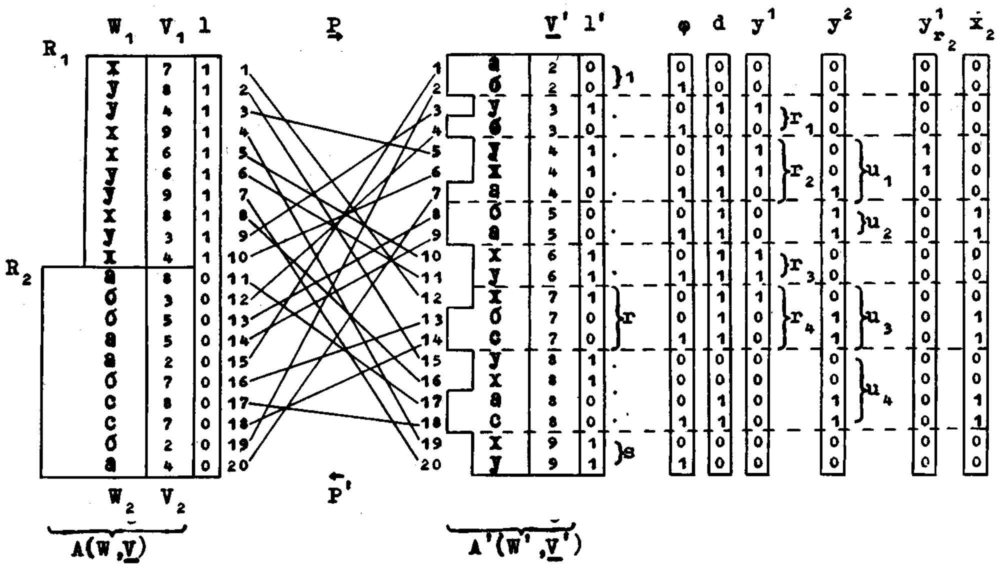

Рис.8. Выполнение операции соединения $R_1[V_1 < V_2]R_2$.

го массива $R(j = \overline{1,p})$. Кортежи r_j-го (u_i-го) подмассива A',отмеченные единицами r_j-го (u_i-го) сегмента y^1 (y^2), составляют r_j-й (u_i-й) подмассив $R_1^{r_j}$ $(R_2^{u_i})$ массива R_1 (R_2). Тогда

$$R^j = R_1^{r_j} \otimes \bigcup_{u_i > r_j} R_2^{u_i} = \bigcup_{u_i > r_j} (R_1^{r_j} \otimes R_2^{u_i}).$$

Результат операции соединения $R_1[V_1 < V_2]R_2 = R = \bigcup_{j=1}^{p} R^j$.

Введем необходимые определения. Пусть вектор φ, содержащий s единиц, разделяет некоторый двоичный вектор ω на s сегментов, при этом вектор ω такой, что в каком-то одном r-м сегменте он содержит хотя бы одну единицу, а во всех остальных - нули $(r = \overline{1,s})$. О п е р а ц и е й $\pi(\varphi,\omega) = t$ (или $\pi'(\varphi,\omega) = t$) назовем процедуру, которая по векторам φ и ω строит вектор t такой, что в его первых $r(r-1)$ сегментах содержатся только нули, а во всех остальных - только единицы $(r = \overline{1,s})$. О п е р а ц и е й $\nu(\varphi,\omega) = v$ назовем процедуру, которая по векторам φ и ω строит вектор v такой, что в его r-м сегменте содержатся только нули, а во всех остальных - только единицы $(r = \overline{1,s})$.

Теперь приведем последовательность операций для реализации соединения $R_1[V_1 < V_2]R_2$. После операции $\mathcal{U}\ A(W,\underline{V}),1] \to A'(W',\underline{V}')$, $1'$,φ выполняются вспомогательные операции $d = \tau(\varphi,1')$; $y^1 = 1' \cdot d$; $y^2 = \overline{1}' \cdot d$. После этого реализация соединения сводится к выполнению p итераций (где p - количество сегментов y^1, содержащих хотя бы одну единицу). В j-й итерации выполняются следующие операции:

$$z_{r_j} = z_{r_{j-1}} \cdot \overline{y^1_{r_{j-1}}};\quad y^1_{r_j} = \mu(\varphi,z_{r_j});\quad t_j = \pi(\varphi,y^1_{r_j});\quad x_j = t_j \cdot y^2;$$

$$R^j = R_1^{r_j} \otimes \bigcup_{u_i > r_j} R_2^{u_i},$$

где $j = \overline{1,p}$; $z_{r_0} = y^1$; $y^1_{r_0} = \underline{0}$. Здесь вектор $y^1_{r_j}$ отмечает единицами кортежи подмассива $R_1^{r_j}$ из A', а вектор x_j - кортежи массива R_2, находящиеся в подмассивах u_i $(i = \overline{1,q})$ массива A',для которых $u_i > r_j$, и составляющие массив $\bigcup_{u_i > r_j} R_2^{u_i}$. Результат $R =$

$= \bigcup_{j=1}^{p} R^j$. (На рис.8 приведены значения векторов $y^1_{r_j}$ и x_j для 2-й итерации.)

Соединение $R_1[V_1 < V_2]R_2$ можно выполнить, также используя операцию $\mathcal{B}$. Кортежи исходных массивов R_1 и R_2 остаются при этом на прежних местах. После операций $\mathcal{B}[A(W,\underline{V})] \to P, \varphi; 1' = P(1)$; $d = \tau(\varphi, 1')$; $y^1 = 1' \cdot d$; $y^2 = \overline{1'} \cdot d$ реализация соединения сводится к выполнению p итераций. В j-й итерации выполняются следующие операции:

$$z_{r_j} = z_{r_{j-1}} \cdot \overline{y^1_{r_{j-1}}};\ y^1_{r_j} = \mu(\varphi, z_{r_j});\ t_j = \pi(\varphi, y^1_{r_j});\ x_j = t_j \cdot y^2;$$

$$g_{r_j} = P'(y^1_{r_j});\ x'_j = P'(x_j);\ R^j = R_1^{r_j} \otimes \bigcup_{u_i > r_j} R_2^{u_i},$$

где $j = \overline{1,p}$; $z_{r_0} = y^1$; $y^1_{r_0} = 0$. Здесь вектор $g_{r_j} = P'(y^1_{r_j})$ отмечает единицами кортежи подмассива $R_1^{r_j}$ из исходного массива R_1, а вектор $x'_j = P'(x_j)$ – кортежи из исходного массива R_2, составляющие массив $\bigcup_{u_i > r_j} R_2^{u_i}$. Результат $R = \bigcup_{j=1}^{p} R^j$.

С о е д и н е н и е $R = R_2[V_1 \le V_2]R_2$. Последовательности операций для реализации этого соединения аналогичны последовательностям операций для соединения $R_1[V_1 < V_2]R_2$ с той разницей, что вместо операции $d = \tau(\varphi, 1')$ используется операция $d = \tau'(\varphi, 1')$, вместо $t_j = \pi(\varphi, y^1_{r_j})$ используется $t_j = \pi'(\varphi, y^1_{r_j})$, а декартово произведение $R^j = R_1^{r_j} \otimes \bigcup_{u_i > r_j} R_2^{u_i}$ заменяется на $R^j = R_1^{r_j} \otimes \bigcup_{u_i \ge r_j} R_2^{u_i}$.

С о е д и н е н и е $R = R_1[V_1 > V_2]R_2$. Реализация этого соединения во многом аналогична выполнению соединения $R_1[V_1 < V_2]R_2$. При этом $d = \tau(\varphi, \overline{1'})$ (где $\overline{1'}$ – инверсия вектора $1'$), $y^1 = 1' \cdot d$, $y^2 = \overline{1'} \cdot d$, а $r_1, \dots, r_j, \dots, r_p$ и $u_1, \dots, u_i, \dots, u_q$ – те же обозначения сегментов для полученных векторов y^1 и y^2 соответственно. В этом случае декартово произведение кортежей массива R_2, находящихся в u_i-м подмассиве A', с кортежами массива R_1, находящимися в подмассивах r_j $(j = \overline{1,p})$ массива A', для которых $r_j > u_i$, составляет i-й подмассив R^i результирующего массива R $(i = \overline{1,q})$, так как для этих кортежей выполняется условие соединения $(V_1 > V_2)$. То есть $R^j = (\bigcup_{r_j > u_i} R_1^{r_j}) \otimes R_2^{u_i} = \bigcup_{r_j > u_i} (R_1^{r_j} \otimes R_2^{u_i})$.

Результат соединения $R_1[V_1 > V_2]R_2 = R = \bigcup_{i=1}^{q} R^i$. Таким образом, после операций $\mathcal{U}[A(W,\underline{V})] \to A'(W',\underline{V}'),1',\varphi$; $d = \tau(\varphi,\overline{1}')$; $y^1 = 1' \cdot d$; $y^2 = \overline{1}' \cdot d$ реализация этого соединения сводится к выполнению q итераций (где q – количество сегментов вектора y^2, содержащих хотя бы одну единицу). В i-й итерации выполняются следующие операции:

$$z_{u_1} = z_{u_{i-1}} \cdot \overline{y^2_{u_{i-1}}};\ y^2_{u_i} = \mu(\varphi, z_{u_i});\ t_i = \pi(\varphi, y^2_{u_i});\ x_i = t_i \cdot y^1;$$

$$R^i = (\bigcup_{r_j > u_i} R_1^{r_1}) \otimes R_2^{u_i},$$

где $i = \overline{1,q}$; $z_{u_0} = y^2$; $y^2_{u_0} = \underline{0}$. Здесь вектор $y^2_{u_i}$ отмечает единицами кортежи подмассива $R_2^{u_i}$ из A', а вектор x_i – кортежи массива R_2 из A', составляющих массив $\bigcup_{r_j > u_i} R_1^{r_j}$. Результат $R = \bigcup_{i=1}^{q} R^i$.

Это соединение можно выполнить, также используя операцию $\mathcal{B}$, при этом кортежи R_1 и R_2 остаются на прежних местах.

С о е д и н е н и е $R = R_1[V_1 \geq V_2]R_2$. Выполнение такого соединения аналогично реализации соединения $R_1[V_1 > V_2]R_2$ с той разницей, что вместо операции $d = \tau(\varphi,\overline{1}')$ используется операция $t_i = \pi'(\varphi,\overline{1}')$, вместо $t_i = \pi(\varphi, y^2_{u_i})$ используется $t_i = \pi'(\varphi, y^2_{u_i})$, а декартово произведение $R^i = (\bigcup_{r_j > u_i} R_1^{r_j}) \times R_2^{u_i}$ заменяется на $R^i = (\bigcup_{r_j \geq u_i} R_1^{r_j}) \otimes R_2^{u_i}$.

С о е д и н е н и е $R = R_1[V_1 \neq V_2]R_2$. Пусть после выполнения операции $\mathcal{U}[A(W,\underline{V}),1] \to A'(W',\underline{V}'),1',\varphi$ $r_1,\dots,r_j,\dots,r_p(u_1,\dots,u_1,\dots,u_q)$ – номера тех сегментов в порядке возрастания, каждый из которых содержит хотя бы одну единицу вектора $1'$ $(\overline{1}')$. В этом случае декартово произведение кортежей массива R_1, находящихся в r_j-м подмассиве A', с кортежами массива R_2, находящимися ся в подмассивах u_i $(i=\overline{1,q})$ массива A', для которых $u_i \neq r_j$, составляет j-й подмассив R^j результирующего массива R $(j = \overline{1,p})$, так как для этих кортежей выполняется условие соединения $(V_1 \neq V_2)$. Кортежи r_j-го (u_i-го) подмассива $A'(W', \underline{V}')$, отмеченные единицами r_j-го (u_i-го) сегмента $1'(\overline{1'})$, составляют r_j-й (u_i-й) подмассив $R_1^{r_j}(R_2^{u_i})$ массива $R_1(R_2)$. Тогда $R^j = R_1^{r_j} \otimes \bigcup_{u_i \neq r_j} R_2^{u_i} = \bigcup_{u_i \neq r_j} (R_1^{r_j} \otimes R_2^{u_i})$. Результат соединения $R_1[V_1 \neq V_2]R_2 = R = \bigcup_{j=1}^{p} R^j$.

Таким образом, после операции $\mathcal{U}$ реализация данного соединения сводится к выполнению p итераций (где p - количество сегментов вектора 1', содержащих хотя бы одну единицу). В j-й итерации выполняются следующие операции:

$$z_{r_j} = z_{r_{j-1}} \cdot \overline{y_{r_{j-1}}}; \; y_{r_j} = \mu(\varphi, z_{r_j}); \; v_j = \nu(\varphi, y_{r_j}); \; x_j = \overline{1'} \cdot v_j;$$

$$R^j = R_1^{r_j} \otimes \bigcup_{u_i \neq r_j} R_2^{u_i},$$

где $j = \overline{1,p}$; $z_{r_0} = 1'$; $y_{r_0} = \underline{0}$. Здесь вектор y_{r_j} отмечает единицами кортежи подмассива $R_1^{r_j}$ из A', а вектор x_j - кортежи массива R_2 из A', составляющие массив $\bigcup_{u_i \neq r_j} R_2^{u_i}$. Результат $R = \bigcup_{j=1}^{p} R^j$.

Соединение $R_1[V_1 \neq V_2]R_2$ можно выполнить,также используя операцию $\mathcal{B}$

Итак, реализация соединения $R_1[V_1 \theta V_2]R_2$ (после операций упорядочения $\mathcal{U}$ или $\mathcal{B}$) сводится к выполнению определенных последовательностей операций над двоичными векторами и операций декартовых произведений над соответствующими подмассивами.

Таким образом, был показан единый подход к выполнению операций α-алгебры, основанный на предлагаемом методе. Реализация любой операции α-алгебры (кроме декартова произведения) после применения $\mathcal{U}$ и/или $\mathcal{B}$ сводится к выполнению некоторых операций над двоичными векторами (в случае соединения используется также декартово произведение). Отсюда следует, что метод может быть положен в основу для реализации реляционных баз данных. Возможно и непосредственное представление произвольных выражений реляционного исчисления через операции данного метода.

Общность предлагаемого метода заключается в том, что для выполнения различных информационно-логических операций над массивами исходные метки элементов этих массивов подвергаются перегруппировке (перестановке) в соответствии с упорядочением элементов массивов по выделенным доменам и вся дальнейшая обработка сводится к выполнению некоторых (зависящих от заданной операции) последовательностей операций над полученными двоичными векторами меток. Результатом в общем случае являются двоичные векторы, отмечающие единицами те элементы массивов, которые составляют результат обработки или непосредственно участвуют в формировании результата.

Обработку массивов в соответствии с этим методом можно организовать на обычной ЭВМ, а также на различного рода параллельных вычислительных системах, в которых имеется возможность параллельной реализации процедуры упорядочения. Однако вследствие того, что предлагаемый метод обеспечивает возможность выполнения широкого круга важных и трудоемких информационно-логических задач над массивами, а операции $\mathcal{U}$ или $\mathcal{B}$ при этом являются основными (именно они определяют основное время обработки), то целесообразно для повышения производительности эти операции реализовать на специализированных аппаратных средствах. Для параллельной реализации операции упорядочения имеется целый ряд достаточно мощных и доступных для современной технологии аппаратных средств [7.8].Наиболее подходящими из них являются сортирующие сети Бэтчера [9] и Стоуна [10], обладающие наибольшим быстродействием при умеренных расходах оборудования. Отметим, что приведенный выше набор специальных операций над двоичными векторами достаточно просто реализуется аппаратно.

Таким образом, предлагаемый метод может быть положен в основу при построении высокопроизводительных спецпроцессоров для решения невычислительных, информационно-логических задач над массивами, в частности, для построения спецпроцессора реляционных баз данных.

В заключение автор выражает глубокую признательность Я.И.Фету за полезные обсуждения работы.

Л и т е р а т у р а

1. КНУТ Д. Искусство программирования для ЭВМ. Т.3. Сортировка и поиск. -М.: Мир, 1978. - 848 с.

2. SCHWARTZ J.T. Ultracomputers. - ACM Trans.on Progr. Lang. and Syst.,1980,v.2,N 4,p.484-521.

3. ДЕЙТ К. Введение в системы баз данных. -М.: Наука,1980. - 384 с.

4. CODD E.F. A relational model for large shared data banks. - Comm.ACM,1970,v.13,N 6,p.377-387.

5. CODD E.F. Relational completeness of data base sublanguages.-In: Data Base Systems N.Y.,Prentice-Hall,1972,p.79-90.

6. ЦАЛЕНКО М.Ш. Реляционные модели баз данных (обзор).-В кн.: Алгоритмы и организация решения экономических задач. Вып. 9, М., 1977, с. 18-36.

7. THOMPSON C.D. The VLSI complexity of sorting. - IEEE Trans. Comput.,1983,v.C-32,N 12,p.1171-1184.

8. ФЕТ Я.И. Массовая обработка информации в специализированных однородных процессорах. - Новосибирск: Наука, 1976. - 200 с.

9. BATCHER K.E. Sorting networks and their applications.-In: AFIPS Confer.Proc.,1968,SJCC,v.32,p.307-314.

10. STONE H.S. Parallel processing with perfect shuffle.-IEEE Trans.Comput.,1971,v.C-20,N 2,p.153-161.

Поступила в ред.-изд.отд.
25 июля 1984 года