

УДК 004.8
ББК 32.813

КОНЕЧНЫЕ МУЛЬТИМНОЖЕСТВА КАК ОБРАЗЫ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ НА ОСНОВЕ КОЛОНОК

Чесноков А. М.¹

(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)

Рассматривается работа интеллектуальных систем на основе колонок с образами, которые содержат кратные элементы, т.е. представляют собой конечные мультимножества. Рассматривается представление таких образов и показывается существование решения базовых задач. Приводится решение этих задач с помощью метода пересечений.

Ключевые слова: искусственный интеллект, интеллектуальные системы на основе колонок, колонка, мультимножество.

1. Введение

Интеллектуальные системы на основе колонок представляют собой системы, рассматриваемые в рамках следующей модели (более подробно см. [3, 5]).

Имеется пусть и очень большое, но *конечное множество имен U* , предназначенных для наименования объектов произвольной природы. Не ограничивая общности, считается, что множество имен U является подмножеством множества целых чисел. В множестве имен U выделяются непересекающиеся подмножества, получившие название *областей имен*. Причины, которые в реальных предметных областях приводят к выделению областей имен, могут быть совершенно различными.

¹ Александр Михайлович Чесноков, старший научный сотрудник, кандидат технических наук (alex-ches@yandex.ru).

Например, это может быть связано с типизацией. Одной из важнейших причин является необходимость обеспечить отсутствие случайных совпадений имен в различных частях большой системы.

Любое конечное множество имен, принадлежащих тем или иным областям имен, называется *образом*.

Образы любого множества образов P можно перенумеровать, используя для этого имена некоторой области имен U' :

$$P = \{p_i \mid i \in U'\},$$

где $|U'| = |P|$, $|\cdot|$ – мощность множества.

Упорядоченная пара (i, p_i) получила название *колонки*. Колонка обозначается как $(i \mid p_i)$, где i – имя колонки, p_i – образ, содержащийся в колонке. Также используется обозначение $i \rightarrow p_i$. В этом случае говорится, что имя колонки i является *ссылкой*, или *указателем* на содержащийся в колонке образ p_i . В свою очередь, про сам образ в колонке p_i часто будет говориться, что это образ, известный под именем i . Отображение $\varphi: i \rightarrow p_i$ называется *отображением наименования*.

Имя i , которое еще не использовалось для наименования образов, называется *чистым*, или *пустым* именем. Его можно представить как колонку, имеющую пустой образ, т.е. колонку вида $(i \mid \emptyset)$ или $i \rightarrow \emptyset$.

В образы колонок могут входить имена других колонок, а также чистые имена. Таким образом, можно считать, что в образе одной колонки содержатся имена других колонок, каждое из которых служит указателем на соответствующий образ, возможно, пустой. В результате образуется показанная на рис. 1 сложная структура колонок.

Индексом называется любое конечное множество колонок. Состав любого индекса может меняться за счет добавления или удаления колонок. Эти операции называются сложением и вычитанием индексов и обозначаются через $+$ и $-$.

Индекс может быть представлен в виде таблицы, состоящей из вертикальных колонок (столбцов) переменной высоты. В нижней строке таблицы, под чертой, имена колонок. Над именем каждой колонки перечислены все имена, входящие в образ

колонки. По умолчанию считается, что имена колонок и имена в образах принадлежат различным областям имен.

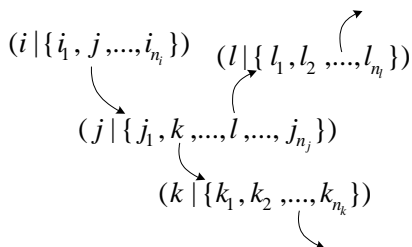


Рис. 1.

Если образы представляют собой неупорядоченные множества имен, то порядок записи имен в образах колонок может быть произвольным. Если же образы упорядочены, то запись имен в образах колонок выполняется в определенном порядке, например, снизу вверх, т.е. первое имя образа в первой строке над чертой, второе – во второй и т.д.

В качестве простейшего примера на рис. 2 показан индекс А, состоящий из трех колонок (1 | {1, 3}), (2 | {2, 3, 4}) и (3 | {4, 5}).

А
4
3 3 5
1 2 4

1 2 3

Рис. 2.

Интеллектуальная система на основе колонок представляет собой один или несколько индексов и работающий с ними механизм (машина колонок), который, получая информацию о внешнем мире в виде образов, формирует новые колонки, изме-

няет уже существующие, удаляет ненужные и выполняет другие необходимые операции (рис. 3).

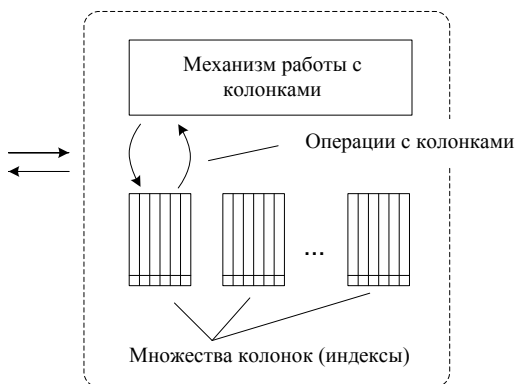


Рис. 3. Система на основе колонок

Знания в рассматриваемых системах представлены с помощью колонок, а в основе процесса накопления знаний лежит запоминание новых образов под определенными именами. При этом *элементарными базовыми задачами*, без которых невозможно функционирование системы, очевидно, являются *прямая задача* – по образу получить его имя, и *обратная задача* – по имени получить соответствующий образ.

Базовые задачи служат основой, на которой строится решение других задач. В [3, 4] рассматривалось решение базовых задач в условиях полной и неполной информации, в частности, для образов, представляющих собой конечные неупорядоченные множества имен. По определению каждое такое множество может содержать лишь единственный экземпляр любого элемента. В то же время в реальных условиях многократное вхождение элементов в образы является обычным явлением. В связи с этим данная работа посвящена решению прямой и обратной задачи для образов в виде конечных неупорядоченных множеств с кратным вхождением элементов, т.е. конечных (неупорядоченных) мультимножеств [1, 2]. В следующем разделе рассматривается представление конечных мультимножеств в

системах на основе колонок и показывается существование решения базовых задач. Затем приводится решение этих задач с помощью метода пересечений.

2. Образы с кратными элементами и общий метод решения базовых задач

2.1. ПРЕДСТАВЛЕНИЕ ОБРАЗОВ С КРАТНЫМИ ЭЛЕМЕНТАМИ

Конечный неупорядоченный образ с повторяющимися элементами, например, образ, который можно записать в виде совокупности $\{a, a, b, b, b, c\}$, представляет собой мультимножество $[1, 2]$. Формально мультимножество можно определить как пару (A, m) , где A – некоторое обычное (*основное*) множество, $m : A \rightarrow N$ – отображение A в множество $N = \{1, 2, 3, \dots\}$, которое для каждого элемента $a \in A$ определяет число вхождений или *кратность* $m_a = m(a)$. Используя определение отображения m в виде множества упорядоченных пар (аргумент, значение), мультимножество можно представить в виде $\{(a, m_a) \mid a \in A\}$, где A – основное множество, m_a – кратность элемента $a \in A$, $m_a \geq 1$. В соответствии с этим приведенное выше в качестве примера мультимножество $\{a, a, b, b, b, c\}$ будет определяться как $\{(a, 2), (b, 3), (c, 1)\}$.

Итак, пусть имеется множество P конечных неупорядоченных образов с кратными элементами вида

$$p = \{(i_k, m_k) \mid k = 1, \dots, n_p\},$$

где $p_0 = \{i_1, \dots, i_{n_p}\}$ – основной образ, $i_k \in U_1$, U_1 – некоторая область имен, $p_0 \in P_0$ – множество всех подмножеств множества U_1 , исключая пустое, m_k – имя, равное кратности имени i_k , $m_k \in U_m$, U_m – область имен, представляющих кратности.

Любая пара (i_k, m_k) является образом в виде двумерного вектора из множества образов $P^2 = U_1 \times U_m$. Образы множества P^2 можно наименовать, используя для этого некоторую область имен U_2 . В результате будут получены колонки вида $(j_k \mid (i_k, m_k))$, где $j_k \in U_2$ – имя образа (i_k, m_k) . Любое имя j_k играет роль «обозначения» пары (i_k, m_k) , так как отображение наименования

$\varphi_m : j_k \rightarrow (i_k, m_k)$ устанавливает взаимно однозначное соответствие между ним и образом (i_k, m_k) . Заменяя в образе p все образы (i_k, m_k) их именами, можно получить образ p' , который представляет собой обычное конечное неупорядоченное множество имен и для которого обычным образом решаются базовые задачи [3], в том числе при неполной информации [4]. Таким образом, решение базовых задач для образов с кратными элементами сводится к решению обычных базовых задач с дополнительным преобразованием образа p в образ p' при решении прямой задачи и образа p' в образ p при решении обратной задачи (рис. 4).

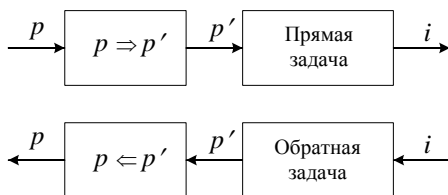


Рис. 4.

Далее везде вместо пар $(i_k, 1)$ в образах p будут записываться просто имена i_k , причем такие имена при преобразовании образов p в p' и p' в p будут оставаться без изменений. Например, образ $p = \{(1, 1), (2, 1), (3, 2)\}$ будет записываться в виде $p = \{1, 2, (3, 2)\}$, и если образ $(3, 2)$ известен под именем 5, то соответствующий образ p' будет равен $\{1, 2, 5\}$.

2.2. УНИВЕРСАЛЬНЫЙ МЕТОД РЕШЕНИЯ БАЗОВЫХ ЗАДАЧ

Для прямой и обратной задачи существует общий метод решения, применимый к любым типам образов. Этим методом является метод, основанный на поэлементном сравнении [3, 4].

Постановка прямой и обратной задачи для образов в виде конечных мультимножеств ничем не отличается от обычной [3, 4]. При решении прямой задачи для образа $p \in P$ необходимо определить его имя i . Если системе это удастся, то имя i являет-

ся решением прямой задачи. В противном случае образ p является новым, и система его запоминает под некоторым именем, которое в данном случае и является решением прямой задачи. При решении обратной задачи необходимо для имени i определить образ p , известный под этим именем. Если образ p найден, то он является решением обратной задачи. В противном случае имя i – это чистое имя.

Для образов с кратными элементами метод, основанный на поэлементном сравнении, будет выглядеть следующим образом.

Система использует два индекса A_m и A . Индекс A_m , в котором хранятся известные системе образы (i, m_i) , $m_i > 1$, представляет собой множество колонок вида $(j_i | (i, m_i))$, где (i, m_i) – образ по имени j_i . Известные системе образы p' хранятся в индексе A , который представляет собой множество колонок вида $(i_{p'} | p')$, где p' – образ, известный под именем $i_{p'}$. В исходном состоянии $A_m = \emptyset$ и $A = \emptyset$.

Любой образ $p \in P$, для которого надо решить прямую задачу, сначала с помощью индекса A_m преобразуется в образ p' . Все входящие в образ p имена с кратностью 1 остаются без изменений. Каждая пара (i_k, m_k) , $m_k > 1$, поэлементно сравнивается с образами всех колонок индекса A_m . Если найден совпадающий образ (i, m_i) , то его имя j_i , т.е. имя колонки $(j_i | (i, m_i)) \in A_m$, заменяет образ (i_k, m_k) в образе p' . Если совпадающий образ не найден, то образ (i_k, m_k) является новым. Для его запоминания выбирается некоторое чистое имя j_k из соответствующей области имен и выполняется сложение $A_m + (j_k | (i_k, m_k))$, т.е. к индексу A_m добавляется колонка $(j_k | (i_k, m_k))$. Пара (i_k, m_k) заменяется в образе p' на имя j_k . Если образ (i_k, m_k) появится снова, то поэлементное сравнение даст для него имя j_k .

Полученный после всех замен образ p' поэлементно сравнивается с образами всех колонок индекса A . Если найден совпадающий образ p_i , то его имя i , т.е. имя колонки $(i | p_i) \in A$, является именем образа p' и решением прямой задачи для образа p . Если совпадающий образ не найден, то образ p' является новым. Для его запоминания выбирается

некоторое чистое имя $i_{p'}$ из соответствующей области имен и выполняется сложение $A + (i_{p'} | p')$, т.е. к индексу A добавляется колонка $(i_{p'} | p')$. Если образ p' появится снова, то поэлементное сравнение даст для него имя $i_{p'}$. Имя $i_{p'}$ является решением прямой задачи для образа p .

Так же просто решается и обратная задача. Если имеется имя i , для которого необходимо решить обратную задачу, то соответствующий образ p' равен образу p_i колонки $(i | p_i) \in A$. Если колонки с таким именем не существует, то i – чистое имя. В противном случае найденный образ p' с помощью индекса A_m преобразуется в образ p . Для каждого имени $i'_k \in p'$ ищется колонка индекса A_m с таким именем. Если такой колонки не существует, то имя i'_k остается без изменений. Если же такая колонка найдена, то имя i'_k заменяется на образ этой колонки. После всех замен будет получен образ p , который является решением обратной задачи для имени i .

ПРИМЕР. Пусть имеются показанные на рис. 5 индексы A и A_m . Легко видеть, что в индексе A хранятся образы $\{1, 3\}$ под именем 1, образ $\{2, 3, 4\}$ под именем 2 и образ $\{3, 5, 6\}$ под именем 3. В индексе A_m под именем 5 хранится образ $(1, 2)$, а под именем 6 – образ $(2, 3)$.

A	A _m
4 6	
3 3 5	2 3
1 2 3	1 2
1 2 3 4	5 6 7

Рис. 5.

Предположим, что прямая задача решается для образа $p = \{1, 3\}$. Очевидно, соответствующий образ $p' = \{1, 3\}$. Поэлементное сравнение показывает, что он равен образу колонки 1 индекса A . Следовательно, образ $p = \{1, 3\}$ известен системе под именем 1. Аналогично для образа $p = \{2, 3, 4\}$ будет получен

образ $p' = \{2, 3, 4\}$, который равен образу колонки 2 индекса A , т.е. $p = \{2, 3, 4\}$ – это образ по имени 2.

Пусть системе предъявлен образ $p = \{(1, 2), (2, 3), 3\}$. Он состоит из имени 1 с кратностью 2, имени 2 с кратностью 3 и имени 3 с кратностью 1. При преобразовании образа p в образ p' имя 3 останется без изменений. Поэлементное сравнение пар $(1, 2)$ и $(2, 3)$ с образами колонок индекса A_m показывает, что первая – это образ по имени 5, а вторая – образ по имени 6. Следовательно, образ $p' = \{5, 6, 3\}$. Его сравнение с образами колонок индекса A показывает, что это образ по имени 3, т.е. образ p известен системе под именем 3.

Пусть теперь для имени $i = 3$ решается обратная задача. Образ p' равен образу колонки 3 индекса A , т.е. $p' = \{3, 5, 6\}$. При преобразовании в образ p имя 3 останется без изменений, а имена 5 и 6 заменяются образами колонок с именами 5 и 6 индекса A_m . В результате для имени $i = 3$ будет получен образ $p = \{(1, 2), (2, 3), 3\}$.

Достоинством метода на основе поэлементного сравнения является его универсальность, недостатком – низкая эффективность. Далее для образов в виде конечных неупорядоченных множеств с кратными элементами рассматривается более эффективный метод пересечений [3, 4].

3. Метод пересечений для конечных неупорядоченных образов с кратными элементами

При решении базовых задач с помощью метода пересечений в системе используются индексы A , $A_m = \{A_{m1}, A_{m2}\}$, B и B_m , а также заданная в виде множества упорядоченных пар (i, n_i) функция $n(i)$, которая содержит мощности известных образов p' .

В начале работы $A = \emptyset$, $A_m = \emptyset$, $B = \emptyset$, $B_m = \emptyset$ и $n(i) = \emptyset$.

Пусть прямая задача решается для некоторого образа $p = \{(i_k, m_k) \mid k = 1, \dots, n_p\} \in P$. Как и ранее каждая из пар $(i_k, 1)$ заменяются именем i_k . Для остальных пар (i_k, m_k) , $m_k > 1$, при

преобразовании образа p в образ p' решаются базовые задачи как для образов в виде двумерных векторов [3, 4].

Для каждой пары (i_k, m_k) , $m_k > 1$, вычисляется покоординатное пересечение $\eta((i_k, m_k)) = a_{i_k} \cap a_{m_k}$, где a_{i_k} – образ колонки $(i_k | a_{i_k}) \in A_{m_1}$, a_{m_k} – образ колонки $(m_k | a_{m_k}) \in A_{m_2}$. Если покоординатное пересечение $\eta((i_k, m_k)) \neq \emptyset$, то оно содержит единственное имя j_k , под которым известен образ (i_k, m_k) [3, 4]. На это имя заменяется пара (i_k, m_k) в образе p' . Если же пересечение $\eta((i_k, m_k)) = \emptyset$, то образ (i_k, m_k) является новым. Для его запоминания выбирается любое чистое имя $j \in U_2 \setminus U_{(i, m)}$, где U_2 – область имен для наименования образов $(i, m_i) \in P^2$, $U_{(i, m_i)}$ – множество имен всех известных образов (i, m_i) . Затем выполняются сложения

$$\begin{aligned} A_m + ((i_k, m_k) | \{j\}) &= \{A_{m_1} + (i_k | \{j\}), A_{m_2} + (m_k | \{j\})\}, \\ B_m + (j | (i_k, m_k)), \end{aligned}$$

т.е. к индексу A_{m_1} добавляется колонка $(i_k | \{j\})$, к индексу A_{m_2} – колонка $(m_k | \{j\})$, а к индексу B_m – колонка $(j | (i_k, m_k))$. Пара (i_k, m_k) в образе p' заменяется на имя j .

Если в дальнейшем опять появится пара (i_k, m_k) , то пересечение $\eta((i_k, m_k))$ будет содержать единственное имя j , под которым известен образ (i_k, m_k) .

После того как получен образ p' , для него решаются обычные базовые задачи как для образов в виде конечных неупорядоченных множеств [3, 4].

Вычисляется пересечение $\eta(p') = \bigcap_{i' \in p'} a_{i'}$, где $a_{i'}$ – образ колонки $(i' | a_{i'}) \in A$. Если $\eta(p') = \emptyset$ или если $\eta(p') \neq \emptyset$, но $n(i) \neq p' |$ для $\forall i \in \eta(p')$, то образ p' является новым. Для него выбирается любое чистое имя $i_{p'} \in U' \setminus U_{p'}$, где U' – область имен для наименования образов p' , $U_{p'}$ – множество имен всех известных образов p' . Затем выполняются сложения

$$\begin{aligned} A + (p' | \{i_{p'}\}), \\ B + (i_{p'} | p'), \end{aligned}$$

т.е. к индексу A для всех $i'_k \in p'$ добавляются колонки $(i'_k | \{i_{p'}\})$, а к индексу B добавляется колонка $(i_{p'} | p')$. Кроме того, в опре-

деление функции $n(i)$ добавляется пара $(i_{p'}, |p'|)$. Имя $i_{p'}$ является именем, под которым теперь будет известен образ p' , и представляет собой решение прямой задачи для образа p .

Если пересечение $\eta(p) \neq \emptyset$ и существует имя $i \in \eta(p)$ такое, что $n(i) = |p'|$, то это имя является единственным и представляет собой решение прямой задачи, т.е. является именем, под которым известны образы p' и p [3, 4].

При решении обратной задачи для некоторого имени i образ p' равен образу b_i колонки $(i | b_i) \in B$. Если колонки с таким именем не существует, то i – чистое имя. В противном случае выполняется преобразование образа $p' = b_i$ в образ p . Каждое входящее в образ p' имя, равное имени колонки индекса B_m , заменяется образом той же колонки. Остальные имена образа остаются без изменений. Другими словами, если имя $j \in p'$ – это имя колонки $(j | (i_j, m_j)) \in B_m$, то оно заменяется на пару (i_j, m_j) . После всех подобных замен будет получен образ p , известный системе под именем i .

ПРИМЕР. Пусть имеются показанные на рис. 6 индексы A , $A_m = \{A_{m1}, A_{m2}\}$, B , B_m и функция $n(i)$. Легко видеть, что в индексе A запомнены образы $\{1, 3\}$ под именем 1 и $\{2, 4\}$ под именем 2. Индексы $A_m = \emptyset$, $B_m = \emptyset$.

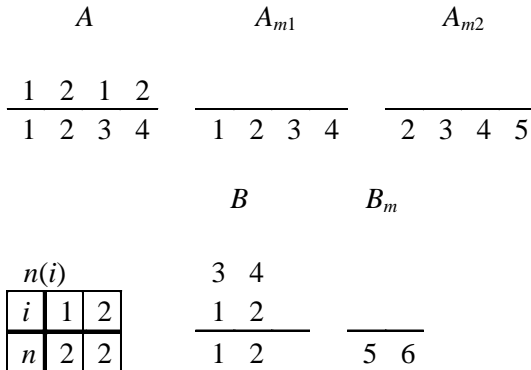


Рис. 6.

При решении прямой задачи сначала входной образ p преобразуется с помощью индекса A_m в образ p' . При этом имена i с кратностью $m_i = 1$ остаются без изменений. Затем с помощью индекса A определяется имя образа p' .

Пусть на вход поступил образ $p = \{1, 3\}$. Оба элемента образа имеют кратность 1. Следовательно, образ $p' = \{1, 3\}$. Пересечение $\eta(p')$ колонок 1 и 3 индекса A равно $\{1\}$, причем $n(1) = 2$, т.е. входной образ $p = \{1, 3\}$ известен системе под именем 1.

Аналогично, если на входе образ $p = \{2, 4\}$, то $p' = \{2, 4\}$. Пересечение $\eta(p') = \{2\}$, причем $n(2) = 2$, т.е. входной образ $p = \{2, 4\}$ – это образ по имени 2.

Пусть теперь рассматривается образ $p = \{(1, 2), 3\}$, т.е. имя 1 имеет кратность 2. Для образа $(1, 2)$ в A_m покоординатное пересечение $\eta_m((1, 2)) = \emptyset$. Поэтому образ $(1, 2)$ является новым и запоминается в A_m под именем 5. Получим $p' = \{5, 3\}$, для которого $\eta(p') = \emptyset$, т.е. он неизвестен системе. После его запоминания под именем 3 будем иметь (рис. 7):

A					A_{m1}				A_{m2}			
3												
1	2	1	2	3	5				5			
1	2	3	4	5	1	2	3	4	2	3	4	5
					B				B_m			
$n(i)$					3	4	5	2				
i	1	2	3	1				1				
n	2	2	2	1	2	3	5 6					

Рис. 7.

Имя 3 является решением прямой задачи для образа $p = \{(1, 2), 3\}$. Если этот образ опять будет предъявлен системе,

то для него $\eta_m((1, 2)) = \{5\}$, т.е. $p' = \{5, 3\}$. Пересечение $\eta(p') = \{3\}$, причем $n(3) = |p'|$, т.е. p – образ, известный системе под именем 3.

Предположим, что на входе образ $p = \{(1, 2), 2, (3, 3)\}$. Пересечения $\eta_m((1, 2)) = \{5\}$, $\eta_m((3, 3)) = \emptyset$. Образ (3, 3) является новым и запоминается под именем 6. Для образа $p' = \{2, 5, 6\}$ пересечение $\eta(p') = \emptyset$, т.е. он неизвестен системе. После его запоминания под именем 4 получим (рис. 8):

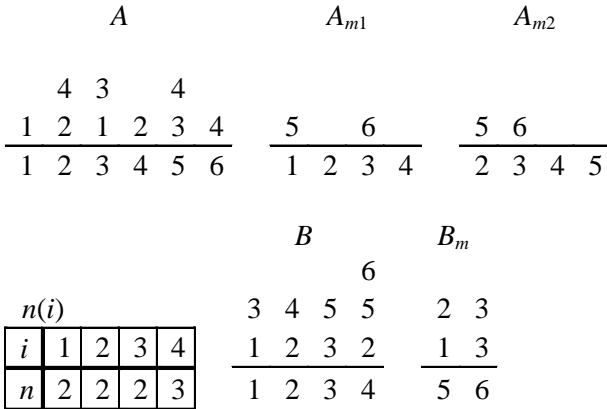


Рис. 8.

Если на входе опять появится образ $p = \{(1, 2), 2, (3, 3)\}$, то $\eta_m((1, 2)) = \{5\}$, $\eta_m((3, 3)) = \{6\}$ и $p' = \{2, 5, 6\}$. Для него $\eta(p') = \{4\}$ и $n(4) = |p'|$, т.е. p – это образ по имени 4.

Пусть теперь решается обратная задача для имени $i = 4$. Образ $p' = b_4 = \{2, 5, 6\}$. С помощью индекса B_m устанавливается, что под именем 5 известен образ $b_{m5} = (1, 2)$, а под именем 6 – образ $b_{m6} = (3, 3)$. Это означает, что решением обратной задачи для имени $i = 4$ является образ $p = \{(1, 2), 2, (3, 3)\}$.

Литература

1. АЙГНЕР М. *Комбинаторная теория*. – М.: Мир, 1982. – 558 с.
2. ПЕТРОВСКИЙ А.Б. *Основные понятия теории мультимножеств*. – М.: Едиториал УРСС, 2002. – 80 с.
3. ЧЕСНОКОВ А.М. *Интеллектуальные системы на основе колонок* // Управление большими системами. – 2013. – №46. – С. 118–146.
4. ЧЕСНОКОВ А.М. *Интеллектуальные системы на основе колонок при неполной информации* // Управление большими системами. – 2014. – №50. – С. 84–98.
5. ЧЕСНОКОВ А.М. *Введение в общую теорию колонок*. – М.: ИПУ РАН. – 2012. – 141 с.

FINITE MULTISSETS AS PATTERNS IN COLUMNS-BASED INTELLIGENT SYSTEMS

Alexander Chesnokov, Institute of Control Sciences of RAS, Moscow, Cand. Sc. (alex-ches@yandex.ru).

Abstract: We consider columns-based intelligent systems with patterns containing duplicates, i.e., representing the finite multisets. The representation of such patterns is suggested and solutions to the direct and the inverse problems are discussed. We use the intersections technique to solve both problems.

Keywords: artificial intelligence, columns-based intelligent systems, column, multiset.

Статья представлена к публикации членом редакционной коллегии М.В. Губко

*Поступила в редакцию 10.08.2014.
Опубликована 30.11.2014.*