EESTI NSV TEADUSTE AKADEEMIA TOIMETISED. 23. KOIDE
FÜÜSIKA * MATEMAATIKA. 1974, NR. 1

ИЗВЕСТИЯ АКАДЕМИИ НАУК ЭСТОНСКОЙ ССР. ТОМ 23.
ФИЗИКА * МАТЕМАТИКА. 1974, № 1

https://doi.org/10.3176/phys.math.1974.1.04

УДК 681.3.06

И. МЕЛЬНИКОВ, Б. ТАММ

ОБ ОДНОМ МЕТОДЕ ФОРМАЛИЗАЦИИ СТРУКТУР ДАННЫХ

От эффективности представления данных в большой мере зависит и эффективность их обработки на ЭВМ. В связи с этим у нас и за рубежом опубликовано множество работ, посвященных вопросам представления данных. Многие из них содержат интуитивно хорошо понимаемое понятие структуры данных, но, к сожалению, с разным содержанием для различных авторов. В настоящей работе даются формализированные понятия как структуры данных, так и связанных с ней определений. При формализации используются основные понятия теории множеств [1], теории графов [2] и теории структур [3].

Пусть $X = \{x_1, x_2, \ldots, x_n\}$ множество переменных, связанных между собой при решении задач определенной проблемы. Переменной является объект или множество объектов, обрабатываемых совместно при решении задач проблемы, и она обозначает определенное понятие проблемы.

Каждая переменная x_i может иметь несколько значений $\bar{x}_i$. Областью определения $\mu(x_i)$ переменной x_i называется множество допустимых значений $\bar{x}_i$. Число различных значений $n_i = |\mu(x_i)|$ называется длиной переменной x_i.

Значение переменной характеризуется длиной, типом и идентификатором. Длина значения $l(\bar{x}_i)$ определяет объем запоминающего устройства (ЗУ), требуемый для записи $\bar{x}_i$, а тип значения $\tau(\bar{x}_i)$ (напр., строка символов, десятичное число и т. д.) — семантику операции для этого значения.

Переменная x_i считается однородной, если все ее значения имеют одинаковую длину и относятся к одному типу.

Областью определения структуры данных называется множество $\mu(X) = \bigcup_{i=1}^{n} \mu(x_i)$ и структурой данных — подмножество $C(X) \subset \subset \mu(X) \times \mu(X)$.

Структура*, у которой область определения пустое множество, является пустой структурой и обозначается Ф. Пара $(\bar{x}_i, \bar{x}_j)$ называется связью, которая определяет местонахождение $\bar{x}_j$ в ЗУ по местонахождению $\bar{x}_i$. Связанностью значения $\varrho(\bar{x}_i)$ называется число всех связей $(\bar{x}_j, \bar{x}_i)$.

При $\varrho(\bar{x}_i) = 0$ значение $\bar{x}_i$ считается прямо идентифицируемым. $\underline{\varrho}(\bar{x}_i)$ обозначает число всех связей $(\bar{x}_i, \bar{x}_j)$.

Схемой структуры является множество

* В дальнейшем под понятием структуры понимается структура данных.

$$G(X)=\{(x_i,x_j)\,|\,\forall x_i,x_j\in X\,\exists\,\bar{x}_i,\bar{x}_j\in\mu(X)\,((\bar{x}_i,\bar{x}_j)\in C(X))\}.$$

Переменные x_i и x_j (значения $\bar{x}_i$ и $\bar{x}_j$) считаются сравнимыми и обозначаются $x_i > x_j (\bar{x}_i > \bar{x}_j)$, если существует $(x_i, x_j) \in G(X)$ (соответственно $(\bar{x}_i, \bar{x}_j) \in C(X)$). Переменная x_i называется входной переменной, если $\forall x_j \in X (x_j \not> x_i)$, и простой переменной, если $\forall x_j \in \in X (x_i \not> x_j)$. Множество простых переменных обозначается $\Pi(X)$, а множество входных переменных $Q(X)$. Множества $\overline{\Pi}(X) = \bigcup_{x_i \in \Pi(X)} \mu(x_i)$ и $\overline{Q}(X) = \bigcup_{x_i \in Q(X)} \mu(x_i)$ называются соответственно множествами простых и входных значений.

Маршрутом переменных называется упорядоченное множество

$$w(x_i,x_j)=\{x_k\,|\,\forall x_k\in X(x_i>x_k\vee x_k>x_j\vee\exists\,x_l,x_m\in w(x_i,x_j)\,(x_k\in w(x_l,x_m)))\}$$

и маршрутом значений — множество

$$w(\bar{x}_i,\bar{x}_j)=\{\bar{x}_k\,|\,\forall\bar{x}_k\in\mu(X)\,(\bar{x}_i>\bar{x}_k\vee\bar{x}_k>\bar{x}_j\vee\exists\,\bar{x}_l,\bar{x}_m\in w(\bar{x}_i,\bar{x}_j)\,(\bar{x}_k\in w(\bar{x}_l,\bar{x}_m)))\}.$$

В любом маршруте первый элемент называется началом, а последний элемент — концом маршрута. Маршруты $w(a,b)$ и $w(a,c)$ называются расходящимися, если $w(a,b) \cap w(a,c)$ — маршрут. Переменные x_i и x_j (значения $\bar{x}_i$ и $\bar{x}_j$) называются связанными и обозначаются $x_i \rightarrow x_j$ $(\bar{x}_i \rightarrow \bar{x}_j)$, если существует маршрут переменных $w(x_i, x_j)$ (значений $w(\bar{x}_i, \bar{x}_j)$).

Маршрут, в котором начало и конец совпадают, называется циклом. Маршрут $w(x_i, x_j)$ называется цепью переменных, если $x_i \in Q(X)\ \&\ x_j \in \in \Pi(X)$, и маршрут значений $w(\bar{x}_i, \bar{x}_j)$ — цепью значений, если $\bar{x}_i \in \in \overline{Q}(X)\ \&\ \bar{x}_j \in \overline{\Pi}(X)$. Номер позиции элемента цепи от начала цепи определяет уровень иерархии элемента цепи, а число элементов в маршруте — длину маршрута.

Схема структуры называется постоянной, если длины переменных и связи между переменными не изменяются в ходе решения любой задачи проблемы.

Структура с постоянной схемой называется постоянной, если при решении любой задачи проблемы связи между значениями не изменяются. Непостоянные структуры называются абстрактными структурами. Структура является простой, если $\forall x_i \in X (x_i \in \Pi(X))$, и иерархической, если все маршруты переменных структуры являются попарно расходящимися цепями. Иерархическая структура называется деревообразной, если все цепи значений попарно расходящиеся. Структура является однородной, если все переменные однородные.

Чтобы сравнивать вычислительные ресурсы, употребляемые структурами, определяется понятие эквивалентности структур.

Идентифицируемые значения $\bar{x}_i \in \mu(X)$ и $\bar{y}_j \in \mu(Y)$ называются эквивалентными и обозначаются $\bar{x}_i = \bar{x}_j$, если они при обработке значений имеют одинаковое содержание**.

Структура $C(X)$ называется подструктурой структуры $C(Y)$ и обозначается $C(X) \subset C(Y)$, если $\forall \bar{x}_i \in \mu(X)\ \exists\,\bar{y}_j \in \mu(Y)\,(\bar{x}_i = \bar{y}_j)$.

Структуры $C(X)$ и $C(Y)$ называются эквивалентными и обозначаются $C(X) = C(Y)$, если $C(X) \subset C(Y) \subset C(X)$.

** Так как эквивалентные значения в разных структурах могут иметь неодинаковые идентификаторы, то они обозначаются различно.

Нетрудно показать, что отношение эквивалентности структур имеет следующие свойства:

1) рефлексивность $C(X) = C(X)$,

2) симметричность $C(X) = C(Y) \Rightarrow C(Y) = C(X)$,

3) транзитивность $C(X) = C(Y) \,\&\, C(Y) = C(Z) \Rightarrow C(X) = C(Z)$.

Между структурами определяются операции умножения, сложения и вычитания.

Умножением структур $C(X_1)$ и $C(X_2)$ называется операция

$$C(X_1) \cap C(X_2) = C(X) \Rightarrow \exists\, C(X)\, (C(X_1) \supset C(X) \,\&\, C(X_2) \supset \supset C(X) \,\&\, \mu(X) = \mu(X_1) \cap \mu(X_2)).$$

Структуры данных называются изолированными, если $C(X_1) \cap \cap C(X_2) = \Phi$, и связанными, если $C(X_1) \cap C(X_2) \neq \Phi$.

Сложением структур $C(X_1)$ и $C(X_2)$ называется операция

$$C(X_1) \cup C(X_2) = C(X) \Rightarrow \exists\, C(X)\, (C(X) \supset C(X_1) \,\&\, C(X) \supset \supset C(X_2) \,\&\, \mu(X) = \mu(X_1) \cup \mu(X_2))$$

и вычитанием — операция

$$C(X_1) / C(X_2) = C(X) \Rightarrow \exists\, C(X)\, (C(X) \subset C(X_1) \,\&\, C(X) \cap \cap C(X_2) = \Phi \,\&\, \mu(X) = \mu(X_1) / \mu(X_2)).$$

Логическим продолжением понятия эквивалентности структур является понятие эквивалентного преобразования структур. Эквивалентным преобразованием структур считается преобразование одной структуры в другую, эквивалентную ей структуру. Для обозначения эквивалентного преобразования структуры $C(X)$ в структуру $C(Y)$ используется запись $C(Y) = A[C(X)]$, где A — эквивалентное преобразование. Так как отношение эквивалентности является однозначным соответствием, все эквивалентные преобразования состоят из последовательности преобразований следующих типов. Если $C(Y) = A[C(X)]$, то возможны следующие подстановки.

1. Значение переменной одной структуры заменяется значением другой структуры

$$\bar{y}_k = \bar{x}_i.$$

2. Связь одной структуры заменяется значением другой структуры

$$\bar{y}_k = (\bar{x}_i, \bar{x}_j).$$

3. Связь одной структуры заменяется связью другой структуры

$$(\bar{y}_k, \bar{y}_l) = (\bar{x}_i, \bar{x}_j).$$

4. Значение переменной одной структуры заменяется связью другой структуры

$$(\bar{y}_k, \bar{y}_l) = \bar{x}_i.$$

Для задания структуры данных можно использовать следующие способы.

1. Перечисление. При этом способе перечисляются все значения и связи структуры.

2. Задание матрицы структуры. Пусть $B = (l_{ij})$ матрица структуры, в которой каждому столбцу и каждой строке соответствует одно значение из области определения структуры. Если значению $\bar{x}_v^m$ соответствует столбец с номером j, а значению $\bar{x}_u^l$ строка с номером i и, если $\bar{x}_u^l > \bar{x}_v^m$, то $b_{ij} = 1$, в противном случае $b_{ij} = 0$.

3. Задание графа структуры. Ориентированный граф $T = (U, V)$ с множеством вершин U и множеством ребер V называется графом структуры $C(X)$, если $U = \mu(X)$ и $V = \{v_i \mid \forall \bar{x}_i,\ \bar{x}_j \in \mu(X)\ \exists v_i \in V\ (\bar{x}_i > \bar{x}_j)\}$.

4. Задание множества цепей и циклов. При этом способе перечисляются все цепи и циклы структуры.

5. Задание классов эквивалентности. Классом эквивалентности значения $\bar{x}_i$ называется множество $c_i = \{\bar{x}_j \mid \forall \bar{x}_j \in \mu(X)\ (\bar{x}_i > \bar{x}_j)\}$. Структура задается перечислением всех классов эквивалентности.

6. Задание подструктур структуры. При этом способе перечисляются определенные подструктуры и связываются в одну структуру при помощи операций, определенных между структурами. Мы будем из всевозможных подструктур различать некоторые структуры и называть их элементарными структурами. Элементарными структурами являются множество, вероятностное множество, разветвление, вероятностное разветвление, цепь, вероятностная цепь. Если $\mu(Y) = \{y_0, y_1, \ldots, y_n\}$ область определения структуры, то множеством называется структура $C_1(Y) = \{y_0, y_1, \ldots, y_n\}$, вероятностным множеством — структура $C_2(Y) = \{y'_i \mid (y'_i = y_i \vee y'_i \in \Phi)\}$, разветвлением — структура $C_3(Y) = \{(y_0, y_i) \mid \forall y_i \in \mu(Y)\}$, вероятностным разветвлением — структура $C_4(Y) = \{(y_0, y'_i) \mid (y'_i = y_i \vee y'_i \in \Phi)\}$, цепью — структура $C_5(Y) = \{(y_i, y_{i+1}) \mid \forall y_i \in \mu(Y)\}$, вероятностной цепью — структура $C_6(Y) = \{(y'_i, y'_{i+1}) \mid (y'_i = y_i \vee y'_i \in \Phi)\}$. Вероятностные элементарные структуры следует понимать так, что каждое значение структуры может отсутствовать. Значение y_0 называется обозначением элементарной структуры. Для описания элементарных структур используются следующие виды записей.

1. Множество $y_0 = (y_1, y_2, \ldots, y_n)$.
2. Вероятностное множество $y_0 = \vee(y_1, y_2, \ldots, y_n)$.
3. Разветвление $y_0 \downarrow (y_1, y_2, \ldots, y_a)$.
4. Вероятностное разветвление $y_0 \downarrow \vee\ (y_1, y_2, \ldots, y_n)$.
5. Цепь $y_0 \rightarrow (y_1, y_2, \ldots, y_n)$.
6. Вероятностная цепь $y_0 \rightarrow \vee\ (y_1, y_2, \ldots, y_n)$.

Любую структуру можно задать при помощи элементарных структур. Полное описание структуры при помощи элементарных структур называется формулой структуры.

Простым примером структуры данных могут служить данные, связанные между собой при решении задач геометрии многоугольников. Множеством переменных этой проблемы является $L = \{O, P, X, Y\}$. Каждый многоугольник O_j из $\mu(O)$ определен множеством граничных точек из $\mu(P)$, а каждая точка P_i задана своими координатами из $\mu(X)$ и $\mu(Y)$. На рис. 1 представлены начальные данные конкретной задачи этой проблемы, на рис. 2 приведен граф схемы структуры и на рис. 3 граф структуры этих данных.

Кроме графа структуры, можно задавать структуру данных на рис. 1 и другими способами.

1. Перечисление

$$\mu(L) = \{O_1,\ O_2,\ P_1,\ P_2,\ P_3,\ P_4,\ P_5,\ P_6,\ P_7,\ 2,\ 2.5,\ 3,\ 3.5,\ 5,\ 6.5,\ 7,\ 9,\ 12,\ 13,\ 14\};$$

$$O_1 > P_1,\ O_1 > P_2,\ O_1 > P_3,\ P_1 > 2,\ P_1 > 3,\ P_2 > 3,\ P_2 > 7,\ P_3 > 7,$$
$$P_3 > 2,\ O_2 > P_4,\ O_2 > P_5,\ O_2 > P_6,\ O_2 > P_7,\ P_4 > 9,\ P_4 > 5,$$
$$P_5 > 12,\ P_5 > 6.5,\ P_6 > 14,\ P_6 > 3.5,\ P_7 > 13,\ P_7 > 2.5.$$

2. Матрица структуры

	O_1	O_2	P_1	P_2	P_3	P_4	P_5	P_6	P_7	2	2.5	3	3.5	5	6.5	7	9	12	13	14
O_1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O_2	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
P_1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
P_2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
P_3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0
P_4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
P_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0
P_6	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
P_7	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0

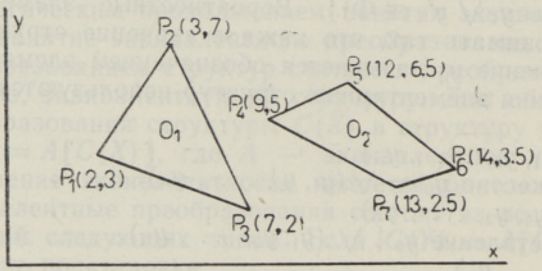

Рис. 1. Начальные данные из задачи геометрии многоугольников.

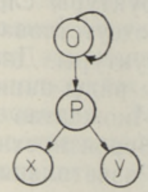

Рис. 2. Граф схемы структуры данных.

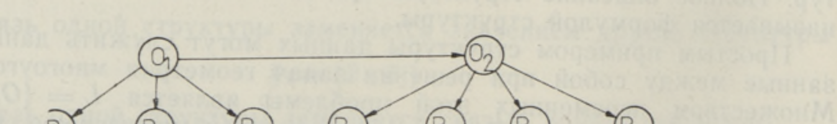

Рис. 3. Граф структуры для данных, приведенных на рис. 1.

Так как все строки, соответствующие простым значениям, заполнены нолями, то они пропущены.

3. Множество цепей и циклов

$O_1 \to P_1 \to 2$, $O_1 \to P_1 \to 3$, $O_1 \to P_2 \to 3$, $O_1 \to P_2 \to 7$, $O_1 \to P_3 \to 7$, $O_1 \to P_3 \to 2$, $O_2 \to P_4 \to 9$, $O_2 \to P_4 \to 5$, $O_2 \to P_5 \to 12$, $O_2 \to P_5 \to 6.5$, $O_2 \to P_6 \to 14$, $O_2 \to P_6 \to 3.5$, $O_2 \to P_7 \to 13$, $O_2 \to P_7 \to 2.5$.

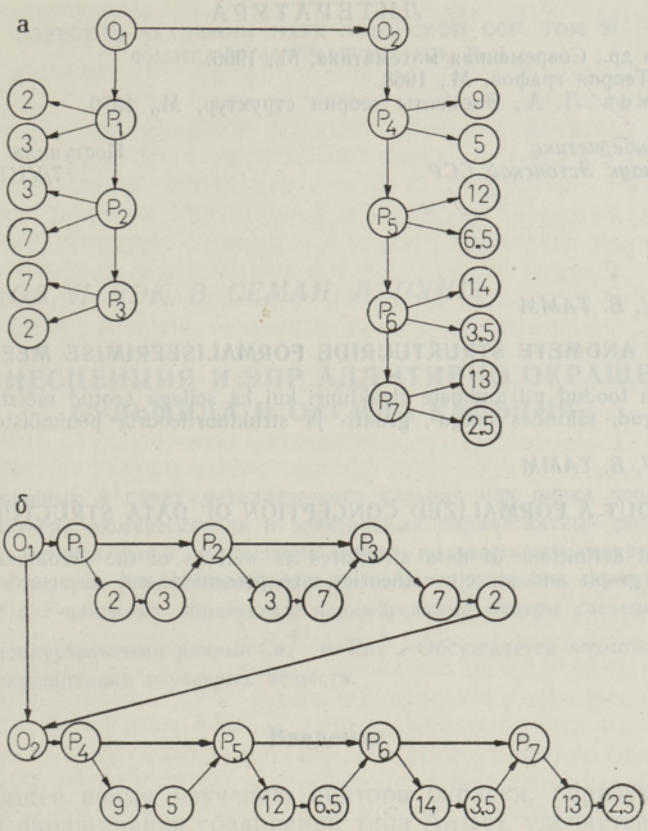

Рис. 4. Графы структур, эквивалентные структуре на рис. 3.

4. Классы эквивалентности значений

$$c(O_1)=\{P_1,\ P_2,\ P_3\},$$
$$c(O_2)=\{P_4,\ P_5,\ P_6,\ P_7\},$$
$$c(P_1)=\{2,\ 3\},\ c(P_2)=\{3,\ 7\},\ c(P_3)=\{7,\ 2\},$$
$$c(P_4)=\{9{,}5\},\ c(P_5)=\{12,\ 6.5\},\ c(P_6)=\{14,\ 3.5\},$$
$$c(P_7)=\{13,\ 2.5\}.$$

5. Задание подструктур структуры

$$c_1(L)=O_1\cup O_2,\ O_1=P_1\cup P_2\cup P_3,\ O_2=P_4\cup P_5\cup P_6\cup P_7,$$
$$P_1=2\cup 3,\ P_2=3\cup 7,\ P_3=7\cup 2,\ P_4=9\cup 5,\ P_5=12\cup 6.5,$$
$$P_6=14\cup 3.5,\ P_7=13\cup 2.5.$$

6. Формула структуры

$$C_1(L)\to(O_1\downarrow(P_1\downarrow(2,\ 3),\ P_2\downarrow(3,\ 7),\ P_3\downarrow(7,\ 2)),\ O_2\downarrow$$
$$\downarrow(P_4\downarrow(9,\ 5),\ P_5\downarrow(12,\ 6.5),\ P_6\downarrow(14,\ 3.5),\ P_7\downarrow(13,\ 2.5))).$$

Структура данных этой проблемы является иерархической, а структура, представленная на рис. 2, — деревообразной.

Приведенные на рис. 4*а*, *б* структуры эквивалентны структуре, приведенной на рис. 3.

ЛИТЕРАТУРА

1. Фор Р. и др., Современная математика, М., 1966.
2. Оре О., Теория графов, М., 1968.
3. Скорняков Л. А., Элементы теории структур, М., 1970.

Институт кибернетики
Академии наук Эстонской ССР

Поступила в редакцию
7/VIII 1973

I. MELNIKOV, B. TAMM

ÜHEST ANDMETE STRUKTUURIDE FORMALISEERIMISE MEETODIST

Artiklis on toodud nii andmete struktuuri kui ka sellega seotud mõistete formaliseeritud määrangud, lähtudes hulga-, graafi- ja struktuuriteooria põhimõistetest.

I. MELNIKOV, B. TAMM

ABOUT A FORMALIZED CONCEPTION OF DATA STRUCTURES

Formalized definitions of data structures as well as of the accompanying concepts based on set, graph and structure theories, are presented and discussed in this paper.