

## КОНТЕКСТНАЯ ГРАММАТИКА ДЛЯ ОПИСАНИЯ ПОТОКОВ ДАННЫХ ЗАДАЧ

А. БАРТКУС, Л. СИМАНАУСКАС

Процесс обработки данных можно представить как процесс их движения с одних массивов в другие (с одних носителей в другие), когда вместе с переносом данных может осуществляться и их преобразование. Экономические задачи, решаемые в АСУ, отличаются большими объемами используемой информации и большим количеством массивов. Процесс обработки таких массивов по характеру выполнения операций разделяется на два этапа: подготовка массивов на машинных носителях и непосредственное решение задачи, в результате чего получаются другие массивы или машинограммы, печатаемые на АЦПУ. В дальнейшем будем считать, что решение задачи в ИВЦ сводится к обработке массивов, уже нанесенных на машинные носители.

Процесс решения задачи на стадии технического проектирования разрабатывается в два этапа. На первом этапе производится разбивка процесса решения на отдельные блоки, которые можно осуществить автономно на ЭВМ и каждый из которых удовлетворяет определенным требованиям. На втором этапе производится детализация каждого блока путем составления для него алгоритма.

Разбивка процесса решения задачи на блоки по сути дела сводится к определению потоков движения данных через массивы (носители). Такие потоки данных, возникающие при решении задачи, можно представить при помощи специальных графов, имеющих вершины четырех видов: источников данных, массивов данных (машинограмм), операций обработки массивов данных, пользователей данных. Граф, изображающий схему потоков данных, является связным ориентированным графом, а его вершины изображаются соответствующими символами по ГОСТу 19428-74. Схему, представленную в виде такого графа, можно назвать схемой потоков данных задачи.

Источником данных для задачи может быть архив (для первичных массивов) или другая задача (для результатно-промежуточных массивов). Поэтому начальных вершин графа может быть несколько.

Одной вершине массива данных поставлен в соответствие один массив, который для задачи может быть исходным, промежуточным, результатным или результатно-промежуточным. В каждую такую вершину входит только одна дуга — из источника данных или вершины операции, формирующей этот массив. Из каждой такой вершины исходит по меньшей мере одна дуга. Число исходящих дуг равно числу пользователей массива (машинограммы). Пользователями могут быть блоки других операций, другие задачи или подразделения предприятия. В последних двух случаях пользователь указывается при помощи специального символа «Архив» или «Приемник данных». Из таких вершин не исходит ни одна дуга, и для графа они являются конечными. Таких конечных вершин может быть несколько, а в каждую из них может входить любое

число дуг. Число входящих дуг соответствует количеству массивов или машинограмм, поступающих к этому пользователю.

Каждой вершине обработки массивов данных задачи соответствует алгоритм (программа) или работа, выполняемые как автономные. При построении схем потоков данных наиболее сложной работой как раз и является определение блоков операций обработки массивов данных, так как количество таких блоков и их содержание зависит от исходных и формируемых массивов, формируемых машинограмм, используемых средств вычислительной техники и ряда других факторов. Исходными данными для построения схем потоков данных и выделения блоков операций являются номенклатура и структура массивов и машинограмм и соответствующие расчетные формулы, определяющие связи между показателями, помещенными в соответствующих массивах.

В настоящее время основную техническую базу АСУП составляют ЭВМ с внешними запоминающими устройствами последовательного доступа (на магнитных лентах). При разработке схем потоков данных для реализации их на таких ЭВМ каждый блок операций обработки массивов данных должен удовлетворять следующим требованиям: 1) количество одновременно обрабатываемых и получаемых массивов (машинограмм) не должно превышать наличного количества соответствующих устройств ЭВМ (лентопротяжных механизмов, устройств ввода и вывода); 2) одновременно обрабатываемые массивы должны быть рассортированы по тем же ключевым признакам (ключам сортировки). При наличии в ЭВМ запоминающих устройств с прямым доступом эти требования становятся менее категоричными: можно одновременно обрабатывать массивы, находящиеся на одном дисковом устройстве; при обработке одновременно нескольких массивов, когда не совпадают главные ключи сортировки и один из них имеет небольшое число различных значений (при совпадении других ключей сортировки), то и второе требование может быть не удовлетворено. Аналогично можно поступать и при обработке массивов, помещенных на магнитных лентах, но, как показывает опыт, такой способ обработки массивов оправдывается лишь при наличии нескольких различных значений главных ключей сортировки (двух, трех) и ниже рассматриваться не будет. Таким образом, в отдельном блоке допускается только одноразовый ввод массива в оперативную память ЭВМ.

Рассмотрим несколько отличающуюся от общепринятой формализованную структуру данных. Под показателем будем понимать кортеж реkvизитов, определяющий количественную характеристику экономической сути. Каждый показатель характеризует один количественный реkvизит (основание показателя) и адресная часть — качественные реkvизиты (признаки). Кроме показателей, выделим кортежи реkvизитов, имеющие определенный смысл, но состоящие лишь из качественных реkvизитов — фасеты. Они как показатели могут не иметь специального названия. Примером фасета может быть сочетание трех реkvизитов «Код оборудования», «Код детали», «Номер технологической операции». Этот фасет указывает, на каком оборудовании выполняется данная операция по производству детали и, следовательно, характеризует ее в определенном аспекте, т. е. дает качественную характеристику сути. Аналогично фасет составляет запись массива данных о рабочих днях предприятия. При машинной обработке экономической информации с показателем можно производить арифметические действия, а с фасетами — структурные.

Кортеж реkvизитов, составляющих показатель или фасет, можно определить как один более общий элемент — параметр. Один или несколько параметров, имеющих общую адресную часть, образуют запись

(фразу) массива. В нотации Бэкуса элементы экономической информации можно определить следующим образом:

(параметр) : = (показатель) | (фасет),  
 (показатель) = (количественный реквизит) (адресная часть),  
 (фасет) :: = (качественный реквизит) (адресная часть),  
 (адресная часть) = (качественный реквизит) | (адресная часть)  
 (качественный реквизит).

В адресной части параметра различаем левую часть, в которую входит главный ключ упорядочения и могут входить также и другие, непосредственно примыкающие к нему ключи упорядочения, что можно записать следующим образом:

(левая часть) :: = (главный ключ упорядочения) | (левая часть)  
 (количественный реквизит),  
 (главный ключ упорядочения) = (качественный реквизит).

Перейдем к более строгому определению процесса построения схем потоков данных задач. Для этого будем считать, что множества элементов схем потоков данных образуют формальный язык этих схем и построим грамматику, порождающую такой язык.

Граматику схем потоков данных зададим указанием следующих объектов: конечного множества символов, образующих основной словарь, конечного множества вспомогательных символов, конечного множества правил, задаваемых с использованием нотации Бэкуса.

Основные символы используются для обозначения реквизитов. Количественные реквизиты будем обозначать заглавными буквами (кроме буквы Р), а качественные — строчными. Тогда массив, содержащий два основания показателей А, В, основание фасета с и имеющий два ключа упорядочения i и j может быть представлен следующим образом:

A B c [ij],

где скобки указывают, что i и j являются ключами упорядочения.

Вспомогательные символы используются для специальных обозначений и не подлежат преобразованию (обработке). К ним относятся символы:

[ ] — используемые для выделения упорядочения,  
 { } — используемые для обозначения элементов множества,  
 + × / — знаки арифметических операций,  
 Σ — знак суммирования.

Все правила грамматики можно разделить на две группы. К первой группе относятся правила об арифметических преобразованиях, которые, как показывает обследование схем потоков данных конкретных задач АСУП, могут быть четырех видов: сложение (в алгебраическом смысле), умножение, деление и суммирование. Мы сознательно выделили суммирование как многократное повторение действия сложения, так как в нем происходит последовательная обработка показателя одного наименования, расположенного в отдельных записях одного массива в противовес сложению, где совместно обрабатываются показатели из разных массивов.

Правило сложения записывается следующим образом:

$$\langle A [i] \rangle = \langle A [i] \rangle + \langle A [ij] \rangle | \langle A [i] \rangle + \langle B [i] \rangle, \quad (1)$$

где (i) = (левая часть),  
 (j) = (реквизит) | (j) (реквизит).

Правило умножения:

$$\langle A [ij] \rangle = \langle A [ij] \rangle \times \langle B [ij] \rangle | \langle A [ij] \rangle \times \langle B [i] \rangle | \langle B [i] \rangle \times \langle A [ij] \rangle. \quad (2)$$

Правило деления:

$$\langle A [ij] \rangle = \langle A [ij] \rangle / \langle B [ij] \rangle | \langle A [ij] \rangle / \langle B [i] \rangle. \quad (3)$$

Правило суммирования:

$$(A [ij]) : \Sigma (A [ij]). \quad (4)$$

Ко второй группе правил относятся правила структурных преобразований массивов. Под структурным преобразованием будем понимать изменение порядка в иерархии ключей упорядочения массива (пересортировка массива), замену ключей упорядочения новыми или дополнение системы ключей упорядочения новыми реквизитами, выведение из ключей упорядочения отдельных реквизитов, объединение массивов и разделение массивов.

Совокупность ключей упорядочения определенного массива составляет множество качественных реквизитов, т. е. множество с конечным числом элементов  $m: \{a\} = \{a_1, \dots, a_m\}$ , причем  $a_j \neq a_s$ , где  $(j \neq s)$ . Множество всех перестановок обозначим через  $P(a)$ , а любой элемент этого множества через  $\alpha_j^p$ , где  $j = 1, 2, \dots, m!$ . Тогда правило пересортировки можно записать следующим образом:

$$(\{A\} \{b\} [\alpha_j^p]) = (\{A\} \{b\} [\alpha_s^p]), \text{ где } j \neq s. \quad (5)$$

Естественно, что множества  $\{A\}$  и  $\{b\}$  частично могут быть и пустыми.

Массив может быть упорядочен по новым ключам, которых не было в первоначальной структуре массива. Рассмотрим следующий пример. Пусть имеется массив пооперационных трудовых нормативов, в котором имеются следующие основные реквизиты: код детали  $d$ , номер технологической операции  $S$ , код цеха  $c$ , код профессии  $p$ , трудоемкость  $A$ . Массив упорядочен по возрастанию  $d$  и  $S$ . Массив «месячный план выпуска деталей» состоит из реквизитов: код детали  $d$ , код цеха  $c$ , месячный план  $N$ . Для определения месячной трудоемкости плана необходима совместная обработка массивов, а это возможно только тогда, когда массивы будут упорядочены по возрастанию  $c$  и  $d$ . Поэтому необходимо главным ключом первого массива сделать  $c$ , а второстепенным  $d$ . В общем случае некоторые основы фасетов могут перейти в разряд ключей упорядочения и наоборот. Правило замены можно формально представить следующим образом:

$$(\{A\} \{b\} [\alpha_j^p]) = (\{A\} \{a\} [\beta_s^p]), \quad (6)$$

где  $\alpha_j^p \in P(a)$ ,

$\beta_s^p \in P(\beta)$ ,

$\{b\} \cup \{a\} = \{a\} \cup \{b\}$ .

Правило (5) является частным случаем (6).

Объединение массивов возможно, если у них системы ключей упорядочения одинаковы, поэтому правило объединения можно записать следующим образом:

$$(\{D\} \{e\} [ij]) = (\{A\} \{b\} [ij]) (\{B\} \{c\} [ij]), \quad (7)$$

где  $\{A\} \cup \{B\} = \{D\}$ ,

$\{b\} \cup \{c\} = \{e\}$ .

Правило разъединения является обратным правилу объединения и может быть представлено так:

$$(\{A\} \{b\} [ij]) (\{c\} \{d\} [ij]) = (\{E\} \{f\} [ij]), \quad (8)$$

где  $\{A\} \cup \{c\} = \{E\}$ ,

$\{b\} \cup \{d\} = \{f\}$ .

Применение этих восьми правил дает возможность формализовать составление схем потоков данных, автоматизировать этот процесс, разработать алгоритмы определения оптимальных схем потоков данных.

Вильнюсский государственный  
университет им. В. Калпакаса  
Кафедра экономической  
информации

Редколлегии вручено  
в апреле 1976 г.

## UŽDAVINIŲ DUOMENŲ SRAUTŲ APRĄŠYMO KONTEKSTINĖ GRAMATIKA

A. BARTKUS, L. SIMANAUSKAS

### Reziumė

Straipsnyje nagrinėjami kai kurie teoriniai uždavinių sprendimo procesų projektavimo automatizavimo klausimai. Parodoma, kad duomenų apdorojimo procesą galima suprasti kaip tų duomenų judėjimą iš vienu masyvų į kitus. Perduodant duomenis, su jais galima atlikti įvairias operacijas, t. y. jie gali būti papildomai apdorojami. Todėl prieš sudarant detalius algoritmus, tikslinga sudaryti bendrą duomenų srautų schemą. Tąsioje schemoje nurodoma, kokių nuoseklumu apdorojami masyvai ir kokios operacijos su jais atliekamos. Taigi visas sprendimo procesas suskaidomas į atskirus blokus.

Konkrečių AVS programų tyrimas rodo, kad visas operacijas galima suskirstyti į dvi grupes: aritmetines operacijas ir masyvų struktūros keitimo operacijas. Aritmetinės operacijos yra šios: algebrinės sudėties, daugybos, dalybos ir sumavimo; masyvų struktūros keitimo operacijos — tai rūšiavimas, rūšiavimo raktų eilės papildymas, masyvų sujungimas ir masyvų išskyrimas.

Šios operacijos sudaro gramatikos, generuojančios duomenų srautų schemas, taisykles, kurios straipsnyje pateiktos naudojant Bekuso notaciją.

Pagal šias gramatikos taisykles galima sugeneruoti duomenų srautų schemas, realizuojančias duotąją priklausomybę tarp pradinių ir rezultatinų rodiklių.