

# Алгебраический подход к операциям на элементах базы данных

С. Лебедева

База данных представлена в качестве формализованной системы; представлено постулаты этой системы, подано определения операций на данных и примеры применения этих операций. Показано практическое применение представленной системы для конкретной задачи: многостепенной идентификации объекта. Указывается связь между алгебраическими операциями на данных и инструкциями Языка Манипулирования Данными. Операции на данных сравниваются с операциями на отношениях в реляционных базах данных.

§

## 1. Введение

Во время разработки базы данных для многостепенного эксперимента [1] появилась необходимость формализации некоторых проблем, связанных с базой данных и манипулированием данными. База данных для многостепенного эксперимента и способ её использования имеют ряд особенностей. Во-первых, база данных имеет численный характер: элементами базы данных являются двумерные матрицы, векторы, системы векторов и отдельные числа. Во-вторых, элементами базы данных являются регулярные структуры [5] или же структуры, полученные при помощи операций на данных из регулярных структур. В-третьих, в процессе эксперимента экспериментатор может принять решение об увеличении числа измерений, или же может появиться необходимость использования измерений, полученных в других лабораториях, что повлечёт за собой не только увеличение количества данных, но также необходимость объединения данных, находящихся в разных физических областях.

Полученные результаты имеют довольно общий характер и справедливы для любых баз данных с иерархическими и сетевыми структурами.

## 2. Язык. Первичные понятия и постулаты

Обозначим символом  $N$  счётное множество имён данных базы данных. Элементы множества имён будем обозначать символами  $n_1, n_2, n_3, \dots$ , элементы множества значений-символами  $v, v_1, v_2, \dots$ . В множестве  $V$  выделим некоторое подмножество  $V_S$  структурных значений. Предполагаем, что на множестве  $N$  определено отображение  $f$ , значения которого принадлежат мно-

жеству  $V$ . Множество всех упорядоченных пар  $(n, v)$  выполняющих условие  $n \in N, v \in V$  и  $v = f(n)$  будем называть *данными*, множество всех данных обозначим символом  $D$ . Элементами множества  $D$  будем обозначать символами  $d, d_1, d_2, \dots$ . Символами  $\cdot, /, \#$  обозначим некоторые операции, определённые на множестве имён. Операции  $\cdot, /, \#$  позволяют получать новые имена из имён, принадлежащих множеству  $N$ . Предполагаем, что на множестве имён определено отношение частичного упорядочения  $<$ . Символами  $\sim, \wedge$  и  $\Rightarrow$  будем обозначать логические связки: отрицание, конъюнкцию и импликацию соответственно, символ  $\forall$  обозначает квантор общности, символ  $\exists$  — квантор существования. Угловые скобки  $\langle$  и  $\rangle$  обозначают последовательность данных. Свойства базы данных  $\mathcal{D}$  описывают постулаты P1—P14 [3].

$$P1 \quad \forall (n \in N) \exists (v \in V) \{f(n) = v\}$$

$$P2 \quad \forall (n_1, n_2) \{n_1 = n_2 \Rightarrow f(n_1) = f(n_2)\}$$

$$P3 \quad \forall (n_1, n_2 \in N) \{n_1 \cdot n_2 \in N\}$$

$$P4 \quad (n_1 \cdot n_2) \cdot n_3 = n_1 \cdot (n_2 \cdot n_3)$$

$$P5 \quad \forall (n_1, n_2 \in N) \{n_1 \# n_2 \in N\}$$

$$P6 \quad \forall (n_1, n_2 \in N) \{n_1/n_2 \in N\}$$

$$P7 \quad f(n_1 \cdot n_2) = \langle f(n_1), f(n_2) \rangle$$

$$P8 \quad f(n_1 \# n_2) = \langle (n_1, f(n_1)), (n_2, f(n_2)) \rangle$$

$$P9 \quad (n, v) \in D \Leftrightarrow n \in N \wedge v \in V \wedge f(n) = v$$

$$P10 \quad d \in V_S$$

$$P11 \quad v_1, v_2 \in V_S \Rightarrow \langle v_1, v_2 \rangle \in V_S$$

$$P12 \quad d = (n, v) \wedge v = \langle d_1, \dots, d_k \rangle \wedge d_i \in v \Rightarrow d_i < d$$

$$P13 \quad d_1 < d_2 \wedge d_2 < d_3 \Rightarrow d_1 < d_3$$

$$P14 \quad \sim (d < d)$$

В силу постулатов P1—P2 на множестве имён определено отображение  $f$  значения которого принадлежат множеству значений  $V$ . Постулаты P3—P6 определяют некоторые отношения на именах. Операцию  $\cdot$  будем называть *соединением имён*, операцию  $/$  — *вычитанием имён*, операцию  $\#$  — *конструкцией имён*.

Постулаты P7 и P8 определяют отображение  $f$  для соединения и конструкции имён соответственно. Постулат P9 является необходимым и достаточным условием принадлежности данного множеству данных базы данных. Постулаты P10—P11 описывают свойства множества структурных значений, из постулатов P12—P14 следует, что на элементах данных базы данных определено отношение частичного упорядочения  $<$ . Отношение  $<$  будем называть отношением предшествования. Из постулатов P9, P1 и P2 сейчас же следует.

$$\text{Следствие 2.1.} \quad d_1 = (n_1, v_1) \wedge d_2 = (n_2, v_2) \Rightarrow (n_1 = n_2 \Rightarrow v_1 = v_2).$$

Следствие 2.1 гарантирует, что в базе данных нет двух данных, имена которых идентичны, а значения разные.

Данные, значения которых принадлежат множеству структурных значений  $V_S$ , будем называть сложными данными, данные, значения которых принадлежат множеству  $V/V_S$  — данными элементарными. Множество сложных данных обозначим символом  $D_S$ , множество элементарных данных — символом  $D_E$ .

### 3. Операции на элементах базы данных

Основными операциями на элементах базы данных являются следующие операции: операция  $\beta$  извлечения значения данного, операция  $\delta$  экстракции или извлечения данного из некоторого подмножества данных операция  $\oplus$  конструкции данных [5], операция  $\odot$  конкатенации (соединения данных), операция  $\sum$  ограниченного вычитания данных, операция  $\gamma$  изменения значения данного [3, 4].

**Определение 3.1.** Пусть  $d=(n, v)$  — данное, где  $n$  — имя, а  $v$  — значение данного  $d$ . Тогда

$$\beta(n, v) = v.$$

Операция извлечения значения данного  $\beta$  ставит в соответствие каждому данному его значение. Выполнимость операции извлечения значения гарантирует постулат P1, однозначность операции следует из следствия 2.1. Заметим, что результат операции извлечения значения, вообще говоря, не данное, а элементарное значение или последовательность данных. Например, допустим, что данные  $d_1=(\text{измерение } 3,1.05)$  и  $d_2=(\text{координаты точки } a, \langle(\text{координата } x, 3.5), (\text{координата } y, 5.0)\rangle)$ . Тогда  $\beta(d_1)=1,5$ ;  $\beta(d_2)=\langle(\text{координата } x, 3.5), (\text{координата } y, 5.0)\rangle$ .

Операция  $\delta$  экстракции или извлечения данного из некоторого подмножества данных базы данных позволяет получить данное  $d=(n, v)$ , если нам известно имя этого данного и имя подмножества данных, элементом которого является данное  $d$ . Операция  $\delta$  зависит от двух аргументов: первым аргументом является имя  $n$  данного  $d$ , вторым аргументом — имя подмножества  $X$ .

**Определение 3.2.** Пусть  $n$  — имя данного  $d=(n, v)$ ,  $X$  — имя подмножества данных базы данных, которому принадлежит данное  $d$ . Тогда

$$\delta(n, X) = (n, v).$$

Операция  $\delta$  определена на основании постулата P1, однозначность операции следует из следствия 2.1. Пусть  $X$  будет именем множества векторов  $e_1=(\text{век } 1, \langle(x_1, 1), (y_1, 2)\rangle)$ ,  $e_2=(\text{век } 2, \langle(x_2, 3), (y_2, 1.5)\rangle)$ ,  $e_3=(\text{век } 3, \langle(x_3, 5), (y_3, 2.5)\rangle)$ , имена этих векторов соответственно век 1, век 2, век 3, значениями являются последовательности данных  $\langle(x_1, 1), (y_1, 2)\rangle$ ,  $\langle(x_2, 3), (y_2, 1.5)\rangle$ ,  $\langle(x_3, 5), (y_3, 2.5)\rangle$  где  $x_i$  и  $y_i$  — имена компонент вектора  $\text{век}_i$ ,  $i=1, 2, 3$ . Тогда

$$\begin{aligned} \delta(\text{век } 1, X) &= (\text{век } 1, \langle(x_1, 1), (y_1, 2)\rangle) \\ \delta(\text{век } 2, X) &= (\text{век } 2, \langle(x_2, 3), (y_2, 1.5)\rangle) \\ \delta(\text{век } 3, X) &= (\text{век } 3, \langle(x_3, 5), (y_3, 2.5)\rangle). \end{aligned}$$



Выполнимость операции  $\hat{\#}$  гарантируют постулаты P5, P10 и P11, однозначность — следствие 2.1. Предположим, что в базе данных хранятся матрицы являющиеся входными и выходными измерениями некоторого эксперимента. Имена этих матриц соответственно — «входные измерения» и «выходные измерения». В результате операции конструкции данных мы получим данное, имя которого — «входные измерения  $\#$  выходные измерения», а значение — последовательность матриц входных и выходных измерений. Многократное применение операции конкатенации и конструкции к элементам базы данных делает возможным создание новых множеств данных. Множество данных можно интерпретировать как данное, значением которого является последовательность элементов этого множества. Операция конструкции сохраняет отношение предшествования.

Операцией обратной относительно операции конструкции является операция ограниченного вычитания  $\hat{\sim}$ . Перед определением этой операции введём понятие отношения непосредственного предшествования  $\hat{<}$ .

**Определение 3.5.**  $d_1 \hat{<}^* d_2 \Leftrightarrow d_1 \hat{<}^* d_2 \wedge \sim \exists (d) \{d_1 \hat{<} d \wedge d \hat{<}^* d_2\}$  например, если  $d = (n, \langle d_1, d_2 \rangle)$ , то  $d_1 \hat{<}^* d$  и  $d_2 \hat{<}^* d$ .

Теперь можно определить операцию ограниченного вычитания данных:

**Определение 3.6.** Если  $d = (n, v)$ ,  $v = \langle d_1, d_2, \dots, d_n \rangle$  и  $d_i = (n_i, v_i)$ ,  $i = 1, 2, \dots, n$ , то

$$d \hat{\sim} d_i = (n/n_i, \langle d_1, d_2, \dots, d_{i-1}, d_{i+1}, \dots, d_n \rangle).$$

Заметим, что операция  $\hat{\sim}$  определена не для всех пар данных, а только для таких пар  $d_k, d_i$ , для которых выполнено условие  $d_i \hat{<}^* d_k$ . Выполнимость и однозначность операции гарантируют постулаты P1, P2, P7, P11 и P9. Покажем действие операции  $\hat{\sim}$  на данных из предыдущего примера. Пусть именами данных  $d, d_1$  и  $d_2$  будут соответственно имена «входные измерения  $\#$  выходные измерения», «входные измерения», «выходные измерения». Тогда результаты операции  $d \hat{\sim} d_2$  будет данное  $d_1$ , имя которого «входные измерения» а значение — матрица значений входных измерений. Операция  $\hat{\sim}$  является избыточной операцией, данное  $d_1$ , может быть получено из данного  $d$  при помощи операции экстракции. Вообще, каждое данное полученное при помощи операции ограниченного вычитания может быть получено при помощи суперпозиции операции экстракции, конструкции и конкатенации. Эта операция введена только для удобства пользователя. Операция  $\gamma$  изменения значения данного присваивает данному  $d = (n, v)$  новое значение  $v_1$ .

**Определение 3.7.** Если  $(n, v) \in D$  и  $v_1 \in V$  то

$$\gamma((n, v), v_1) = (n, v_1).$$

Операция изменения значения может изменить структуру базы данных. Чтобы этого избежать, можно потребовать, чтобы старое и новое значения были одинаковы в структурном отношении, в случае, когда оба значения определены. Постулаты P1—P14 не предусматривают случая, когда значение данного не определено. Но такая ситуация может иметь место. Например, пользо-

ватель вводит в базу данных имена матриц, информацию относительно числа строк и столбцов. Система управления базой данных резервирует место для записи значений. До момента введения данного в базу данных, значение данного не определено. Чтобы учесть эту ситуацию допустим существование выделенного элемента  $\Phi$  множества значений  $V$ .

P15  $\Phi \in V$ .

Если  $f(n) = \Phi$ , то значение данного не определено. Результат операций конкатенации и конструкции для случая, когда значение одного из аргументов не определено, определяют формулы (3.1)—(3.2)

$$(n_1, v_1) \odot (n_2, \Phi) = (n_2, \Phi) \odot (n_1, v_1) = (n_1, v_1) \quad (3.1)$$

$$(n_1, v_1) \# (n_2, \Phi) = (n_2, \Phi) \# (n_1, v_1) = (n_1, v_1) \quad (3.2)$$

Результатом конкатенации или конструкции любого данного  $d$  с данным, значение которого не определено, является данное  $d$ , следовательно данное, значение которого не определено является нейтральным элементом относительно операций конкатенации и конструкции. Частным случаем операции изменения значения является операция *присваивания значения* данным  $\gamma^*$ . Операция  $\gamma^*$  определяется формулой

$$\gamma^*((n, \Phi), v) = (n, v). \quad (3.3)$$

В результате операции присваивания значения значение данного становится определённым.

#### 4. Практическое применение

Система базы данных, удовлетворяющая постулатам P1—P15 была разработана для специальной задачи: многостепенного эксперимента (многостепенной идентификации объекта). Отличительной чертой базы данных для многостепенного эксперимента является её динамический характер: число данных в базе данных постоянно возрастает [1, 2]. Экспериментатору может понадобиться матрица измерений, проводимых в разное время и записанных в разных физических областях, эту потребность удовлетворяют операции конкатенации и конструкции. Экспериментатору может понадобиться только часть матрицы входных (или выходных) измерений, это требование выполняется при помощи операции ограниченного вычитания или суперпозиции операций экстракции и конструкции. Заметим, что операция ограниченного вычитания была введена только для удобства пользователя. Система Управления Базой Данных (СУБД) резервирует место для данных пользователя и содержит процедуры, являющиеся реализациями перечисленных операций.

Доступ до базы данных и операции на данных и множествах данных реализует Язык Манипулирования Данными (ЯМД). Инструкции ЯМД можно поделить на две группы: инструкции типа WRITE осуществляющие запись информации в базу данных и инструкции типа READ, осуществляющие введение информации, находящейся в базе данных, в оперативную память [4]. Инструкции типа WRITE являются реализациями операций присваивания значения данным и изменения значения данных. Инструкции типа READ

реализуют операции извлечения значения данного, суперпозицию операций экстракции и извлечение значения, операции конкатенации, конструкции, ограниченного вычитания и суперпозиции этих операций. Описанные операции гарантируют независимость данных. Пользователь должен знать только имя данного и имя множества, которому принадлежит это данное, физическая организация базы данных и физические адреса данных пользователю неизвестны.

Описанная система базы данных является действующей системой, она внедрена в Лаборатории Технической Механики Вроцлавского политехнического института на ЭВМ ODRA-1325.

## 5. Заключительные замечания

Представленные операции не исчерпывают всех операций, которые можно производить на элементах базы данных. Важный класс операций составляют операции выборки данных, выполняющие некоторое логическое условие, например, нужно выбрать все измерения, значения которых находятся в определенном интервале и т. п. Логические условия обычно зависят от конкретного применения. В настоящем сообщении операции, реализующие выборку данных по заданному критерию, не обсуждаются.

В заключение несколько слов о сравнении представленной модели базы данных с реляционной моделью Кодда. Известно, что базы данных с древовидными и простыми сетевыми структурами можно преобразовать в реляционную базу данных (с некоторой избыточностью). Заметим, что при таком преобразовании в общем случае не сохраняется отношение порядка. Нетрудно заметить, что существует и обратный переход от реляционной модели к сетевой структуре. Действительно, элементами реляционной базы данных являются двумерные таблицы, поэтому реляционные базы данных выполняют постулаты P1—P14 при интерпретации отношения (таблицы) и строки таблицы как сложного данного и элемента таблицы  $(A, g)$  где  $A$  — имя атрибута,  $g$  — конкретное значение атрибута) как данного элементарного. Основные операции реляционной базы данных — *объединение отношений* и *проекция* выполнимы в представленной модели. Операция объединения отношений выполняется при помощи операции конкатенации, операция проекции — при помощи экстракции (если нам нужен один столбец отношения) или же при помощи суперпозиции операций экстракции, конструкции и конкатенации.

Представленные постулаты непротиворечивы, существует простая интерпретация этих постулатов в теории множеств. Представленная модель базы данных однозначна (результаты операций однозначны), полна в том смысле, что при помощи конечного числа операций может быть получено любое элементарное или структурное данное, а также замкнута. Замкнутость состоит в том, что результаты всех операций на элементах базы данных принадлежат базе данных.

ВРОЦЛАВСКИЙ ПОЛИТЕХНИЧЕСКИЙ ИНСТИТУТ  
ИНСТИТУТ ТЕХНИЧЕСКОЙ КИБЕРНЕТИКИ  
ЛАБОРАТОРИЯ СИСТЕМ УПРАВЛЕНИЯ  
ПОЛЬША

### Литература

- [1] BUBNICKI, Z., On the multistage identification, *Systems Sci.*, v. 3, No 2, 1977.
- [2] LEBIEDIEWA, S., Rola bazy danych w procesie sterowania eksperymentem wielostopniowym, *Prace Naukowe ICT Politechniki Wrocławskiej, Studia i Materiały*, NR 46, 1978.
- [3] Лебедева, С., Аксиоматический подход к операциям на элементах базы данных, 2 Symposium Grundlagen und Anwendung der Informationsverarbeitung, Technische Hochschule Karl-Marx-Stadt, Tagungsberichte, 1978.
- [4] LEBIEDIEWA, S., Rola bazy danych i Języka Manipulacji Danymi w procesie sterowania eksperymentem wielostopniowym, ICT Politechniki Wrocławskiej, Raport 279, Wrocław, 1978.
- [5] TURSKI, W., *Struktury danych*, WNT, Warszawa, 1971.

(Поступило 3-ого декабря 1981 г.)