ФОРМАЛИЗАЦИЯ МОДЕЛЕЙ ОБРАБОТКИ ДАННЫХ

В.И. Мунерман
Смоленский государственный университет
vimoon@gmail.com, vim@munerman.ru
910-7887980, 481-2-607980

Рассматриваются вопросы

- Анализ одного класса алгебраических концептуальных моделей.
- Выявление общности в описании данных и операций над ними.
- Построение метатеории, позволяющей исследовать общие свойства моделей.
- Доказательство возможности однозначной трансляции запросов с языка одной модели на языки других моделей.

Иерархия моделей

Модель данных

Концептуальная модель

Модель вычислений

Реляционная алгебра	Алгебра файлов	Алгебра многомерных матриц
неявная операция	*сортировка*	*транспонирование*
selection	*выборка*	*m-кратное сечение*
projection	*сжатие*	*свертка*
теоретико-множественные операции	*слияние строго упорядоченных файлов*	*сложение*
join	*слияние нестрого упорядоченных файлов*	(λ, μ)-*свернутое произведение*

Примеры концептуальных моделей их рассматриваемого класса

SELECT <список полей> FROM R UNION SELECT список полей> FROM S

SELECT R.a1, R.a2 FROM R,S WHERE R.a1=S.b1 AND R.a2=S.b2

СЛИЯНИЕ СТРОГО УПОРЯДОЧЕННЫХ ФАЙЛОВ: строго упорядоченный файл Z_K получается в результате слияния строго упорядоченных файлов X_K и Y_K, если:

–все три файла строго упорядочены по одному и тому же множеству ключей K;

–классы эквивалентности файла Z_K задаются соотношением

$$Z_{K^*} = f(X_{K^*}, Y_{K^*})$$

Объединение в алгебре файлов

$$f(X_{K^*}, Y_{K^*}) = \begin{cases} Y_{K^*}, \textit{ если } X_{K^*} = \Theta_{K^*} \\ X_{K^*}, \textit{ в противном случае} \end{cases}$$

Объединение в алгебре многомерных матриц

$$C_{i_1 \dots i_p} = A_{i_1 \dots i_p} \vee B_{i_1 \dots i_p}$$

Метатеория обработки баз данных дает возможность обсуждения свойств формальных систем, в том числе точное их описание (определение правил образования и преобразования) и исследование относящихся к ним результатов, производить в некоторой другой теории, которая называется ***метатеорией***. Сами же формальные системы называются ***предметными*** (или ***объектными***) теориями.

Рассмотрим теорию **Т**, в которой множество символов состоит из:

1. цифр, строчных и заглавных букв $a, b, \dots, A, B, \dots$ или $a_1, a_2, \dots, A_1, A_2, \dots$;
2. специальных символов двух классов:
   a. связки: (,), <запятая>, $=, \neq, \in, \subset, \wedge, \forall, \exists, \to$;
   b. знаки операций: $'$, λ, σ, $\oplus$, $\otimes$, **К**.

Термы (***объекты***) теории **Т** определяются как буквы или последовательности букв.

Формулы (***соотношения***) теории **Т** определяются следующим образом:

1. Любой терм есть формула;
2. $(A_1, \dots, A_n)$, $(a_1, \dots, a_n)$ – формулы;
3. $B \subset A$, $a \in A$, $(a_1, \dots, a_n) = (b_1, \dots, b_n)$, $(a_1, \dots, a_n) \neq (b_1, \dots, b_n)$ – формулы;
4. $\forall\ (a_1, \dots, a_n)$, $\exists\ (a_1, \dots, a_n)$, $f(A_1, \dots, A_n)$ – формулы;
5. $R(A_1, \dots, A_n)$, $sR(A_1, \dots, A_n)$, $\lambda R(A_1, \dots, A_n)$, $\sigma R(A_1, \dots, A_n)$, **К** $(A_1, \dots, A_k)R(A_1, \dots, A_k, A_{k+1}, \dots, A_n)$, $(R_1(A_1, \dots, A_n) \oplus R_2(A_1, \dots, A_n))$, $(R_1(A_1, \dots, A_n) \otimes R_2(A_1, \dots, A_n))$ – формулы;
6. Если F, F_1, F_2 – формулы в смысле п.5, то $F_1 = F_2$, $F_1 \neq F_2$, sF, λF, σF, **К** $(A_1, \dots, A_k)F$, $F_1 \oplus F_2$, $F_1 \otimes F_2$ – формулы;
7. Других формул нет.

Примеры аксиом

A$^{\sigma}$.

$$R(A_1,\dots, A_k, A_{k+1},\dots, A_n) \wedge \mathsf{K} \quad (A_1,\dots, A_k) \rightarrow$$

$$\rightarrow {}^{\sigma}R(A_1,\dots, A_k, A_{k+1},\dots, A_n) = R(A_1,\dots, A_k, f(A_{k+1},\dots, A_n))$$

A$^{\oplus}$.

$$^{\mathsf{K}} \ (A_1,\dots, A_k)R_1(A_1,\dots, A_k, B_{k+1},\dots, B_m) \wedge$$

$$^{\mathsf{K}} \ (A_1,\dots, A_k)R_2(A_1,\dots, A_k, C_{k+1},\dots, C_n) \wedge$$

$$\forall(a_1,\dots, a_k)(((a_1,\dots, a_k, b_{k+1},\dots, b_m) = (a_1,\dots, a_k, b'_{k+1},\dots, b'_m)) \wedge$$

$$((a_1,\dots, a_k, c_{k+1},\dots, c_n) = (a_1,\dots, a_k, c'_{k+1},\dots, c'_n))) \rightarrow$$

$$\rightarrow (R_1(A_1,\dots, A_k, B_{k+1},\dots, B_m) \oplus R_2(A_1,\dots, A_k, C_{k+1},\dots, C_n)) =$$

$$R(A_1,\dots, A_k, f(B_{k+1},\dots, B_m, C_{k+1},\dots, C_n)))$$

Теория Т	Алгебра файлов	Алгебра многомерных матриц	Реляционная алгебра
$A(a_1, \ldots, a_n)$	$F_K(a_1, \ldots, a_n)$ файл с записью $(a_1, \ldots, a_n)$ и множеством ключей K	A многомерная матрица с индексами $i_1, \ldots, i_n$	$R(a_1, \ldots, a_n)$ отношение со схемой $(a_1, \ldots, a_n)$
$'A$	*Sort* F_K сортировка файла A_K	$'A$ транспонирование матрицы A	*Index* R упорядочивание отношения R

λA	*Sel* F_K выборка из файла A_K	${}^{\lambda}A$ сечение матрицы A	*Select R* реляционная операция селекции
σA	$\sum_{K'} F_K$ сжатие файла F_K по подмножеству ключей K' множества ключей K	${}^{\mu}A$ – свертка матрицы A	*Select group R* реляционная операция селекции с группировкой

$\oplus AB$	$Ms\ F_K G_K$ слияние строго упорядоченных файлов	$A+B$ сложение многомерных матриц	$\cup AB,\ \cap AB,\ \backslash AB$ любая из теоретико-множественных операций
$\otimes AB$	$Mns\ F_K G_L$ слияние нестрого упорядоченных файлов	$\underset{\times}{\underline{(\lambda,\ \mu)}} AB$ произведение многомерных матриц	*Join RQ* – операция соединения отношений R и Q

Теоремой в теории **T** называется конструкция вида: **F** $= R(A_1, \ldots, A_p)$,
где **F** – формула теории **T**, в смысле п.5 определения формул.

Вывод: *запрос к базе данных – это формула, выводимая из аксиом или* ***теорема****,* ***в теории*** **T**.

Теорема об эквивалентности теорий
(Н. Бурбаки, Теория множеств).

Пусть **T** – теория, **A** $_1$, …, **A** $_n$ – ее явные аксиомы, $a_1, \ldots, a_h$ – ее константы, $T_1, \ldots, T_h$ – ее термы. Если $(T_1|a_1) \ldots (T_h|a_h)$ **A** $_i$, (i =1, ..., n) являются теоремами теории **T** ', и знаки теории **T** являются знаками теории **T** ', и схемы теории **T** являются схемами теории **T** ', то если **A** теорема теории **T** , то $(T_1|a_1) \ldots (T_h|a_h)$ **A** есть теорема теории **T** '.

$(T_1|a_1) \ldots (T_h|a_h)$ – подстановки термов $T_1, \ldots, T_h$ вместо констант $a_1, \ldots, a_h$ в формулах **A** $_i$, (i =1, ..., n) и **A** .

Следствие.

Запрос в одной модели обработки данных однозначно транслируется в запрос другой модели.

Средства доказательства изоморфизма моделей

Абстрактные типы данных (АТД), определяются как многоосновные алгебраические системы.

(В.М.Глушков, Е.Г.Цейтлин, Е.Л.Ющенко Алгебра.Языки. Программирование)

Абстрактный тип данных – математическая модель с совокупностью операторов, определенных в рамках этой модели.

(А.Ахо, Дж. Хопкрофт, Дж. Ульман Структуры данных и алгоритмы)

Среди множества произвольных АТД выделяется один специфический вид – двухосновные алгебраические системы $\boldsymbol{E}$=<{S, T}; Ω; П>.

Основа S – ***структура***, основа T – ***тип***.

Важная особенность этих АТД состоит в том, что суть операций над элементами структуры S не изменяется при изменении сути операций над элементами типа T.

Результаты

На основе анализа алгебраических моделей обработки баз данных выявлена общность в описании данных и операций над ними.

Построена метатеория, позволяющая исследовать общие свойства моделей.

Доказана возможность однозначной трансляции запросов с языка одной модели на языки других моделей.