О ТИПИЗАЦИИ ИЕРАРХИЧЕСКИХ СТРУКТУР ДАННЫХ

Г.В. Чернышев

ФГБУН Институт информатики и проблем регионального управления КБНЦ РАН,
И. Арманд 37а, 360000 Нальчик, Россия chern_gen@mail333.com

Алгебраические методы, с учетом опыта их применения в теории баз данных реляционного типа [1,2], оказываются полезными в исследованиях информационных структур и других типов. В данной работе предлагается подход к типизации одного класса иерархических структур, основанный на теоретико-категорной характеристике элементов иерархии.

Понятие структуры данных является фундаментальным понятием, во многом определяющим построение алгоритмов обработки данных. Важнейшими свойствами структур данных, влияющих на качество алгоритмов, являются их однородность и регулярность.

Однородность структуры связывают с совокупностью используемых в ней (типов) элементов, а регулярность — с повторяемостью связей между ними (существованием закономерностей). Естественным представлением совокупности объектов, структуры каждого объекта, является иерархическое представление.

Будем рассматривать корневые деревья, структура которых содержит регулярности специального вида, в связи с чем будем называть их иерархическими структурами. Иерархические структуры представляем категорией путей $\mathbb{C}_{\mathrm{P}}$ [3], которая содержит:

1. класс объектов $\mathbf{O}_{\mathrm{P}} = \{v_1, v_2, \ldots, v_i, \ldots\}$ с выделенным объектом v_{R};

2. для каждой упорядоченной пары (v_i, v_j), $v_i, v_j \in \mathbf{O}_{\mathrm{P}}$ — множество морфизмов вида $\mathbf{M}_{\mathrm{P}}\,(v_i, v_j) = \{\langle v_i, v_{i_1}, \ldots v_{i_k}, v_j\rangle \,|\, \{v_i, v_{i_1}, \ldots v_{i_k}, v_j\} \subseteq \mathbf{O}_{\mathrm{P}}\}$, такое, что $|\mathbf{M}_{\mathrm{P}}\,(v_i, v_j)| = 1$ (поэтому $\mathbf{M}_{\mathrm{P}}(a, b)$ — единственный морфизм между объектами a и b);

3. для каждой упорядоченной тройки объектов (v_i, v_j, v_k) — функцию (композицию)

   $$\mathbf{M}_{\mathrm{P}}\,(v_j, v_k) \circ \mathbf{M}_{\mathrm{P}}\,(v_i, v_j) \to \mathbf{M}_{\mathrm{P}}\,(v_i, v_k)\,,$$

   задаваемую правилом (конкатенация) $\langle v_i, \ldots, v_j\rangle \cdot \langle v_j, \ldots, v_k\rangle = \langle v_i, \ldots, v_j, \ldots, v_k\rangle$;

4. для каждого $v \in \mathbf{O}_{\mathrm{P}}$ единичный морфизм (единицу) $\mathbf{1}_v = \mathbf{M}_{\mathrm{P}}(v, v) = \langle v\rangle$.

Путь между объектами v_i и v_j в категории $\mathbb{C}_{\mathrm{P}}$ — это последовательность объектов в представлении морфизма $\mathbf{M}_{\mathrm{P}}\,(v_i, v_j)$.

Важнейшее понятие категории путей — характеристический терминальный путь χ_{v^t}, интерпретируется как последовательность вершин дерева от терминальной вершины v^t к v_R. Совокупность всех характеристических путей однозначно определяет $\mathbb{C}_{\mathrm{P}}$. Для χ-путей вводятся понятия обратного пути, делителя, общего и наиболее общего делителя.

Разработка теоретико-категорных положений для иерархических структур ([4]) привела к описанию регулярностей, позволяющих обосновать введение для этих структур схемы. Схема, по сути, является типом соответствующей иерархической структуры, что позволяет все действия над иерархическими структурами и их составляющими выражать через действия со схемой.

Для категории $\mathbb{C}_{\mathrm{P}}$ схемой называется категория $\mathbb{S}_{\mathrm{P}}$, такая, что каждый терминальный χ-путь из $\mathbb{C}_{\mathrm{P}}$ изоморфен ровно одному терминальному χ-пути из $\mathbb{S}_{\mathrm{P}}$.

Любой нетерминальный χ-путь χ_v однозначно определяет множество χ-путей, для которых χ_v является наибольшим общим делителем: $\mathcal{S}(\chi_v) = \{\chi_{v_i} = \langle v_i, v, \ldots, v_R\rangle \mid i \in I_v\}$, где

I_v — индексное множество для первых компонентов этих путей. При этом, χ_v назовем χ-путем типа *структура*, а элементы множества $\{\mathbf{M}_{\mathrm{P}}(v_i, v) \mid i \in I_v\}$ — *элементами структуры*. Здесь каждый объект структуры v_i может определять, либо терминальный χ-путь, либо χ-путь типа структура (структурный тип).

Введенные понятия структуры и ее элементов являются достаточными для представления произвольного дерева, в котором структурный тип характеризует нетерминальные вершины, а характеристистические терминальные пути — висячие вершины. Данные типы являются основными «структурообразующими» элементами как для категории $\mathbb{C}_{\mathrm{P}}$, так и для ее схемы $\mathbb{S}_{\mathrm{P}}$.

Любое подмножество χ-путей категории $\mathbb{C}_{\mathrm{P}}$ образует подкатегорию этой категории. Множество изоморфных подкатегорий $\mathcal{R}(\mathbb{C}_{\mathrm{P}}) = \left\{\mathbb{C}^i_{\mathrm{P}} \mid \mathbb{C}^i_{\mathrm{P}} \cong \mathbb{C}^j_{\mathrm{P}},\ \mathbb{C}^i_{\mathrm{P}}, \mathbb{C}^j_{\mathrm{P}} \subseteq \mathbb{C}_{\mathrm{P}},\ i \neq j\right\}$ категории $\mathbb{C}_{\mathrm{P}}$ назовем *регулярностью* в $\mathbb{C}_{\mathrm{P}}$. Регулярность в категорию $\mathbb{C}_{\mathrm{P}}$ вносят структуры. Изоморфными в структуре являются объекты, определяющие терминальные χ-пути.

Определим еще один вид регулярности. Если χ_{v_i} для фиксированного v_i не является терминальным χ-путем, то χ_{v_i} будет наибольшим общим делителем некоторого множества χ-путей, которое образует подкатегорию категории $\mathbb{C}_{\mathrm{P}}$. Обозначим эту подкатегорию, для каждого такого v_i, через $\mathbb{C}^{v_i}_{\mathrm{P}}$. Если все подкатегории из этой совокупности изоморфны друг другу, т.е. $\forall i, j \in I_v\ (i \neq j)$, $\mathbb{C}^{v_i}_{\mathrm{P}} \cong \mathbb{C}^{v_j}_{\mathrm{P}}$, то χ_v назовем χ-путем типа *массив*, а элементы множества $\left\{\mathbb{C}^{v_i}_{\mathrm{P}} \mid i \in I_v\right\}$ — *элементами массива*, где каждый v_i определяет, либо терминальный χ-путь, либо χ-путь типа массив. На самом деле, тип массива (как объект) можно ввести только в $\mathbb{S}_{\mathrm{P}}$.

Основные соотношения между категорией $\mathbb{C}_{\mathrm{P}}$, ее схемой $\mathbb{S}_{\mathrm{P}}$ и их подкатегориями $\mathbb{C}'_{\mathrm{P}} \subseteq \mathbb{C}_{\mathrm{P}}$ и $\mathbb{S}'_{\mathrm{P}} \subseteq \mathbb{S}_{\mathrm{P}}$ иллюстрирует следующая коммутативная диаграмма:

$$\begin{tikzcd}
\mathbb{S}_{\mathrm{P}} \arrow[r, shift left, "\mathsf{F}_{cons}"] \arrow[d, "\mathsf{F}_{pr}"'] & \mathbb{C}_{\mathrm{P}} \arrow[l, shift left, "\mathsf{F}_{abs}"] \arrow[d, "\mathsf{F}_{flt}"] \\
\mathbb{S}'_{\mathrm{P}} \arrow[r, shift left, "\mathsf{F}'_{cons}"] & \mathbb{C}'_{\mathrm{P}} \arrow[l, shift left, "\mathsf{F}'_{abs}"]
\end{tikzcd}$$

где F_{abs} — функтор абстрагирования, задающий типизацию категории $\mathbb{C}_{\mathrm{P}}$ посредством категории $\mathbb{S}_{\mathrm{P}}$, F_{cons} — функтор конкретизации, задающий конструктивное представление категории $\mathbb{S}_{\mathrm{P}}$ категорией $\mathbb{C}_{\mathrm{P}}$, F'_{cons} и F'_{abs} — аналогичные функторы, но действующие на соответствующих подкатегориях, F_{pr} — функтор проекции, F_{flt} — функтор фильтрации.

Смыл функтора проекции заключается в выделении подсхем в схеме $\mathbb{S}_{\mathrm{P}}$, при этом согласованное с ним действие функтора фильтрации выделяет соответствующую подкатегорию $\mathbb{C}'_{\mathrm{P}}$ в категории $\mathbb{C}_{\mathrm{P}}$. Согласование заключается в определении так называемого расширенного естественного преобразования $\alpha : \mathsf{F}_{pr} \Rightarrow \mathsf{F}_{flt}$, которое позволит функтору F_{flt} выделить в $\mathbb{C}_{\mathrm{P}}$ конкретную подкатегорию $\mathbb{C}'_{\mathrm{P}}$, изоморфную подсхеме $\mathbb{S}'_{\mathrm{P}}$.

Литература

1. Плоткин Б.И. *Универсальная алгебра, алгебраическая логика и базы данных.* М.: Наука, 1991.

2. Бениаминов Е.М. *Алгебраические методы в теории баз данных и представлении знаний.* М.: Научный мир, 2003.

3. Чернышев Г.В. *Теоретико-категорное описание иерархических структур* // Материалы Второго Международного Российско-Узбекского симпозиума "Уравнения смешанного типа и родственные проблемы анализа и информатики". Нальчик: Издательство КБНЦ РАН, 2012. С. 278-280.

4. Чернышев Г.В. *Теоретико-методологические основы типизации иерархических структур* // Известия КБНЦ РАН, №5, 2013. С. 21–28.