Доклады Академии наук СССР
1974. Том 216, № 2

УДК 51:801 МАТЕМАТИКА

В. Ю. МЕЙТУС

АБСТРАКТНОЕ ЗАДАНИЕ СИНТАКСИСА ФОРМАЛЬНЫХ ЯЗЫКОВ

(Представлено академиком В. М. Глушковым 25 IX 1973)

1. В настоящей заметке описывается алгебраический способ задания синтаксиса широкого класса формальных языков. Этот класс включает в себя контекстно-свободные языки, языки, порождаемые параллельными грамматиками ([1], [2]), и подкласс языков, порождаемых синхронизируемыми грамматиками ([3]).

2. Предлагаемый способ задания синтаксиса включает описание нескольких формальных категорий и связывающих их функторов.

Универсальная категория $\mathfrak{M}$ задается следующим образом. Ее множество объектов $\mathscr{A}_{\mathfrak{M}}$ состоит из чисел (с индексами или без) вида $\bar{m}=\prod_{i=1}^{m}\mathbf{p}(i)^{a_i}$, где a_i — целое неотрицательное число, а $\mathbf{p}(i)$ — i-тое простое число. По предположению, категория конечных множеств является подкатегорией категории $\mathfrak{M}$. Если $\bar{m}, \bar{n}\in\mathscr{A}_{\mathfrak{M}}$, $\bar{n}=\prod_{j=1}^{n}\mathbf{p}(j)^{b_j}$, то через $\mathscr{H}_{\mathfrak{M}}(\bar{m}, \bar{n})$ обозначим множество морфизмов объекта $\bar{m}$ в $\bar{n}$. $\mathscr{H}_{\mathfrak{M}}(\bar{m}, \bar{n})$ — это множество всех функций φ, сопоставляющих каждому числу $p(i)^{a_i}$ число $\prod_{b_j\in A_i}\mathbf{p}(j)^{b_j}$, $\bigcup_{i=1}^{m}A_i=\{b_j|1\leqslant j\leqslant n\}$. При этом некоторые множества A_i могут быть пустыми.

Пусть $\mathfrak{T}$ — множество типов, $\mathfrak{T}=\{\tau_i|i\geqslant 0\}$. Категория типов $\mathbf{T}$ задается множеством объектов $\mathscr{A}_{\mathbf{T}}=\{t_{n,i}|1\leqslant i\leqslant 2^k, k\geqslant 1\}$ и множествами морфизмов $\mathscr{H}_{\mathbf{T}}(t_m, t_n)$ для всех объектов $t_m, t_n\in\mathscr{A}_{\mathbf{T}}$ (индекс i опущен). Каждый объект t_n — это конечное подмножество множества $\hat{t}_n$, состоящего из всех возможных наборов длины n, которые можно построить из $\mathfrak{T}$, $t_n=\{t_n'\}$.

Морфизм $\theta\in\mathscr{H}_{\mathbf{T}}(t_m, t_n)$ каждому $t_m'\in t_m$ сопоставляет некоторое подмножество множества t_n. Для различных t_m' и t_m'' соответствующие подмножества t_n' и t_n'' могут быть одинаковыми. Морфизм θ можно представить как совокупность таких функций $\{\theta^{(i)}\}$, что если $(t_m', t_n')\in\theta$, то найдется единственная функция $\theta^{(i)}$, определенная на каждом $\tau_k\in t_m'$ и сопоставляющая ему подмножество $P_k=\{\tau_{j_1}, \ldots, \tau_{j_r}\}\subseteq t_n'$ и $\bigcup_{k=1}^{m}P_k=t_n'$. При этом структура всех функций $\theta^{(i)}$ одинакова, т. е. если подстановка π_m преобразует множество t_m' в t_m'', заменяя элемент τ_k' на k-м месте в t_m' элементом τ_k'', стоящим в t_m'' также на k-ом месте, а π_n — аналогичная подстановка для элементов t_n', t_n'', то для функций $\theta^{(')}: t_m'\to t_n'$, $\theta^{('')}: t_m''\to t_n''$, принадлежащих $\{\theta^{(i)}\}$, справедливо равенство $\pi_n\theta^{(')}=\theta^{('')}\pi_m$, выражающее свойство коммутативности соответствующей диаграммы (см. рис. 1). Множество всех объектов $\{t_n\}$ и всех морфизмов θ для этих объектов задают категорию типов $\mathscr{T}$, для которой композиция морфизмов определяется следующим образом. Если θ: $t_k\to t_p$ и θ': $t_p\to t_q$, то $\theta\circ\theta'$: $t_k\to t_q$, $(\theta\circ\theta')(t_k^{(i)})=$ $=\theta'(\theta(t_k^{(i)}))$ с очевидным выполнением условия ассоциативности композиции.

3. Произвольный функтор T из подкатегории $\mathfrak{N}$ универсальной категории $\mathfrak{M}$ в категорию типов $\mathscr{T}$, определенный условиями $T\bar{n}=t_n$, $T(\bar{p}, \bar{q})(\varphi)=$

$=T\varphi=\theta\in H_{\mathscr{T}}(t_p, t_q)$ и структурой морфизма θ, совпадающей со структурой морфизма φ, называется синтаксическим функтором (S-функтором). Образ категории $\mathfrak{N}$ в $\mathscr{T}$, определенный функтором T, обозначим через $\mathscr{T}_T$.

Языковой S-схемой Λ, определенной S-функтором T и объектом $\bar{m}\in\mathscr{A}_{\mathfrak{N}}$, называется полная подкатегория $\mathscr{T}_\Lambda$ категории $\mathscr{T}_T$, множество объектов которой $\mathscr{A}_\Lambda=\{t_n | H_{\mathscr{T}}(t_m, t_n)=\emptyset \& t_n\in\mathscr{A}_{\mathscr{T}_T}\}$.

$$\begin{tikzcd} t_m' \arrow[r, dash, "\theta^{(I)}"] \arrow[d, dash, "\pi_m"'] & t_n' \arrow[d, dash, "\pi_n"] \\ t_m'' \arrow[r, dash, "\theta^{(II)}"] & t_n'' \end{tikzcd}$$

Рис. 1

Из категории $\mathscr{T}_\Lambda$ можно построить новую категорию $\hat{\mathscr{T}}_\Lambda$, называемую разложением категории $\mathscr{T}_\Lambda$. Объекты $\hat{\mathscr{T}}_\Lambda$ — это различные наборы t_n', каждый из которых принадлежит одному $t_n\in\mathscr{A}_\Lambda$. Если $t_n'\in t_n$, $t_q'\in t_q$, то для каждого $\theta\in H_{\mathscr{T}_\Lambda}(t_n, t_q)$ в множестве $\hat{H}(t_n', t_q')$ существует морфизм $\hat{\theta}$, совпадающий с θ, ограниченным на пару (t_n', t_q'). Множество $\hat{\mathscr{A}}_\Lambda$ объектов категории $\hat{\mathscr{T}}_\Lambda$ называется расширенным языком, порождаемым S-схемой Λ; это множество обозначается через L_{ext}.

Пусть $\Sigma=(\Sigma_1, \ldots, \Sigma_N)$ — конечный набор подмножеств множества типов $\mathfrak{T}$. S-языком ΣL_{ext}, определяемым множеством Σ и расширенным языком L_{ext} называется такое подмножество множества L_{ext}, для которого выполнены условия: 1) если $t\in\Sigma L_{\text{ext}}$, то $\hat{H}(t, t')=\emptyset$ для всех t', отличных от t из L_{ext}; 2) $t=(a_{k_0}, \ldots, a_{k_1-1}, a_{k_1}, \ldots, a_{k_2}, \ldots, a_{k_N})$, где $\{a_{k_{i-1}}, \ldots, a_{k_i-1}\}\in\Sigma^*$, $1\leqslant i\leqslant N$, причем для некоторых Σ_i допускается, чтобы наборы $\{a_{k_{i-1}}, \ldots, a_{k_i-1}\}$ были пустыми.

4. Для проблем, связанных с анализом языковых структур, чрезвычайно важным является изучение системы морфизмов, связывающих объекты расширенного языка. Для выделения морфизмов вводится специальный функтор D_T, который называется S-анализатором. D_T — это ковариантный функтор из категории $\hat{\mathscr{T}}_\Lambda$ в категорию множеств, сопоставляющий каждому объекту $t_n\in\hat{\mathscr{A}}_\Lambda$ объект $D_T(t_n')=d_n'=\bigcup\limits_{t_m'\in t_m}\hat{H}(t_m', t_n')$, а каждому морфизму ω: $t_n'\to t_q'$ — морфизм $D_T(\omega)$: $d_n'\to d_q'$, $D_T(\omega)(\theta)=\theta\circ\omega$. Тождественному морфизму сопоставляется тождественный морфизм, а композиции $\omega\circ\omega'$ — композиция $D_T(\omega)\circ D_T(\omega')$. Образ категории $\hat{\mathscr{T}}_\Lambda$ при отображении D_T обозначим через $\mathfrak{S}_\Lambda$. Категорию $\mathfrak{S}_\Lambda$ также можно разложить и получить категорию $\hat{\mathfrak{S}}_\Lambda$. Функтор P^η: $\mathfrak{S}_\Lambda\to\hat{\mathfrak{S}}_\Lambda$, сопоставляющий каждому d_n морфизм $\theta_n\in d_n$, а каждому морфизму $\omega\in H(d_n, d_q)$ — морфизм $P^\eta(\omega)$: $\theta_n\to\theta_q$, где $\theta_q=\theta_n\circ\omega$, называется представляющим функтором S-анализатора D_T, определяющим представление η.

5. Введенные выше понятия обобщают целый ряд известных подходов к заданию синтаксиса формальных языков. Рассмотрим следующие классы формальных языков: конечно-автоматные, обобщенно-автоматные ([4]) контекстно-свободные, параллельные ([1]), синхронизируемые ([3]). В частности, как показано в работе ([3]), синхронизируемые языки включают в себя класс контекстно-связанных языков.

Пусть грамматика G порождает язык $L(G)$, принадлежащий одному из перечисленных выше классов. Тогда для грамматики G и языка $L(G)$ справедливы следующие результаты.

Теорема 1. *По грамматике G, порождающей язык $L(G)$, можно построить подкатегорию $\mathfrak{N}_G$ универсальной категории $\mathfrak{M}$, языковую S-схему $\mathscr{T}_G$, S-функтор T_G: $\mathfrak{N}_G\to\mathscr{T}_G$ и множество Σ_G такие, что S-язык $\Sigma_G L_{\text{ext}}$ совпадает с языком $L(G)$.*

Теорема 2. *Для каждого S-функтора T_G существует представимый S-анализатор D_{T_G} с представляющим функтором P_G^η.*

Используя понятие вычислимых категорий и функторов, можно доказать следующие результаты.

Теорема 3. *Для любой грамматики G, принадлежащей определенным выше классам, категории $\mathfrak{R}_G$, $\mathcal{T}_G$, $\hat{\mathcal{T}}_G$, $\mathfrak{S}_G$, $\hat{\mathfrak{S}}_G$ и функторы T_G, D_{TG}, P_G^{η} вычислимы.*

Теорема 4. *Для любой грамматики G по вычислимому заданию функторов D_{TG} и P_G^{η} можно построить автомат, который перерабатывает любое слово x в терминальном алфавите грамматики G в морфизм $\theta_x \in H(t_m', x)$ тогда и только тогда, когда $x \in L(G) = \Sigma_G L^{G}_{\mathrm{ext}}$.*

Институт кибернетики
Академии наук УССР
Киев

Поступило
23 VII 1973

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] *J. Kràl,* Kybernetica, № 1 (1969). [2] *В. Ю. Мейтус,* Кибернетика, № 6 (1972). [3] *В. Ю. Мейтус,* Сборн. Теоретич. киберн., в. 1, Киев, 1970. [4] *J. Thatcher, J. Wright,* Math. Syst. Theory, v. 2, № 1 (1968).