ՀԱՅԿԱԿԱՆ ՍՍՀ ԳԻՏՈՒԹՅՈՒՆՆԵՐԻ ԱԿԱԴԵՄԻԱՅԻ ԶԵԿՈՒՅՑՆԵՐ
ДОКЛАДЫ АКАДЕМИИ НАУК АРМЯНСКОЙ ССР

LXXX 1985 4

УДК 681.142.2

МАТЕМАТИКА

Г. В. Джулакян

Восходяще-анализируемые подклассы индексных языков

(Представлено чл.-корр. АН Армянской ССР Р. Р. Варшамовым 3/IV 1984)

Понятие индексных языков введено в ([1]) и определено с помощью порождающего механизма индексных грамматик. В работах ([2-4]) введены классы гнездных стековых автоматов, многоуровневых магазинных автоматов и индексных магазинных автоматов соответственно. Односторонние недетерминированные гнездные стековые автоматы, магазинные автоматы уровня 2 и индексные магазинные автоматы допускают в точности класс индексных языков. В данной работе определяется класс нового типа автоматов (D_M-автомат) и выделяются подклассы индексных языков, для которых возможно построить детерминированный D_M-автомат. Существенной особенностью D_M-автомата является возможность моделирования восходящих алгоритмов разбора цепочек индексных языков, тогда как автоматами из работ ([2-4]) моделируются нисходящие алгоритмы разбора. Частные модели анализаторов индексных языков, работающих восходящим методом, введены в работах ([5,6]).

Индексной грамматикой называется пятерка $G=(N, \sum, F, P, S)$, где N, $\sum$ и F конечные множества нетерминалов, терминалов и индексов соответственно, P—конечное множество правил вида $x\to\chi$, где либо $x\in N$, $\chi\in(NF\cup N\cup\sum)^*$, либо $x\in NF$, $\chi\in(N\cup\sum)^*$, причем для каждого $f\in F$ существует хотя бы одно правило вида $Af\to\chi$. Пусть α, $\beta\in(NF^*\cup\sum)^*$. Говорят, что в грамматике G из α непосредственно выводится β, если либо

1) $\alpha=\gamma A\xi\delta$, $(A\to X_1\eta_1 \dots X_k\eta_k)\in P$ и $\beta=\gamma X_1\theta_1 \dots X_k\theta_k\delta$, где $\theta_l=\varepsilon$ при $X_l\in\sum$, $\theta_l=\eta_l\xi$ при $X_l\in N$, $1\leqslant l\leqslant k$, либо

2) $\alpha=\gamma Af\xi\delta$, $(Af\to X_1 \dots X_k)\in P$ и $\beta=\gamma X_1\theta_1 \dots X_k\theta_k\delta$, где $\theta_l=\varepsilon$ при $X_l\in\sum$, $\theta_l=\xi$ при $X_l\in N$, $1\leqslant l\leqslant k$.

В этих обозначениях $f\in F$, $\xi\in F^*$, $\eta_l=\varepsilon$ при $X_l\in\sum'$, $\eta_l\in F\cup\{\varepsilon\}$ при $X_l\in N$, γ, $\delta\in(NF^*\cup\sum)^*$. Непосредственный вывод в G обозначается $\Rightarrow$. Рефлексивное и транзитивное замыкание отношения $\Rightarrow$ над множеством $(NF^*\cup\sum)^*$ называется выводом и обозначается $\Rightarrow^*$. Языком, порождаемым индексной грамматикой G, называется множество $L(G)=\{w/S\Rightarrow^+w,\ w\in\sum^*\}$.

D_M-автомат состоит из управляющего устройства с конечным числом состояний и тремя ввод-выводными головками и работает над памятью, представляющей собой совокупность трех магазинных лент, одна из которых является входной, а в течение работы становится также рабочей лентой, вторая является рабочей лентой и во многом по своей структуре похожа на гнездный стек, а третья лента служит для хранения индексов. D_M-автомат работает над алфавитами: Σ—множество входных элементов, N—множество вспомогательных символов, F—множество индексов, $\Gamma_0=NF\cup N\cup F\cup\Sigma\cup\{\bar{\bar{s}}, \bar{\bar{c}}, \perp\}$—множество рабочих элементов, где $\bar{\bar{s}}$, $\bar{\bar{c}}$, $\perp$ специальные символы, служащие граничными маркерами над рабочей и входной лентами. Ниже приводится формальное определение D_M-автомата. Для краткости в некоторых обозначениях положим: $\Gamma_1=N\cup NF$; $\Gamma_2=\Gamma_1\cup\Sigma\cup\cup\{\perp\}$; $A_1=(\Gamma_2/\{\perp\})^+$; $A_2=A_1\{\bar{\bar{c}}\}F^*$; $A_3=A_2(\{\bar{\bar{s}}\}A_2)^*$; $A_4=A_3(\{\bar{\bar{s}}\}A_1)$;

$E=\{\perp, \bar{\bar{c}}\}F^*\{\bar{\bar{s}}\}(\Gamma_2/\{\perp\})^+$.

Формально D_M-автомат — это семерка $(Q, \Sigma, \Gamma_0, F, \delta, q_0, S)$, где Q—конечное непустое множество состояний управляющего устройства, Σ—конечное множество входных символов, Γ_0—конечное непустое множество рабочих элементов, F—конечное множество индексов, $q_0\in Q$—начальное состояние управляющего устройства, $S\in\Gamma_0$—специальный символ, показывающий успешное завершение работы D_M—автомата, $\delta=(\delta_1\vee\delta_2\vee\delta_3\vee\delta_4)$—функция переходов управляющего устройства, где (а) δ_1—отображение множества $(Q\times\Gamma_2\times\Sigma)$ в конечное подмножество множества $(Q\times\Sigma)$; (б) δ_2—отображение множества $(Q\times\Gamma_2\times F^*\times\Gamma_1F^*\{\bar{\bar{c}}\})$ в конечное подмножество множества $(Q\times\Gamma_1\times\times F^*)$; (в) δ_3—отображение множества $(Q\times\Gamma_2\times F^*\times(\Gamma_1F^*\{\bar{\bar{c}}\}\cup\Sigma))$ в конечное подмножество множества $(Q\times\{\bar{\bar{c}}\}F^*\{\bar{\bar{s}}\}(\Gamma_1\cup\Sigma)\times F^*)$; (г) δ_4—отображение множества $(Q\times E\times F^*\times(\Sigma\cup\{\perp\}))$ в конечное подмножество множества $(Q\times\{\perp, \varepsilon\}\times F^*\times\Gamma_1F^*\{\bar{\bar{c}}\})$ и может быть еще $\delta_4(q_0, \perp, \varepsilon, \perp)=(q_0, \perp, \varepsilon, S\perp)$.

Мгновенное состояние D_M-автомата называется конфигурацией. Она определяется состоянием управляющего устройства, содержаниями входного, рабочего и индексного магазинов. Таким образом, конфигурация D_M-автомата — это четверка $(q, \perp\alpha, \xi, \beta_1\beta_2\perp)$, где $q\in Q$, $\xi\in F^*\alpha,\in A_1\cup A_2\cup A_3\cup A_4,\beta_1\in(NF^*\{\bar{\bar{c}}\}\cup\{\varepsilon\})$, $\beta_2\in\Sigma^*$.

Определим над множеством конфигураций отношение перехода $\vdash$. В зависимости от значения текущей конфигурации D_M-автомата и значения функции переходов δ возможны следующие виды переходов*:

1) $(q, \perp\alpha_1X, \xi, Y\beta_2\perp)\vdash(p, \perp\alpha_1XY, \xi, \beta_2\perp)$**, если множество $\delta_1(q, X, Y)$ содержит элемент (p, Y);
2) $(q, \vdash\alpha_1X, \xi, X_1\xi\bar{\bar{c}}\beta_2\perp)\vdash(p, \perp\alpha_1XX_1, \xi, \beta_2\perp)$, если множество $\delta_2(q, X, \xi, X_1\xi\bar{\bar{c}})$ содержит элемент (p, X_1, ξ);
3) $(q, \perp\alpha_1X, \xi, X_1\xi_1\bar{\bar{c}}\beta_2\perp)\vdash(p, \perp\alpha_1X\bar{\bar{c}}\xi\bar{\bar{s}}X_1, \xi_1, \beta_2\perp)$, если множество $\delta_3(q, X, \xi, X_1\xi_1\bar{\bar{c}})$ содержит элемент $(p, \bar{\bar{c}}\xi\bar{\bar{s}}X_1, \xi_1)$, $(q, \perp\alpha_1X, \xi, Y\beta_2\perp)\vdash(p, \perp\alpha_1X\bar{\bar{c}}\xi\bar{\bar{s}}Y, \varepsilon, \beta_2\perp)$, если множество $\delta_3(q, X, \xi, Y)$ содержит элемент $(p, \bar{\bar{c}}\xi\bar{\bar{s}}Y, \varepsilon)$;

* В этих обозначениях приняты следующие соглашения: $\alpha_1\in A_1\cup A_4\cup\{\varepsilon\}$; $\alpha_2\in(\{\varepsilon\}\cup\cup\{\&\}A_2)^*A_1$; $\alpha_3\in A_1$; $X\in\Gamma_2/\{\perp\}$; $X_1\in\Gamma_1$; $Y\in\Sigma$.

** В случае, когда на лентах остается только символ $\perp$, переходы определяются аналогичным образом.

4) $(q, \perp\alpha_2\bar{\bar{c}}\xi\bar{\bar{s}}\alpha_3X, \xi_1, Y\beta_2\perp) \vdash (p, \perp\alpha_2, \xi, X_1\xi_1\bar{\bar{c}}Y\beta_2\perp)$, если множество $\delta_4(q, \bar{\bar{c}}\xi\bar{\bar{s}}\alpha_3X, \xi_1, Y)$ содержит элемент $(p, \varepsilon, \xi, X_1\xi_1\bar{\bar{c}})$.

Транзитивное замыкание отношения $\vdash$ над множеством конфигураций обозначим через $\vdash^+$, а транзитивное и рефлексивное замыкание $\vdash^*$. Будем говорить, что $D_И$-автомат допускает цепочку $w\in\Sigma^*$, если он, начиная свою работу с начальной конфигурации $(q_0, \perp, \varepsilon, w\perp)$, может переходить в заключительную конфигурацию $(p, \perp, \varepsilon, S\perp)$ для некоторого $p\in Q$, т. е. имеет место отношение $(q_0, \perp, \varepsilon, w\perp) \vdash^* (p, \perp, \varepsilon, S\perp)$.

Языком, допускаемым $D_И$-автоматом M, называется множество всех допустимых цепочек, которое обозначается через $T(M)$:

$$T(M) = \{w/(q_0, \perp, \varepsilon, w\perp) \vdash^+ (p, \perp, \varepsilon, S\perp), \text{ где } p\in Q, w\in\Sigma^*\}.$$

Пусть $G=(N, \Sigma, F, P, S)$ индексная грамматика и пусть $A\in N\cup \cup NF$; $B, C\in N$; $X, Y\in(NF\cup N\cup\Sigma)$; $f\in F$; $f_1, f_2\in F\cup\{\varepsilon\}$, $\xi, \xi'\in F^*$; $\alpha, \beta, \gamma, \gamma_1, \gamma_2, \gamma_3\in(NF^*\cup\Sigma)^*$.

Следующие бинарные отношения, определенные над множеством $N\cup NF\cup\Sigma\cup F$, называются отношениями предшествования:

1) $X \doteq Y$, если существует некоторое правило $(A\to\alpha XY\beta)\in P$;
2) $X \cdot\!> Y$, если существует правило $(A\to\alpha Bf_1Cf_2\beta)\in P$, хотя бы один из выводов $B\Rightarrow^+\gamma X$, $Bf_1\Rightarrow^+\gamma_1 X$ и хотя бы один из выводов $C\Rightarrow^* Y\gamma_2$, $Cf_2\Rightarrow^* Y\gamma_3$;
3) $X <\!\cdot Y$, если существуют правило $(A\to\alpha XBf_1\beta)\in P$ и хотя бы один из выводов $B\Rightarrow^+ Y\gamma$, $Bf_1\Rightarrow^+ Y\gamma_1$;
4) $X \,|\Xi|\, f$, если существует вывод $S\Rightarrow^+\alpha Xf\beta$ такой, что в $\alpha Xf\beta$ элементы X и f появились не одновременно с применением некоторого правила $B\to\alpha' Xf\beta'$.

Будем говорить, что G удовлетворяет F-условию, если для всех $X\in N$ одновременно не имеют место $(X, f)\in\{|\Xi|\}$ и $(A\to\alpha Xf\beta)\in P$.

Пусть $G=(N, \Sigma, F, P, S)$ некоторая индексная грамматика. Над множеством $(N\cup\Sigma\cup F)^*\cup\{\to\}$ определим гомоморфную функцию h такую, что $h(X)=X$, если $X\in(N\cup\Sigma\cup\{\to\})$, и $h(X)=\varepsilon$ в противном случае. Обозначим $P'=\{h(x\to\chi)/(x\to\chi)\in P\}$. КС-грамматику $BASE(G)= =(N, \Sigma, P', S)$ назовем базовой грамматикой для индексной грамматики G.

Пусть $G=(N, \Sigma, F, P, S)$ приведенная грамматика (отсутствуют бесполезные нетерминалы, выводы вида $A\Rightarrow^+A$ для $A\in N\cup NF$ и G грамматика без ε-правил). G называется:

1) индексной грамматикой простого предшествования, если а) грамматика $BASE(G)$ является КС-грамматикой простого предшествования, б) G удовлетворяет F-условию;

2) индексной грамматикой предшествования, если а) G обратима, б) $\{<\!\cdot\}\cap\{\cdot\!>\}\cup\{\doteq\}\cap\{\cdot\!>\}=\varnothing$, в) когда $\{<\!\cdot\}\cap\{\doteq\}=\{(X_i, Y_i)\}$, где $i=1, 2, \ldots$, то $X_i, Y_i\in N\cup NF$ и для всех сентенциальных форм $\rho_i= =\alpha_iX_i\xi_{i_1}Y_i\xi_{i_2}\beta_i$ из условия $\xi_{i_1}=\xi_{i_2}$ вытекает, что в ρ_i символы X_i и Y_i появились одновременно применением некоторого правила $A\to\alpha_1'X_iY_i\beta_1'$; г) если в G имеются правило $A\to\alpha BC\beta$ и хотя бы одно из правил $B\to w_1$, $C\to w_2$ таких, что $w_1, w_2\in\Sigma^+$ и $(A, f)\in\{|\Xi|\}$ для некоторого $f\in \in F$, то $(B, C)\notin\{<\!\cdot\}$, д) G удовлетворяет F-условию;

3) LR индексной грамматикой предшествования, если а) $FIRST_0^k(A) \cap FIRST_0^k(B) = \varnothing$ для всех тех $A, B \in N \cup NF$, для которых существуют правила вида $A \to \alpha$ и $B \to \alpha$, б) G удовлетворяет условиям б)—д) пункта 2).

Пусть $\perp$ некоторый символ, не принадлежащий множеству $N \cup \cup NF \cup \Sigma$. Для всех $X, Y \in (N \cup NF \cup \Sigma)$, если $S \Rightarrow_G^* X\alpha$, то положим $\perp \lessdot X$, и если $S \Rightarrow_G^* \beta Y$, то положим $Y \gtrdot \perp$. В последующих теоремах $a_i \in \Sigma$; $X_i \in N \cup NF$; $Y_i \in N \cup NF \cup \Sigma$; $f_i \in F \cup \{\varepsilon\}$; $\xi \in F^*$.

Теорема 1. *Пусть* $\alpha X_1 f_1 \xi_1 X_2 f_2 \xi_2 \beta$ *некоторая сентенциальная форма в грамматике* G. *Тогда существуют* $Y_1 \in \{X_1, X_1 f_1\}$ *и* $Y_2 \in \{X_2, X_2 f_2\}$ *такие, что между ними имеет место хотя бы одно из отношений предшествования* 1—3.

Теорема 2. *Если* $\perp S \perp \Rightarrow_{rG}^n X_p X_{p-1} f_{p-1} \xi_{p-1} \ldots X_{k+1} f_{k+1} \xi_{k+1} A \xi a_1 \ldots a_q \Rightarrow_{rG} X_p X_{p-1} f_{p-1} \xi_{p-1} \ldots X_{k+1} f_{k+1} \xi_{k+1} Y_k \xi_k \ldots Y_1 \xi_1 a_1 \ldots a_q$, *где* $X_p = a_q = \perp$, *и в последнем шаге вывода применилось правило* $A \to Y_k \ldots Y_1$, *то существуют* $Y_l \in \{X_l, X_l f_l\}$, $p-1 \geqslant l \geqslant k$, *такие, что*

(1) $Y_{l+1} \lessdot Y_l$ *или* $Y_{l+1} \doteq Y_l$ *для* $p-1 \geqslant l > k$ *и* $Y_{k+1} \lessdot Y_k$ *и*

(2) $Y_{l+1} \doteq Y_l$ *для* $k > l \geqslant 1$ *и* $Y_1 \gtrdot a_1$.

Теорема 3. *Пусть* $G = (N, \Sigma, F, P, S)$ *некоторая индексная грамматика. Тогда можно эффективно построить* D_M*-автомат* $M = (Q, \Sigma, \Gamma_0, F, \delta, q_0, S)$ *такой, что* $L(G) = T(M)$.

Схема доказательства. D_M-автомат получается из G с помощью следующих построений: $Q = \{q_0\}$, $\Gamma_0 = N \cup NF \cup \Sigma \cup F \cup \{\bar{\bar{s}}, \bar{\bar{c}}, \perp\}$, а функция δ определяется с помощью отношений предшествования (предполагается, что множества $\{\doteq\}$, $\{\lessdot\}$, $\{\gtrdot\}$ и $\{|\Xi|\}$ для грамматики G известны) по следующей схеме:

(а) $\delta_1(q_0, Z, \xi, a) = (q_0, Za, \xi, \varepsilon)$, если $Z \doteq a$;

(б) $\delta_2(q_0, Z, \xi_1, Af\xi_2\bar{\bar{c}}) = (q_0, ZA, \xi)$, если $Z \doteq A$ и

(1) $\xi_1 = f\xi_2 = \xi \neq \varepsilon$, $A|\Xi|f$, или

(2) $\xi_1 = f\xi_2 = \xi = \varepsilon$, или

(3) $f\xi_2 = \varepsilon$, $\xi = \xi_1$, или

(4) $\xi_1 = \varepsilon$, $f\xi_2 \neq \varepsilon$, $A|\Xi|f$, $\xi = f\xi_2$;

(в) $\delta_3(q_0, Z, \xi_1, Af\xi_2\bar{\bar{c}}) = (q_0, Z\bar{\bar{c}}\xi_1\bar{\bar{s}}A, f\xi_2)$, если $Z \lessdot A$ и при $f\xi_2 \neq \varepsilon$, $A|\Xi|f$, $\delta_3(q_0, Z, \xi_1, a) = (q_0, Z\bar{\bar{c}}\xi_1\bar{\bar{s}}a, \varepsilon)$, если $Z \lessdot a$;

(г) $\delta_4(q_0, \perp\alpha Z, \xi, a) = (q_0, \perp, \varepsilon, A\xi\bar{\bar{c}})$, $\delta_4(q_0, \bar{\bar{c}}\xi_1\bar{\bar{s}}\alpha Z, \xi, a) = (q_0, \varepsilon, \xi_1, A\xi\bar{\bar{c}})$, если $Z \gtrdot a$ и $A \to \alpha Z$ некоторое правило в грамматике G. Если $\varepsilon \in L(G)$, то $\delta_4(q_0, \perp, \varepsilon, \perp) = (q_0, \perp, \varepsilon, S\perp)$.

В этих обозначениях $a \in \Sigma$; $Z \in \Gamma_2$; $A \in \Gamma_1$ $\alpha \in (\Gamma_2/\{\perp\})^*$; $f \in F \cup \{\varepsilon\}$.

Доказательство утверждения $L(G) = T(M)$ проводится индукцией по шагам вывода грамматики G и по шагам переходов D_M-автомата M.

Так как в общем случае отношения предшествования 1—3 определяют множества с непустыми пересечениями, правые части некоторых правил могут быть одинаковыми и определения префикса и суффикса начальной цепочки входного магазина выполняются неоднозначно, то построенный в теореме 3 D_M-автомат получается недетерминированным.

Теорема 4. *Пусть* $G = (N, \Sigma, F, P, S)$ *индексная грамматика. Если:* 1) G *индексная грамматика простого предшествования, или*

2) G индексная грамматика предшествования, или 3) G LR индексная грамматика предшествования, то для G можно построить детерминированный D_M-автомат $M=(Q, \sum, \Gamma_0, F, \delta, q_0, S)$ такой, что $L(G)=T(M)$.

В заключение автор выражает благодарность А. А. Ордяну за руководство работой.

Вычислительный центр
Академии наук Армянской ССР и
Ереванского государственного университета

Գ. Վ. ՋՈՒԼՀԱԿՅԱՆ

Ինդեքսային լեզուների ներքևից-վերև վերլուծվող ենթադասեր

Հոդվածում դիտարկվում է վերլուծության խնդիրը ինդեքսային քերականություններով տրվող լեզուների դասի համար։ Այդ նպատակով սահմանվում է ավտոմատների նոր դաս (D_M—ավտոմատ), որոնք ի տարբերություն ինդեքսային լեզուները ճանաչող հայտնի ավտոմատների, կարող են հանդիսանալ ինդեքսային լեզուների ներքևից-վերև վերլուծության ալգորիթմների մոդել։ Սահմանվում են նախորդման հարաբերություններ ինդեքսային քերականությունների համար և բերվում է տրված ինդեքսային քերականությունով ծնվող ինդեքսային լեզուն ճանաչող D_M—ավտոմատ կառուցելու եղանակ։ Ընդհանուր դեպքում կառուցվող D_M—ավտոմատն ստացվում է ոչ դետերմինացված։ Սահմանվում են ինդեքսային քերականությունների (լեզուների) ենթադասեր, որոնց համար հնարավոր է կառուցել դետերմինացված D_M—ավտոմատ։

ЛИТЕРАТУРА—ԳՐԱԿԱՆՈՒԹՅՈՒՆ

[1] *А. Ахо*, в сб.: Языки и автоматы, Мир, М., 1975. [2] *A. V. Aho*, J. ACM, v. 16, 383—406 (1969). [3] *А. Н. Маслов*, Проблемы передачи информации, № 1, 1976. [4] *R. Parchmann, J. Duske, J. Specht*, Inf. and Control, v. 45, 48—67 (1980). [5] *R. W. Sebesta, N. D. Jones*, Int. J. Comput. and Inf. Sci., v. 7, 345—359 (1978). [6] *А. А. Ордян, Г. В. Джулакян*, ДАН АрмССР, т. 69. № 1 (1979). [7] *А. Ахо, Дж. Ульман*, Теория синтаксического анализа, перевода и компиляции, т. 1, Мир, М., 1978.