Cybernetics and Systems Analysis, Vol. 39, No. 1, 2003

A CLASS OF FUNCTIONS COMPUTABLE BY INDEX GRAMMARS

L. P. Lisovik and T. A. Karnaukh UDC 519.716.35

The classes of functions IGF and IGF_0 are investigated that can be computed by means of special index grammars. The class IGF_0 is demonstrated to coincide with the class of strictly increasing functions (of a natural argument), which can be represented by systems of linear recurrence relations with natural coefficients, and to be closed with respect to the operations of addition, multiplication, series summation, and addition of a natural number to their arguments or results.

Keywords: *index grammar, computability, linear set, semilinear set, linear recurrence relations, closure property.*

Index grammars were first considered by A. Aho [1]. Index languages generated by them form a class of languages that is intermediate in the Chomsky hierarchy between the classes of context-free and context-dependent languages. It is well known [2] that the class of index languages coincides with the class of languages recognized by nested pushdown (stack) automata. Later on, it was shown in [3, 4] that any index language is recognized by a determined linearly bounded automaton.

In this article, we continue the investigation of index grammars from the viewpoint of their computational capabilities. We first consider an analog of the Parikh theorem for index languages and then pass to the consideration of the classes of functions *IGF* and IGF_0 that are computable and absolutely computable by index grammars. The class IGF_0 is demonstrated to coincide with the class of strictly increasing functions from N_+ into N_+ that are represented by systems of linear recurrence relations with natural coefficients and is closed with respect to the operations of addition, summation, multiplication, and addition of a natural constant to their argument or result.

We briefly recall the definition of an index grammar. An index grammar is understood to be a quintuple $G=(V,\Sigma,F,P,\sigma)$, where V is a finite set, Σ is its subset, $\sigma\,(\sigma\in\Delta)$ is the initial symbol, $\Delta=V\setminus\Sigma$ is the set of nonterminals, F is a finite set of indices (pointers), and P is a finite system of productions of the form $A\to\xi$, where $A\in\Delta$ and $\xi\in(\Delta F^*\cup\Sigma)^*$. Each index from the set F is a finite set of productions, which are called index production and are of the form $A\to\xi$, where $A\in\Delta$, $\xi\in V^*$.

On the set $(\Delta F^*\cup\Sigma)^*$, the relation of direct derivability $\Rightarrow$ is defined as follows.

Assume that $\xi_0=\alpha_0X_1\psi_1\alpha_1X_2\psi_2\alpha_2\ldots X_n\psi_n\alpha_n$ and $\xi_1=\alpha_0X_1\psi_1\theta\alpha_1X_2\psi_2\theta\alpha_2\ldots X_n\psi_n\theta\alpha_n$, where $\alpha_0,\alpha_1,\ldots,\alpha_n\in\Sigma^*$, $\psi_1,\psi_2,\ldots,\psi_n,\theta\in F^*$, and $X_1,X_2,\ldots,X_n\in\Delta$. Then we have

(1) if $\alpha A\theta\beta,\alpha,\beta\in(\Delta F^*\cup\Sigma)^*$ and the set P contains the production $A\to\xi_0$, then the relation $\alpha A\theta\beta\Rightarrow\alpha\xi_1\beta$ is fulfilled;

(2) if $\alpha Af\theta\beta,\alpha,\beta\in(\Delta F^*\cup\Sigma)^*$, $f\in F$, and the set f contains the index production $A\to\xi_0$, then the relation $\alpha A\theta\beta\Rightarrow\alpha\xi_1\beta$ is fulfilled.

An index grammar G generates an index language $L(G)=\{w\mid w\in\Sigma^*,\sigma\models\!\!\Rightarrow w\}$, where $\models\!\!\Rightarrow$ is the reflexive transitive closure of the relation $\Rightarrow$.

Taras Shevchenko University, Kiev, Ukraine. Translated from Kibernetika i Sistemnyi Analiz, No. 1, pp. 108-115, January-February 2003. Original article submitted November 27, 2001.

Let $\Sigma=\{a_1, a_2, \ldots, a_n\}$, where $n\in N_+$ and a_i $(i=\overline{1,n})$ are pairwise different symbols. We define a mapping $\varphi:\Sigma^* \to N^n$ as $\varphi(w)=(\varphi_{a_1}(w), \varphi_{a_2}(w), \ldots, \varphi_{a_n}(w))$, where $\varphi_x(w)$ is the number of occurrences of a symbol x in a word w. The mapping φ can be naturally extended to the set 2^{Σ^*}, assuming that $\varphi(L)=\{\varphi(w) | w\in L\}$, where $L\subseteq\Sigma^*$.

We will additionally use the definitions given below. Assume that $c\in N^n$ and $P=\{p_1, p_2, \ldots, p_k\}\subseteq N^n$; then, by definition, we have $L(c; P)=L(c; p_1, p_2, \ldots, p_k)=\{x\in N^n | x=c+\sum_{i=1}^{k}\alpha_i p_i, \ \alpha_i\in N, \ i=\overline{1,k}\}$. We call a set $M\subseteq N^n$ linear if $M=L(c; P)$ for some $c\in N^n$ and some finite set $P\subseteq N^n$. A set $M\subseteq N^n$ is semilinear if it is a finite union of linear sets.

By the Parikh theorem, the necessary condition that a language $L\subseteq a_1^* a_2^* \ldots a_n^*$ is context-free is the semilinearity of the set $\varphi(L)$; however, this condition is not sufficient [5]. For index languages, this condition is sufficient, which will be shown below.

LEMMA 1. If $L\subseteq a_1^* a_2^* \ldots a_n^*$ and $\varphi(L)=L(c; p_1, p_2, \ldots, p_k)$ is a linear set, then the language L is an index language.

Proof. We assume that $c=(c_1, c_2, \ldots, c_n)$ and $p_i=(p_{i,1}, p_{i,2}, \ldots, p_{i,n})$, $i=\overline{1,k}$. Let us consider an index grammar $G=(V, \Sigma, F, P, \sigma)$, where $\Sigma=\{a_1, a_2, \ldots, a_n\}$, $V=\Sigma\cup\{\sigma, S, S_1, S_2, \ldots, S_n\}$, $F=\{g, f_1, f_2, \ldots, f_k\}$, $P=\{\sigma\to Sg, \ S\to S_1 S_2 \ldots S_n\}\cup\{S\to Sf_i \mid i=\overline{1,k}\}$, $f_i=\{S_j\to a_j^{p_{i,j}} S_j \mid j=\overline{1,n}\}$, and $g=\{S_j\to a_j^{c_j} \mid j=\overline{1,n}\}$.

In the grammar G, any derivation from σ is of the form

$$\sigma\Rightarrow Sg \models\!\Rightarrow S\tilde{f}g \Rightarrow S_1\tilde{f}gS_2\tilde{f}g\ldots S_n\tilde{f}g$$

$$\models\!\Rightarrow a_1^{c_1+\sum_{i=1}^{k}\varphi_{f_i}(\tilde{f})\,p_{i,1}} a_2^{c_2+\sum_{i=1}^{k}\varphi_{f_i}(\tilde{f})\,p_{i,2}} \ldots a_n^{c_n+\sum_{i=1}^{k}\varphi_{f_i}(\tilde{f})\,p_{i,n}},$$

where $\tilde{f}\in(F\setminus\{g\})^*$. This implies that $\varphi(L(G))=L(c; p_1, p_2, \ldots, p_k)$. The lemma is proved.

THEOREM 1. If $L\subseteq a_1^* a_2^* \ldots a_n^*$ and the set $\varphi(L)$ is semilinear, then the language L is an index language.

The proof immediately follows from Lemma 1 and the closure of the class of index languages with respect to the union operation.

To study the class of single-letter index languages, we introduce the following concept. A strictly increasing function $h: N_+\to N_+$ is called computable by an index grammar $G=(V, \Sigma, F, P, \sigma)$ if the following conditions are fulfilled: $\Sigma=\{a\}$, $\{a, \sigma, T, X\}\subseteq V$, $F=\{f, g\}$, $P=P'\cup\{\sigma\to Tg, \ T\to Tf, \ T\to X\}$, and the nonterminals σ and T are absent in the set of productions $P'\cup f\cup g$; moreover, we have $L(G)=\{[h(i)] | i\in N_+\}$ (the following agreement is used here: for any natural number m, the designation $[m]$ denotes a word a^m) and, for any of natural number $n\geq 0$, there exists the derivation $\sigma \models\!\Rightarrow Tf^n g \models\!\Rightarrow [h(n+1)]$.

We denote by IGF the class of functions computable by index grammars. A strictly increasing function $h: N_+\to N_+$ is called absolutely computable by the index grammar $G=(V, \Sigma, F, P, \sigma)$ if the following conditions are fulfilled: $\Sigma=\{a\}$, $\{a, \sigma, T, X\}\subseteq V$, $F=\{f, g\}$, $P=\{\sigma\to Tg, \ T\to Tf, \ T\to X\}$, the nonterminals σ and T do not occur in the set of productions $f\cup g$, and each of sets f and g taken separately does not contain different productions with the same left side; moreover, we have $L(G)=\{[h(i)] | i\in N_+\}$, where, for any natural number $n\geq 0$, there exists the derivation $\sigma \models\!\Rightarrow Tf^n g \models\!\Rightarrow [h(n+1)]$.

Denote by IGF_0 the class of functions that are absolutely computable by index grammars. For the introduced classes of functions, the following relation is true: $IGF_0\subseteq IGF$. Moreover, the class IGF_0 is a subclass of functions of a natural argument that are computable by determined transducers with nested pushdown memory.

We will say that a function $f: N_+\to N$ is specified by a system of linear recurrence relations with natural coefficients if f is a solution of a system S_f of the form

$$S_f = \begin{cases} f_i(x+1) = \sum_{j=1}^{n} a_{i,j} f_j(x) + b_i \quad \text{when } x \geq c,\ i = \overline{1, n}, \\ f_i(x) = f_{i,x} \text{ when } x = \overline{1, c},\ i = \overline{1, n}, \end{cases} \tag{1}$$

with respect to f_1, where $a_{i,j}$, b_i, n, c, and $f_{i,x}$ are some fixed natural numbers and $n \geq 1$ and $c \geq 1$.

THEOREM 2. For the function $f : N_+ \to N$ specified by a system of linear recurrence relations with natural coefficients, the language $\{[f(x)] \mid x \in N_+\}$ is an index language.

Proof. Let a function $f : N_+ \to N$ be specified by a system of linear recurrence relations with natural coefficients as a solution of the following system S_f with respect to f_1:

$$S_f = \begin{cases} f_i(x+1) = \sum_{j=1}^{n} a_{i,j} f_j(x) + b_i \text{ when } x \geq c,\ i = \overline{1, n}, \\ f_i(x) = f_{i,x} \text{ when } x = \overline{1, c},\ i = \overline{1, n}, \end{cases}$$

where $a_{i,j}$, b_i, n, c, and $f_{i,x}$ are some fixed natural numbers and $n \geq 1$ and $c \geq 1$.

Let us consider the following index grammar:

$G_{S_f} = (\{a, \sigma, T, S^{(0)}, S^{(1)}, \ldots, S^{(c-1)}, S_1, S_2, \ldots, S_n\},\ \{a\}, \{f, g\},\ \{\sigma \to Tg,\ T \to Tf,\ T \to S^{(0)}\},\ \sigma)$,

where

$f = \{S^{(c-1)} \to S_1^{a_{1,1}} S_2^{a_{1,2}} \ldots S_n^{a_{1,n}} a^{b_1}\} \cup \{S_i \to S_1^{a_{i,1}} S_2^{a_{i,2}} \ldots S_n^{a_{i,n}} a^{b_i} \mid i = \overline{1, n}\} \cup \{S^{(i)} \to S^{(i+1)} \mid i = \overline{0, c-2}\}$ and $g = \{S_i \to a^{f_{i,c}} \mid i = \overline{1, n}\} \cup \{S^{(i)} \to a^{f_{1,i+1}} \mid i = \overline{0, c-1}\}$.

It is easy to see that we have $L(G) = \{[f(x)] \mid x \in N_+\}$. The theorem is proved.

COROLLARY 1. The class IGF_0 contains all the strictly increasing functions from N_+ into N_+ that can be specified by a system of linear recurrence relations with natural coefficients.

THEOREM 3. Any function that is absolutely computable by an index grammar can be specified by a system of linear recurrence relations with natural coefficients.

Proof. Let a function f be absolutely computable by an index grammar $G = (V, \Sigma, \{f, g\}, P, \sigma)$, where $P = \{\sigma \to Tg, T \to Tf,\ T \to X\}$. We assign to each nonterminal $S \in V \setminus \{a, \sigma, T\}$ a functional symbol f_S. Consider the following system S_f of linear recurrence relations with natural coefficients:

$$S_f = \begin{cases} f_S(x+1) = \sum_{R \in V \setminus \{a, \sigma, T\}} \varphi_R(\gamma) f_R(x) + \varphi_a(\gamma) \text{ when } x \geq 1,\ S \in \{X \mid (X \to \gamma) \in f\}, \\ f_S(1) = \varphi_a(\gamma),\ S \in \{X \mid (X \to \gamma) \in g\}. \end{cases}$$

It is easily verified that, for any $x \in N_+$ and any f_S ($S \in V \setminus \{a, \sigma, T\}$), the system S_f contains no more than one determining equation for $f_S(x)$ and can be reduced to the form (1). On the other hand, the solution of the system S_f with respect to f_X is the function f being considered. The theorem is proved.

THEOREM 4. The class IGF_0 coincides with the class of all the strictly increasing functions from N_+ into N_+ that are specified by systems of linear recurrence relations with natural coefficients.

The proof follows from Corollary 1 and Theorem 3.

COROLLARY 2. The class *IGF* contains all the strictly increasing functions from N_+ into N_+ that are specified by systems of linear recurrence relations with natural coefficients.

COROLLARY 3. For any function f from the class IGF_0, a natural number K can be found such that $f(x) < K^x$, $x \in N_+$.

LEMMA 2. Let K be some natural constant, let $f : N_+ \to N_+$, and let $h_1(x) = f(x) + K$, $h_2(x) = f(x + K)$. Then we have

(1) $f \in IGF \Rightarrow h_1 \in IGF$ and $h_2 \in IGF$,

(2) $f \in IGF_0 \Rightarrow h_1 \in IGF_0$ and $h_2 \in IGF_0$.

As a proof, we note that the required index grammars G_{h_1} and G_{h_2} are efficiently constructed from the initial grammar G_f for the function f.

LEMMA 3. If a function $f : N_+ \to N$ can be specified by a system of linear recurrence relations with natural coefficients, then the function $\Sigma f(x) = \sum_{i=1}^{x} f(i)$ can also be specified by a system of linear recurrence relations with natural coefficients.

Proof. Let f be a solution of the system

$$S_f = \begin{cases} f_i(x+1) = \sum_{j=1}^{m} a_{i,j} f_j(x) + b_i \text{ when } x \geq c,\ i = \overline{1, m}, \\ f_i(x) = f_{i,x} \text{ when } x = \overline{1, c},\ i = \overline{1, m} \end{cases}$$

with respect to f_1 ($f = f_1$); then the solution of the system

$$S_{\Sigma f} = \begin{cases} h(x+1) = h(x) + \sum_{j=1}^{m} a_{1,j} f_j(x) + b_1 \text{ when } x \geq c, \\ h(x) = \sum_{j=1}^{x} f_{1,j} \text{ when } x = \overline{1, c}, \\ S_f \end{cases}$$

with respect to h is the function $\Sigma f(x)$.

It is easy to show that the system $S_{\Sigma f}$ can be reduced to the form (1). The lemma is proved.

LEMMA 4. The class of functions specified by systems of linear recurrence relations with natural coefficients is closed with respect to the operations of addition and multiplication.

Proof. Let $f = f_1$ and f_1 satisfy the following system of relations with natural coefficients:

$$S_f = \begin{cases} f_i(x+1) = \sum_{j=1}^{n_f} a_{i,j,f} f_j(x) + b_{i,f} \text{ when } x \geq c_f,\ i = \overline{1, n_f}, \\ f_i(x) = f_{i,x} \text{ when } x = \overline{1, c_f},\ i = \overline{1, n_f}. \end{cases}$$

Let $g = g_1$, and let g_1 satisfies the following system of relations with natural coefficients:

$$S_g = \begin{cases} g_i(x+1) = \sum_{j=1}^{n_g} a_{i,j,g} g_j(x) + b_{i,g} \text{ when } x \geq c_g,\ i = \overline{1, n_g}, \\ g_i(x) = g_{i,x} \text{ when } x = \overline{1, c_g},\ i = \overline{1, n_g}. \end{cases}$$

Without loss of generality, we can consider that $c_f = c_g = c$. The system

$$S_{f+g} = \begin{cases} h(x+1) = \sum_{j=1}^{n_f} a_{i,j,f} f_j(x) + \sum_{j=1}^{n_g} a_{i,j,g} g_j(x) + b_{i,f} + b_{i,g} \text{ when } x \geq c, \\ h(x) = f_{1,x} + g_{1,x} \text{ when } x = \overline{1, c}, \\ S_f, \\ S_g \end{cases}$$

specifies the function $(f+g) = h$ for which, by the construction of the system S_{f+g}, we have $(f+g)(x) \equiv f(x) + g(x)$. The system

$$S_{f\cdot g}=\begin{cases} f_{(i_1,i_2)}(x+1)=\sum\limits_{j_1=1}^{n_f}\sum\limits_{j_2=1}^{n_g}(a_{i_1,j_1,f}\cdot a_{i_2,j_2,g})f_{(j_1,j_2)}(x)+\sum\limits_{j_1=1}^{n_f}(a_{i_1,j_1,f}\cdot b_{i_2,g})f_{j_1}(x) \\ \qquad +\sum\limits_{j_2=1}^{n_g}(b_{i_1,f}\cdot a_{i_2,j_2,g})g_{j_2}(x)+(b_{i_1,f}\cdot b_{i_2,g}) \text{ when } x\ge c, \\ f_{(i_1,i_2)}(x)=f_{i_1,x}\cdot g_{i_2,x} \text{ when } x=\overline{1,c}, \\ S_f, \\ S_g \end{cases}$$

specifies the function $(f\cdot g)=f_{(1,1)}$ for which, by the construction of the system $S_{f\cdot g}$, we have $(f\cdot g)(x)\equiv f(x)\cdot g(x)$.

It is easy to make sure that the systems S_{f+g} and $S_{f\cdot g}$ have natural coefficients and can be easily transformed into the form (1). The lemma is proved.

Based on Lemmas 2–4 and Theorem 4, we obtain the theorem given below.

THEOREM 5. The class of functions IGF_0 is closed with respect to the operations of addition, multiplication, summation, addition of a natural constant to the result, and addition of a natural constant to the argument.

LEMMA 5. Functions x and k^x, $k\in N$, $k\ge 2$, belong to the class IGF_0.

To prove the lemma, we consider the following systems of linear recurrence relations with natural coefficients for the functions x and k^x, $k\in N$, $k\ge 2$:

$$S_x=\begin{cases} f_1(x+1)=f_1(x)+1 \text{ when } x\ge 1, \\ f_1(1)=1, \end{cases}$$

$$S_{k^x}=\begin{cases} f_1(x+1)=k\cdot f_1(x) \text{ when } x\ge 1, \\ f_1(1)=k. \end{cases}$$

COROLLARY 4. All the functions that are finite sums of functions of the form x^k, a^x, and $x^k a^x$, where $k, a\in N_+$ and $a\ge 2$, are contained in the class IGF_0.

LEMMA 6. Let a function of a real argument $\tilde{f}_n(x)=\sum\limits_{i=0}^{n}a_i x^i$, $n\in N_+$, $a_i\in Z$, $i=\overline{0,n}$, $a_n>0$, and all its derivatives up to the nth order inclusively assume positive values on the set $[1,+\infty)$. Then the function $f_n(x)=\tilde{f}_n(x)$, $x\in N_+$, can be specified by a system of linear recurrence relations with natural coefficients.

We prove the lemma by induction in n.

The basis of induction is $n=1$. In this case, we have $f_1(x)=ax+b$, where $a+b$, $a\in N_+$, and f_1 is a solution of the system

$$S_{f_1}=\begin{cases} f_1(x+1)=f_1(x)+a \text{ when } x\ge 1, \\ f_1(1)=a+b \end{cases}$$

with the parameter $c=1$.

The step of induction is as follows: let the statement of the lemma be true for all $k<n$, $k\in N_+$, and let the functions can be specified by systems with the parameter $c=1$.

Let us consider the function $\tilde{\varphi}(x)=\tilde{f}_n(x+1)-\tilde{f}_n(x)=\sum\limits_{i=0}^{n-1}\left(\sum\limits_{j=i+1}^{n}a_j C_j^i\right)x^i=\sum\limits_{i=0}^{n-1}b_i x^i$, where $b_i=\sum\limits_{j=i+1}^{n}a_j C_j^i\in Z$, $i=\overline{0,n-1}$. Since $\tilde{f}_n$ assume positive values on the set $[1,+\infty)$ together with their derivatives up to the nth order inclusively, $\tilde{\varphi}$ assume positive values on the set $[1,+\infty)$ together with their derivatives up to the $(n-1)$th order inclusively and

$b_{n-1} = a_n C_n^{n-1} > 0$. Then the function $\varphi(x)$ is specified on N_+ as a solution (with respect to φ) of some system S_φ of linear recurrence relations with natural coefficients with the parameter $c = 1$. In this case, the corresponding system for f_n is of the form

$$S_{f_n} = \begin{cases} f_n(x+1) = f(x) + \varphi(x) \text{ when } x \geq 1, \\ f_n(1) = \sum_{i=0}^{n} a_i, \\ S_\varphi, \end{cases}$$

and the parameter c for the system is also equal to unity. According to the principle of mathematical induction, the lemma is proved.

We assume that $f_n(x) = \sum_{i=0}^{n} a_i x^i$, $n \in N_+$, $a_i \in Z$, $i = \overline{0, n}$, and $a_n > 0$. Then there exists a natural number K such that the function $f_n(x + K)$ satisfies the conditions of Lemma 6. This implies the validity of the theorem presented below.

THEOREM 6. The class IGF_0 contains all the strictly increasing polynomials (with integer coefficients)on N_+ that assume positive values on N_+.

Corollary 3 implies that the functions 2^{2^n} and $2^{\frac{n(n+1)}{2}}$ do not belong to the class of functions IGF_0. At the same time, as is shown in [6], the condition $\overline{\lim_{n \to \infty}} (f(n))^{1/n} = \infty$ for an increasing function $f : N_+ \to N$ is sufficient in order that the language $L_f = \{[f(n)] | n \in N_+\}$ be not an index language, but then the functions 2^{2^n} and $2^{\frac{n(n+1)}{2}}$ also do not belong to the class of functions IGF. The theorem given below is true.

THEOREM 7. In the general case, the operations of superposition and multiplication Π ($\Pi f(x) = \prod_{i=1}^{x} f(i)$) applied to a function from the class of functions IGF or IGF_0 do not retain the properties of the function to belong to the class of functions IGF.

Note that the class IGF_0 contains not only functions that are of the form of finite sums of functions of the form x^k, a^x, and $x^k a^x$, where $k, a \in N_+$ and $a \geq 2$. For example, the Fibonacci sequence with the initial members 1 and 2 is the domain of values of an IGF_0-function.

In conclusion, it should be added that Theorem 5 does not give the answer to the following question: whether it is correct that $L_+ = \{[h(n)] | n \in N_+\}$ and $L_* = \{[\tilde{h}(n)] | n \in N_0\}$ are index languages in the general case, where $h_i : N \to N$, $i = 1, 2$, are any strictly increasing functions such that $L_i = \{[h_i(n)] | n \in N_+\}$, $i = 1, 2$, are index languages, $h(x) = h_1(x) + h_2(x)$, and $\tilde{h}(x) = h_1(x) \cdot h_2(x)$.

REFERENCES

1. A. V. Aho, "Indexed grammars — an extension of context-free grammars," J. ACM, **15**, No. 4, 647–671 (1968).
2. A. V. Aho, "Nested stack automata," J. ACM, **16**, No. 3, 383–406 (1969).
3. L. P. Lisovik, "Recognition of perfect labeled trees with regularity condition," Kibern. Sist. Anal., No. 2, 176–179 (1996).
4. L. P. Lisovik, "Deterministic acceptors for index languages," Kibern. Sist. Anal., No. 4, 100–114 (1997).
5. S. Ginsburg, The Mathematical Theory of Context-Free Languages [Russian translation], Mir, Moscow (1970).
6. T. Hayashi, "On derivation trees of indexed grammars: An extension of the *uvwxy*-theorem," Publ. RIMS Kyoto Univ., **9**, No. 1, 61–92 (1973).