УДК 681.3.06

ПІДКЛАСИ ПАРАМЕТРИЧНИХ ГРАМАТИК КВАЗІРЕКУРСИВНОГО ТИПУ

В. М. Євладенко

Виділяються окремі підкласи параметричних граматик квазірекурсивного типу, орієнтовані на ефективний синтаксичний контроль.

Separate subclasses of parametric grammars of quazirecursive types oriented to the effective syntactic control are singled out in this paperю

Побудова ефективних систем програмування, орієнтованих на клас вхідних мов, тісно пов’язана з розробкою формалізмів, які є досить зручними для практичного задання реальних мов програмування. До таких формалізмів практика висуває перш за все такі дві вимоги:

1) формалізм повинен бути орієнтованим на побудову ефективного синтаксичного аналізатора для відповідної мови програмування;

2) формалізм повинен мати досить простий апарат, який дає можливість ураховувати погодженість окремих конструкцій і як завгодно далеких фрагментів тексту, що є характерним для мов програмування.

Вказаним двом вимогам не задовольняють ні контекстно-вільні граматики Хомського, які широко застосовуються для опису синтаксису мов програмування, ні їх модифікація – бекусівські нормальні форми (БНФ).

В статті показується, що ці вимоги можуть бути виконані у випадку використання окремих підкласів параметричних граматик квазірекурсивного типу (ПГКвРТ) [1].

Наведемо основні означення. Нехай Я, Я_i, $\overline{\text{Я}}_i$ ($i \in I$, де I – деяка множина індексів) – довільні мови в алфавіті X, а Ω_i ($i \in I$) – довільні бінарні відношення у вільній півгрупі $F = F(X)$, тобто $\Omega_i \subseteq F \times F\ (i \in I)$ – деяка підмножина універсального бінарного відношення.

Під операцією ***породження*** будемо розуміти таке відображення, яке мовам Я, Я_i, $\overline{\text{Я}}_i$ ($i \in I$) і бінарним відношенням Ω_i ($i \in I$) ставить у відповідність мову $R\ (\text{Я};\ \text{Я}_i\ (\Omega_i)\, \overline{\text{Я}}_i,\ i \in I)$, яка є множиною тих і тільки тих слів, які одержуються із слів $\alpha_1^i\ \alpha^i\ \alpha_2^i \subset \text{Я}$, де $\alpha^i \subset \text{Я}_i$, а $(\alpha_1^i, \alpha_2^i) \subset \Omega$, шляхом заміни α^i деякими словами із $\overline{\text{Я}}_i$. ***Породженням мови*** Я за допомогою мов Я_i, $\overline{\text{Я}}_i$ ($i \in I$) при бінарних відношеннях Ω_i ($i \in I$) називається мова $R\ (\text{Я};\ \text{Я}_i\ (\Omega_i)\, \overline{\text{Я}}_i,\ i \in I)$.

Нехай X і $У$ – довільні термінальний і нетермінальний алфавіти відповідно, що не перетинаються, а $\text{Я};\ \text{Я}_i$, $\overline{\text{Я}}_i$, ($i \in I$) –мови в алфавіті $X \cup У$, Ω_i ($i \in I$) – бінарні відношення (можливо і пусті) у вільній півгрупі $F\ (X \cup У)$.

Під ***операцією рекурсивного породження*** будемо розуміти таке відображення, яке мовам $\text{Я};\ \text{Я}_i$, $\overline{\text{Я}}_i$, ($i \in I$) і бінарним відношенням Ω_i ($i \in I$)

ставить у відповідність мову R^* $(Я;\ Я_i\ (\Omega_i)\ \overline{Я}_i,\ i \in I)$, яка є максимальною підмовою (підмножиною слів) в алфавіті X мови $Я^1 \cup Я^2 \cup \ldots$, де $Я^1 = Я$, а $Я^{n+1} = R\ (Я^n\ ,\ Я_i\ (\Omega_i)\ \overline{Я}_i,\ i \in I)$. Мова R^* $(Я;\ Я_i\ (\Omega_i)\ \overline{Я}_i,\ i \in I)$, називається ***рекурсивним породженням мови*** $Я$ за допомогою мов $Я_i$, $\overline{Я}_i$ $(i \in I)$ при бінарних відношеннях Ω_i . Операція рекурсивного породження має ряд цікавих властивостей. Перш за все вона дає можливість в загальному випадку із скінченних мов будувати нескінченні мови. Крім того, ця операція абстрактним способом задає ті композиції, які найбільш часто застосовуються на практиці при побудові нових мов із заданих.

Під ***граматикою квазірекурсивного типу*** відносно мов $Я;\ Я_i$, $\overline{Я}_i$ $(i \in I)$, і бінарних відношень Ω_i будемо розуміти об'єкт R^* $(Я;\ Я_i\ (\Omega_i)\ \overline{Я}_i,\ i \in I)$, що задає рекурсивне породження мови $Я$ за допомогою мов $Я_i$, $\overline{Я}_i$ при бінарних відношеннях Ω_i.

Виявляється, що клас граматик рекурсивного типу [2] є підкласом граматик квазірекурсивного типу. Дійсно, нехай $Я_i$ $(i = 1,\ldots,n)$ – довільні мови в алфавіті $X \cup У$ (X - термінальний, $У = \{y_1, \ldots, y_n\}$ – нетермінальний алфавіти), а Ω - універсальні бінарні відношення в тому ж алфавіті, тоді $y_i ::= Я_i\ (X \cup У)\ (i = 1,\ldots, n) = R^*(y_1;\ y_1(\Omega)\ Я_1,\ldots,y_n(\Omega)Я_n)$. Отже, і бекусівські нормальні форми є підкласом граматик квазірекурсивного типу.

Параметричною граматикою квазірекурсивного типу (ПГКвРТ) називається об'єкт $Г^* = Г^*\ (R^*,\ \Omega^*\ ,\sigma^*)$, де R^* – граматика квазірекурсивного типу, а зміст параметрів Ω^* і σ^* аналогічний змісту відповідних параметрів Ω і σ в параметричних граматиках рекурсивного типу [2].

Поняття ПГКвРТ є надзвичайно загальним. Природно, що з практичної точки зору важливими є окремі їх підкласи. Такі підкласи виділяються шляхом накладання на параметри $R^*,\ \Omega^*\ ,\sigma^*$ додаткових обмежень, що ми далі і зробимо.

Нехай R^* – граматика квазірекурсивного типу, яка задається наступною операцією рекурсивного породження:

$$R^*(y_1;\ y_1(\Omega)\ Я_1,\ldots,y_r(\Omega)Я_r,\ y^i_{r+1}(\Omega)\ Я^i_{r+1},\ldots,y^i_n(\Omega)Я^i_n),\ (i = 1, 2,\ldots),$$

де Ω – універсальне бінарне відношення; X – скінченний термінальний алфавіт; $У = У_1 \cup У_2$ – нетермінальний алфавіт, де $У_1 = \{y_1,\ldots,y_r\}$ – скінченна множина нетермінальних символів (метасимволів), $У_2 = \{y^i_m\}$ $(m = r+1,\ldots, n;\ i = 1,2,\ldots)$ – нескінченна множина метасимволів за рахунок того, що індекс i пробігає всю множину натуральних чисел, тобто $У_2 = \{y^1_{r+1},\ y^2_{r+1},\ldots,\ y^1_n,\ y^2_n,\ldots\}$; $Я_j$ $(j = 1,2,\ldots, r)$ – скінченні мови в алфавіті $X \cup У$; $Я^i_p$ $(p = r+1,\ldots, n)$ скінченні мови в алфавіті $X \cup У$, кожна із яких містить принаймні одне із слів, що має структуру першого виду:

1) $\upsilon_1\ \varphi^i_1\ \upsilon_2\ \varphi^i_2 \ldots \upsilon_r\ \varphi^i_r\ \upsilon_{r+1}$, де $\upsilon_j \in F\ (X \cup У)\ (j = 1,\ldots, r+1)$, $\varphi_j \in F\ (X \cup У) \setminus \varepsilon$ $(j = 1,\ldots, r)$, а $\varphi^i_j = \varphi_j \ldots \varphi_j$ (φ_j береться i раз, ε – пусте слово), або ж структуру другого виду:

2) $\upsilon_1 y_m^{f(i)} \upsilon_2$, де $y_m^{f(i)} \in У_2$ (m=r + 1,…, n); $f(i)$- функція натурального аргументу i, що приймає значення із множини {1,2,… }.

Виділений клас P граматик квазірекурсивного типу будемо називати ***одноіндексними граматиками.***

Одноіндексні граматики зручно задавати тією ж схемою, що і БНФ. Дійсно, нехай R^* (y_1; y_1 (Ω) $Я_1$, $y^i_2(\Omega)$ $Я^i_2$) - граматика із класу P, де термінальний алфавіт X = {a, $в$, (,)}, нетермінальний алфавіт $У = У_1 \cup У_2$, $У_1 = \{y_1\}$, $У_2 = \{y^i_2\}$, а $Я_1 = \{\varepsilon, (ay_1y^1_2\}$, $Я^i_2 = \{в)^i\}$. Тоді, як легко переконатися, ця граматика задається такою схемою БНФ (y_1 - аксіома) :

$$y_1 ::= \varepsilon \mid (a\, y_1 y^1_2$$
$$y^i_2 ::= в)^{\,i}$$

При цьому, наприклад, слово ($a(aв)в$) як у граматиці R^* , так і в наведеній БНФ одержується наступним породженням (виводом):

$$y_1 \Rightarrow (ay_1y^1_2 \Rightarrow (a\ (ay_1y^1_2\, y^1_2 \Rightarrow (a(ay^1_2\, y^1_2 \Rightarrow (a(aв)y^1_2 \Rightarrow (a(aв)в)$$

Очевидно, що клас одноіндексних граматик містить в собі в явному вигляді засоби побудови погоджених конструкцій в мовах, які вони задають. Так, в наведеному вище прикладі $\underset{1}{(a}\,\underset{2}{(a}\,\underset{2}{в)}\,\underset{1}{в)}$ погодженими є фрагменти, що позначені однаковими цифрами.

До того ж цей клас граматик суттєво ширший класу контекстно-вільних граматик. Так, наприклад, мова $Я = \{a^i\ в^i\ c^i\ (\ i = 1,2,\ldots\ )\}$, яку, як відомо [3], не можна задати ніякою контекстно-вільною граматикою, породжується такою одноіндексною граматикою:

$$y_1 ::= y^1_2$$
$$y^i_2 ::= a^i\ в^i\ c^i \mid y_2^{i+1}$$

Із наведеного зв’язку між параметрами R^*, Ω^* і σ^* слідує, що прості одноіндексні граматики мають чітку орієнтацію на ефективний синтаксичний контроль породжених ними мов. Це пов’язано з тим, що в цьому випадку керування синтаксичним аналізом здійснюється за допомогою простих бінарних відношень.

ПГКвРТ будемо називати – ***простою одноіндексною граматикою***, якщо вона має вид: $Г^* = Г^*\ (R^*, \Omega^*, \sigma^*)$, де R^* – одноіндексна граматика, яка задається такою схемою БНФ:

$$y_j ::= Я_j(У \cup X) \quad (j = 1, \ldots, r)$$
$$y^i_m ::= Я^i_m\ (У \cup X) \quad (m = r + 1, \ldots, n;\ i = 1,2, \ldots),$$

де X – термінальний алфавіт, $У = У_1 \cup У_2$ – нетермінальний алфавіт, $У_1 = \{y_1, \ldots, y_r\}$, $У_2 = \{y^i_m\}$, а зв’зок між параметрами R^*, Ω^* і σ^* виражається співвідношеннями :

$$\sigma^*\ (j, \gamma) = \{(\ \upsilon_1 y_j\ \upsilon_2,\ \upsilon_1 x_{k\,j}\ F(X)\},$$

де $\upsilon_1 \in F(X)$, $x_{k\,j} \in X$, $\upsilon_2 \in F\ (X \cup У)$ для всіх $\gamma \in Я_j\ (У \cup X)$ (j =1,…, r);
$\sigma^*\ (i_s, \gamma) = \{(\ \upsilon_1 y^i_s\ \upsilon_2,\ \upsilon_1 x_{k\,j}\ F(X)\},$

де $\upsilon_1 \in F(X)$, $x_{kj} \in X$, $\upsilon_2 \in F(X \cup У)$ для всіх $\gamma \in Я^i_s(У \cup X)$ ($s=r+1,\ldots,n$; i=1, 2,…) .

ПГКвРТ будем називати *k-**простою одноіндексною граматикою***, якщо вона має вид:

$Г^* = Г^*(R^*, \Omega^*, \sigma^*)$, де R^* - одноіндексна граматика, а зв'язок між параметрами R^*, Ω^* і σ^* виражається співвідношеннями:

$$\sigma^*(j, \gamma) = \{(\upsilon_1 y_j \upsilon_2, \upsilon_1(\cup zF(X))\}, \upsilon_1 \in F(X), \upsilon_2 \in F(X \cup Y)\}$$

для всіх $\gamma \in Я_j(У)$ $(j = 1, 2,\ldots, r);$

$$\sigma^*(i_s, \gamma) = \{(\upsilon_1 y_j \upsilon_2, \upsilon_1(\cup zF(X))\}, \upsilon_1 \in F(X), \upsilon_2 \in F(X \cup Y)\}$$

для всіх $\gamma \in Я^i_s(У)$ $(s = r+1, \ldots, n;\ i=1, 2, \ldots);$ z – деяка множина термінальних слів довжиною k, яка ставиться у відповідність слову γ.

Мова, що задається k-простою одноіндексною граматикою, називається ***k-простою одноіндексною граматикою***. У випадку, коли k = 1, одержимо 1-просту одноіндексну граматику (мову).

Клас k – простих одноіндексних граматик має безпосередній зв'зок з *LL(k)* – грататиками і орієнтований на безперебірний синтаксичний контроль методом "розгортки" [4].

БІБЛІОГРАФІЯ

1. Редько В.Н. Параметрические грамматики и проблема синтаксического анализа языков. //Труды II Всесоюзной конференции по прораммированию. – Новосибирск, 1970.
2. Редько В.Н. Параметрические грамматики и проблема параметризации. //Автоматизация программирования. Сборник научных трудов. Вып. I. – К.: ИК АН УССР, 1969.
3. Евладенко В.Н. Алгоритм распознавания для одного класса параметрических грамматик квазирекурсивного типа. //Программирование. Вып 5. – М.: Наука, 1980.
4. Братчиков Н.Л. Синтаксис языков программирования. – Новосибирск: Наука, 1975.