Практическое занятие №3. Преобразования грамматик в нормальную форму Хомского. Устранение левой рекурсии и левая факторизация..

Д.Ю. Чалый

24 декабря 2010 г.

Нормальная форма Хомского используется в алгоритме Кока-Янгера-Касами.

1 Пример 1.

Рассмотрим алгоритм преобразования грамматик в нормальную форму Хомского (НФХ). Грамматики в НФХ имеют продукции, построенные по одному из следующих образцов: $A \to BC$ и $A \to a$, где $A, B, C \in N$, $a \in T$. Преобразуем следующую грамматику в нормальную форму Хомского:

$$\begin{array}{rcl} S & \to & LaM \\ L & \to & LM \mid \varepsilon \\ M & \to & MM \mid ab \mid \varepsilon \end{array}$$

Для этого последовательно проводим следующие преобразования:

1. Удаляем ε-продукции.
2. Удаляем цепные продукции.
3. Удаляем непорождающие символы.
4. Удаляем недостижимые символы.
5. Приводим к нормальной форме Хомского.

После первых четырех преобразований (проделайте их самостоятельно), грамматика примет следующий вид:

$$\begin{array}{rcl} S & \to & LaM \mid La \mid aM \mid a \\ L & \to & LM \mid ab \mid MM \\ M & \to & ab \mid MM \end{array}$$

Преобразуем теперь грамматику в НФХ. Сначала избавимся от терминальных символов в правых частях продукций:

$$\begin{array}{rcl} S & \to & LAM \mid LA \mid AM \mid a \\ L & \to & LM \mid AB \mid MM \\ M & \to & AB \mid MM \\ A & \to & a \\ B & \to & b \end{array}$$

Единственная продукция, которая не находится в НФХ — $S \to LAM$. К НФХ она приводится так:

$$\begin{array}{rcl} S & \to & LS' \\ S' & \to & AM \end{array}$$

В итоге получится следующая грамматика в НФХ:

$$\begin{array}{rcl} S & \to & LS' \mid LA \mid AM \mid a \\ S' & \to & AM \\ L & \to & LM \mid AB \mid MM \\ M & \to & AB \mid MM \\ A & \to & a \\ B & \to & b \end{array}$$

2 Задание 1.

Преобразовать в НФХ следующую грамматику:

$$\begin{array}{rcl} Number & \to & Integer \mid Real \\ Integer & \to & Digit \mid IntegerDigit \\ Real & \to & IntegerFractionScale \\ Fraction & \to & .Integer \\ Scale & \to & eSignInteger \mid Empty \\ Digit & \to & 0 \mid \ldots \mid 9 \\ Sign & \to & + \mid - \\ Empty & \to & \varepsilon \end{array}$$

3 Пример 2.

Удаление левой рекурсии. Простой пример:

$$A \quad \to \quad A\alpha_1 \mid \ldots \mid A\alpha_n \mid \beta_1 \ldots \beta_m$$

Заменяем это на правила:

$$\begin{array}{rcl} A_{head} & \to & \beta_1 \ldots \beta_m \\ A_{tail} & \to & \alpha_1 \ldots \alpha_n \\ A_{tails} & \to & A_{tail} A_{tails} \mid \varepsilon \\ A & \to & A_{head} A_{tails} \mid A_{head} \end{array}$$

Сложный случай. Рассмотрим его на примере грамматики:

$$\begin{array}{lcl} S & \to & ABc \\ A & \to & \varepsilon \\ B & \to & Cd \mid AB\int \\ C & \to & Se \mid e \end{array}$$

Эта грамматика является леворекурсивной несмотря на отсутствие явной левой рекурсии. В этом случае, как зачастую и происходит, левую рекурсию «скрывают» ε-продукции. Например, благодаря им возможен такой вывод: $S \to ABc \to Bc \to Cdc \to Sedc$. В итоге мы получили сентенциальную форму, которая начинается с того же нетерминала, что и начало вывода. Такие грамматики избавляются от левой рекурсии при помощи следующего алгоритма.

Сначала мы удалим ε-продукции и цепные продукции. Цепных продукций не будет и получится следующая грамматика:

$$\begin{array}{lcl} S & \to & Bc \\ B & \to & Cd \mid B\int \\ C & \to & Se \mid e \end{array}$$

Теперь упорядочим нетерминалы грамматики. Например так: S, B, C. Наша цель состоит в том, чтобы в новой грамматике продукции начинались либо с терминального символа, либо с более позднего нетерминала. Рассмотрим нетерминал S. Все нормально, единственная продукция для него начинается с более позднего нетерминала. Ее можно поместить в итоговую грамматику. Рассмотрим следующий нетерминал, B. Для него есть продукция $B \to B\int$, которая является леворекурсивной. Преобразуем ее так:

$$\begin{array}{lcl} B & \to & Cd \mid CdB' \\ B' & \to & \int \mid \int B \end{array}$$

Поместим получившиеся продукции в итоговую грамматику. Рассмотрим теперь нетерминал C. Для него есть продукция $C \to Se$, которая начинает с более «раннего» нетерминала. Поэтому вместо нее кандидатами на помещение в итоговую грамматику будет множество продукций $C \to \beta e$, где $S \to \beta$ — продукции для S, что дает нам кандидата $C \to Bce$. Данная продукция тоже начинается с более «раннего» нетерминала. Повторим предложенное преобразование и получим продукции $C \to Cdce \mid CdB'ce$. Таким образом для C мы получили три продукции:

$$C \to Cdce \mid CdB'ce \mid e$$

Преобразуем их как в простом случае и получим:

$$\begin{array}{lcl} C & \to & e \mid eC' \\ C' & \to & dce \mid dB'ce \mid dceC' \mid dB'ceC' \end{array}$$

В итоге получится следующая грамматика:

$$
\begin{array}{rcl}
S & \to & Bc \\
B & \to & Cd \mid CdB' \\
B' & \to & \int \mid \int B \\
C & \to & e \mid eC' \\
C' & \to & dce \mid dB'ce \mid dceC' \mid dB'ceC'
\end{array}
$$

4 Задача 2.

Убрать левую рекурсию из следующей грамматики:

$$
\begin{array}{rcl}
S & \to & c \mid CAB \\
A & \to & Bc \mid Cc \\
B & \to & Ba \mid CA \\
C & \to & CSe \mid \varepsilon
\end{array}
$$

5 Задача 3.

Пусть дана следующая грамматика:

$$A \to \alpha\beta_1 \mid \ldots \mid \alpha\beta_n$$

Преобразовать ее к виду, где не будет первых символов α.