Лекция №3

Формальные грамматики

Две грамматики G[S], G[Z] называются эквивалентными если совпадают словари терминальных символов V_T и если языки, порождаемые грамматиками L{G[S]} = L{G[Z]} равны.

Перестроение грамматик нужно если:

1. Необходимо придание продукциям грамматики свойств, обеспечивающих корректную работу того или иного алгоритма анализа
2. Целью является исследование (см. далее)

Построение ε–свободной грамматики (e-free)

$L\{G[Z]\} \setminus \{\varepsilon\} = L\{G`[Z]\}$;

$Z \Rightarrow^*_{G[Z]} X$;

$Z \Rightarrow^*_{G`[Z]} X$;

Для любого контекстно–свободного языка L существует контестно-свободая, ε–свободная грамматика G[Z], такая, что $L \setminus \varepsilon = L\{G[Z]\}$

Продукцию вида $U \rightarrow \varepsilon$ назовём ε–продукцией, а не терминальные символы, для которых имеет место быть вывод $A \Rightarrow^+ \varepsilon$ или $Z \Rightarrow^* \varepsilon$.

Алгоритм:

1. В множестве продукций формальной грамматики отыскиваются ε–продукции из которых формируется множество $R\varepsilon$;
2. Среди словаря не терминалов отыскиваются ε–порождающие символы;
3. Каждой продукции формальной грамматики в правой части которой присутствуют один или несколько ε–порождающих не терминальных символов ставится в соответствие продукция у которой, по сравнению с исходной, опущено один или несколько ε–порождающих символов. В результате формируется множество правил R_0;
4. Из множества исходных правил грамматики удаляются ε–продукции и объединяются с множеством с предыдущего шага: $\{R \setminus R\varepsilon\} \cup R_0$.

Пример

S → (E) I E

E → T I T+E I T-E

T → F I T∗F I T/F

F → <iden> I <data> I ε

V_N = {S, E, T, F, <<u>iden</u>>, <<u>data</u>>}

V_T = {(,), +, -, ∗, /, ε}

1. $R\varepsilon = \{F \rightarrow \varepsilon\}$
2. $S \Rightarrow E \Rightarrow T \Rightarrow F \Rightarrow \varepsilon$

   {

   $S \Rightarrow^* S$

   $E \Rightarrow^+ \varepsilon$

   $T \Rightarrow^+ \varepsilon$

   $F \rightarrow \varepsilon$

   }

3. $S \to ()$
   $E \to T+ \mid +E \mid + \mid T- \mid -E \mid -$
   $T \to T* \mid *F \mid * \mid T/ \mid /F \mid /$
   $R_0 = \{S, E, T\}$

   $S \to (E) \mid E \mid ()$
   $E \to T \mid T+E \mid T-E \mid T+ \mid +E \mid + \mid T- \mid -E \mid -$
   $T \to F \mid T*F \mid T/F \mid T* \mid *F \mid * \mid T/ \mid /F \mid /$
   F → <iden> | <data>

Проблема не пустоты языка заключается в том, что бы дать ответ на вопрос «Всегда ли данная грамматика порождает терминальные цепочки или же смесь из терминалов и не терминалов?» и «Все ли правила грамматики используются при получении предложений языка?».

Релевантность и иррелевантность

Продукция будет являться релевантной, если существует хотя бы одна цепочка, выводимая из аксиомы и и принадлежащая языку, в котором эта продукция используется.

$x \in L\{G[Z]\}$

$Z \Rightarrow^* y\,U\,v \Rightarrow yuv \Rightarrow^+ x;\ U \to u$

Нормальная форма Хомского – en. Chomsky normal form (в целях исследования)

Существует конструктивная теорема, которая утверждает, что любой ε–сводобный контекстно свободный язык L может быть порождён формальной грамматикой, представленной в нормальной форме Хомского, продукции которой имеют вид:

$A, B, C \in V_N$

$\alpha \in V_T$

Назовём первичными продукциями правила формальной грамматики вида:

1. Поиск первичных продукций
   $A_i \to B_j$
   $\{A_i, B_j\} \in V_N$
   $U\{A_i\}, N\{A_i\}$

   <ЦБЧ> → <Ц> | <ЦБЧ><Ц>
   <Ц> → 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9
   <ДЧ> →<ЦБЧ>, <ЦБЧ>
   Необходимо произвести замену первичных продукций с не терминалом A_i в левой части на не первичные продукции не терминалов B_j. Таким образом получим множество, где отсутствуют первичные продукции
2. Поиск вторичных продукций
   Правая часть вторичной продукции представляет собой цепочку над словарями терминалов и не терминалов $A_i \to v;\ v \in (V_T \cup V_N)$
   $x_1, x_2, x_3, \ldots, x_n$
   $y_1, y_2, y_3, \ldots, y_n$
   $y_i = \{$
      $x_i,\ x_i \in V_N$
      $Nx_i,\ x_i \in V_T$
      $Nx_i \to x_i \in R$
   $\}$
3. Поиск троичных продукций, в правой части которых помещаются более двух символов
   $A \to B_1\,B_2 \ldots B_m$
   $A, B_i \in V_N$
   $A \to B_1 D_1$

$D_1 \to B_2D_2$

…

$D_{m-2} \to B_{m-1}B_m$

$S \to A \mid ABA$
$A \to aA \mid a \mid B$
$B \to bB \mid b$

1. Поиск первичных продукций

$U(S) = \{S \to A\}$, $N(S) = \{S \to ABA\}$
$U(A) = \{A \to B\}$, $N(A) = \{A \to aA \mid a\}$
$U(B) = \{\}$, $N(B) = \{B \to bB|b\}$
$S \to aA \mid a \mid bB \mid b \mid ABA$
$A \to aA \mid a \mid bB \mid b$
$B \to bB \mid b$
aA $Pa \to a$
bB $Pb \to b$

$S \to PaA \mid a \mid PbB \mid b \mid ABA$
$A \to PaA \mid a \mid PbB \mid b$
$B \to PbB \mid b$
$Pa \to a$
$Pb \to b$
$D \to BA$
$S \to ABA$
{
$S \to AD$
$D \to BA$
}