Глава 1. Элементы теории формальных языков и грамматик

1.8. Свойство самовложения КС-грамматик

Если в КС-грамматике существует нетерминал A, для которого $A \overset{+}{\Rightarrow} \alpha A \beta$, $\alpha, \beta \in (V_T \cup V_N)^+$, то о такой грамматике говорят, что она содержит *самовложение*. Заметим, что строки α и β являются непустыми строками.

Свойство самовложения позволяет эффективно различать контекстно-свободные (нерегулярные) и регулярные языки. Теоретически любая КС-грамматика, не содержащая самовложения, эквивалентна регулярной грамматике и генерирует регулярный язык. Регулярная грамматика не может содержать самовложения.

Пример грамматики, содержащей самовложение

$S \to aSb$, $S \to \varepsilon$

Грамматика арифметических выражений

$E \to E + T \mid T$

$T \to T \times F \mid F$

$F \to (E) \mid i$

также содержит самовложение, поскольку, например, существует вывод $E \Rightarrow T \Rightarrow F \Rightarrow (E)$. В таких случаях говорят, что нетерминал E проявляет свойство самовложения. Нетерминалы T и F также проявляют свойство самовложения:

$T \Rightarrow F \Rightarrow (E) \Rightarrow (T)$

$F \Rightarrow (E) \Rightarrow (T) \Rightarrow (F)$

1.9. Лемма Огдена и лемма о разрастании для КС-языков

Лемма Огдена. Для любой КС-грамматики $G = (V_T, V_N, P, S)$ существует такая константа $k \geq 1$, что если $z \in L(G)$, $|z| \geq k$ и для любых выделенных в z не менее k позиций, строку z можно записать в виде $z = uvwxy$, причем:

1) w содержит хотя бы одну выделенную позицию;

2) либо и u и v содержат выделенные позиции, либо её содержат и x и y;

3) vwx содержит не более k выделенных позиций;

4) существует такой нетерминал $A \in V_N$, что

$S \overset{+}{\Rightarrow} uAy \overset{+}{\Rightarrow} uvAxy \overset{+}{\Rightarrow} \cdots \overset{+}{\Rightarrow} uv^i Ax^i y \overset{+}{\Rightarrow} uv^i wx^i y$ для всех $i \geq 0$.

Следствием леммы Огдена является *лемма о разрастании* (*лемма о накачке*, *лемма о подкачке*; англ. *pumping lemma*) для КС-языков.

Для любого КС-языка L существует такая константа $k \geq 1$, что любую строку $z \in L$, $|z| \geq k$, можно записать в виде $z = uvwxy$, где $|vwx| \leq k$, $vx \neq \varepsilon$, и для любого $i \geq 0$ справедливо $uv^i wx^i y \in L$.

Лемма Огдена и лемма о разрастании для КС-языков полезны для доказательства утверждений о том, что некоторые языки не являются контекстно-свободными.

Для примера рассмотрим язык $L = \{\alpha\alpha \mid \alpha \in \{a, b\}^*\}$. Доказательство проведем методом от противного, используя лемму о разрастании. Предположим, что L – КС-язык. Рассмотрим строку $z = a^k b^k a^k b^k$, где k – константа из леммы о разрастании. Поскольку $z \in L$, ее можно представить в виде $z = uvwxy$, где $|vwx| \leq k$, тогда строка $uv^i wx^i y$ $(i \geq 0)$ также должна принадлежать языку. Проверим для $i = 0$, т. е. будет ли uwy принадлежать языку. Возможны следующие варианты:

1. v и x находятся либо в первой половине z, либо во второй. В этом случае строка uwy образуется путем исключения символов a и b из первой или второй половины z, но не из обеих. Следовательно, $uwy \notin L$.

2. v содержит символы b из первой половины z, x содержит символы a из второй половины z. В этом случае uwy образуется путем исключения символов b из первой половины и символов a из второй половины, т. е. полученная строка $uwy \notin L$.

Таким образом, данная строка не может быть представлена в виде $uvwxy$, чтобы uwy тоже принадлежала языку. Поэтому язык $L = \{\alpha\alpha \mid \alpha \in \{a, b\}^*\}$ не является КС-языком.