Семинирское занятие 3. Формальные грамматики. Часть 1.

Задача 1. Докажите, что КС-грамматика с множеством правил

1. $S \to aSb|bSa|SS|\varepsilon$,
2. $S \to aB|bA \quad A \to aS|bAA|a \quad B \to bS|aBB|b$,

порождает язык всех слов с одинаковым числом вхождений букв a и b.

Задача 2. Постройте КС-грамматику без ε-правил и правил–переименований, эквивалентную грамматике с множеством правил

$$S \to aSbb|T \qquad T \to bTaa|S|\varepsilon\ .$$

Задача 3. Постройте КС-грамматику для языка правильных скобочных выражений с двумя типами скобок $(,)$ и $[,]$. Приведите построенную грамматику к нормальной форме Хомского.

Задача 4. Постройте КС-грамматики в нормальной форме Хомского для следующих языков

1. $\{a^n b^{2n} c^k : k, n \geq 1\}$;
2. $\{a^n b^k c^n : k, n \geq 1\}$;
3. $\{a^k b^m c^n : k, m, n \geq 1, 2k \geq n\}$;
4. все слова в алфавите $\{a, b\}$, не являющиеся палиндромами;
5. все слова в алфавите $\{a, b\}$, в которых буква a встречается не более чем в два раза чаще, чем буква b;
6. $L(a^* b^* c^*) \setminus \{a^n b^n c^n : n \geq 0\}$.

Решение задачи 1.1. ($\Rightarrow$) Индукцией по длине n левостороннего грамматического вывода покажем, что в любом терминальном слове w, выводимом из стартового нетерминала S, число вхождений буквы a равно числу вхождений буквы b.

Базис. $n = 1$. Единственный успешный грамматический вывод длины 1: $S \xrightarrow{G} \varepsilon$. Для пустого слова все очевидно.

Индуктивный переход. $n \to n + 1$. Рассмотрим произвольный успешный левосторонний вывод длины $n + 1$. Возможны 3 варианта вывода.

1. $S \xrightarrow{G} aSb \xrightarrow{G} \cdots \xrightarrow{G} awb$.
   Тогда вывод слова w из стартового нетерминала S проводится за n шагов, и по индуктивному предположению в нем равное число вхождений букв a и b.

2. $S \xrightarrow{G} bSa \xrightarrow{G} \cdots \xrightarrow{G} bwa$.
   Аналогично предыдущему пункту.

3. $S \xrightarrow{G} SS \xrightarrow{G} \cdots \xrightarrow{G} uS \xrightarrow{G} \cdots \xrightarrow{G} uv$.
   Тогда вывод слов u и v из стартового нетерминала S проводится менее, чем за n шагов, и по индуктивному предположению в них равное число вхождений букв a и b.

($\Leftarrow$) Покажем индукцией по n, что любое слово длины $2n$ с равным числом вхождений букв a и b порождается грамматикой G. Рассмотрим произвольное слово w указанного вида.

Базис. Если $|w| = 0$, то очевидно, что $w = \varepsilon \in L(G)$.

Индуктивный переход. Рассмотрим слово w длины $2n + 2$ и допустим, что $w = aw'$ (если $w = bw'$, то рассуждения аналогичны). Если $w' = ub$, то слово u длины $2n$ имеет равное число вхождений букв a и b. По индуктивному предположению $S \xrightarrow{G}_* u$. Поэтому $S \xrightarrow{G} aSb \xrightarrow{G}_* aub = w$.

Если $w = aua$, то существует такое разбиение слова $u = u'u''$, что в подсловах u' и u'' число вхождений b на 1 больше числа вхождений a. Тогда слова au' и $u''a$ имеют длину не большую $2n$ и равное число вхождений в них букв a и b. По индуктивному предположению $S \xrightarrow{G}_* a'u$ и $S \xrightarrow{G}_* u''au$. Поэтому $S \xrightarrow{G} SS \xrightarrow{G}_* au'S \xrightarrow{G}_* au'u''a = w$. $\square$

Решение задачи 2. 1) Избавимся от правил–переименований. Будем добавлять в грамматику для каждой пары правил — правила-переименования $N \to M$ и всякого правила вида $M \to \alpha$ — новое правило $N \to \alpha$. Продолжаем делать это, до тех пор пока новые правила не перестанут появляться. После этого устраним все правила–переименования.

Новые правила: $S \to bTaa,\ S \to S,\ S \to \varepsilon,\ T \to aSbb,\ T \to T$.

Новая грамматика: $S \to aSbb|bTaa|\varepsilon,\ T \to bTaa|aSbb|\varepsilon$.

2) Избавимся от ε-правил. Будем добавлять в грамматику для каждой пары правил — ε-правила $N \to \varepsilon$ и всякого правила вида $M \to \alpha N \beta$ — новое правило $M \to \alpha\beta$. Продолжаем делать это, до тех пор пока новые правила не перестанут появляться. После этого устраним все ε-правила $N \to \varepsilon$, где $N \neq S$.

Новые правила: $S \to abb,\ S \to baa,\ T \to baa,\ T \to abb$.

Новая грамматика: $S \to aSbb|bTaa|abb|baa|\varepsilon,\ T \to bTaa|aSbb|baa|abb$.

Решение задачи 4.1. Нетрудно показать (используя тот же метод доказательства, что и при решении задачи 1), что язык $L = \{a^n b^{2n} c^k : k, n \geq 1\}$ порождается грамматикой с множеством правил:

$$S \to DC,\ D \to aDbb|abb,\ C \to cC|c.$$

Избавимся от всех несущественных правил (их нет).

Введем вспомогательные нетерминалы 1-го вида и правила для них:

$$T_a \to a,\ T_b \to b,\ T_c \to c,$$

и заменим во всех остальных правила терминальные буквы соответствующими нетерминалами:

$$D \to T_a D T_b T_b | T_a T_b T_b,\ C \to T_c C | T_c.$$

Вводя вспомогательные нетерминалы сократим длину правых частей правил:

$$\begin{array}{lcl} D \to T_a D T_b T_b & \Rightarrow & D \to T_a D_1',\ D_1' \to D D_1'',\ D_1'' \to T_b T_b, \\ D \to T_a T_b T_b & \Rightarrow & D \to T_a D_2',\ D_2' \to T_b T_b. \end{array}$$

Избавимся правил переименований (их нет), ε-правил (их нет).

Получим грамматику в нормальной форме Хомского:

$$\begin{array}{lll} S \to DC, & & \\ D \to T_a D_1', & D_1' \to D D_1'', & D_1'' \to T_b T_b, \\ D \to T_a D_2', & D_2' \to T_b T_b, & \\ C \to T_c C, & C \to c, & \\ T_a \to a, & T_b \to b, & T_c \to c. \end{array}$$