1. Приведите примеры не менее чем пяти языков в алфавите
$$\Sigma = \{a, b, c, d, e\}.$$

Теория.

Определение. Языком в алфавите Σ называется множество цепочек в Σ.

Решение.

$$L_1 = \{ab, bc, cd, de\}$$
$$L_2 = \{a^i b^i c^i d^i e^i \mid i > 0\}$$
$$L_3 = \{\omega\omega^R \mid \omega \in \{a, b, c, d, e\}^*\}$$
$$L_4 = \{(ae)^i (bdc)^j \mid j - 1 = i \geqslant 0\}$$
$$L_5 = \{dead, bad\}$$

2. Напишите регулярные выражения для следующих языков:

(а) множество цепочек над алфавитом $\{a, b, c\}$, содержащих хотя бы один символ a и хотя бы один символ b;

(б) множество цепочек из нулей и единиц, в которых десятый от правого края символ равен 1;

(в) множество цепочек из нулей и единиц, содержащих не более одной пары последовательных единиц.

Теория.

Определение. Регулярные выражения в алфавите Σ и регулярные множества, которые они *обозначают*, определяются рекурсивно следующим образом:

(1) $\varnothing$ — регулярное выражение, обозначающее регулярное множество $\varnothing$,

(2) ε — регулярное выражение, обозначающее регулярное множество $\{\varepsilon\}$,

(3) если $a \in \Sigma$, то a — регулярное выражение, обозначающее регулярное множество $\{a\}$,

(4) если p и q — регулярные выражение, обозначающие регулярные множества P и Q соответсвенно, то

  (a) $(p + q)$ — регулярное выражение, обозначающее $P \cup Q$;

  (b) (pq) — регулярное выражение, обозначающее PQ;

  (c) $(p)^*$ — регулярное выражение, обозначающее P^*;

(5) ничто другое не является регулярным выражением.

Решение.

(а) $(a + b + c)^*((a^+(a + c)^*b^+) + (b^+(b + c)^*a^+))(a + b + c)^*$

(б) $(0 + 1)^*1(0 + 1)^9$

(в) $(0 + 10)^*(11)?(0 + 01)^*$

3. Постройте праволинейные грамматики для языков, состоящих из

(а) идентификаторов произвольной длины, состоящих из букв и цифр, но начинающихся с буквы (например, как в языке Си);

(б) идентификаторов, которые должны содержать от одного до шести символов и начинаться с буквы I, J, K, L, M или N;

(в) целочисленные константы языка Си (напомним, что целочисленные константы в Си бывают восьмеричные, десятичные и шестнадцатеричные, а в конце может следовать суффикс, задающий тип и знаковость константы);

(г) вещественных констант языка Си;

(д) всех цепочек из нулей и единиц, имеющих чётное число нулей и чётное число единиц.

Теория.

Грамматика G называется *праволинейной*, если каждое правило из P имеет вид $A \to xB$ или $A \to x$, где $A, B \in N, x \in \Sigma^*$.

Решение.

(а) $S \to aA' \mid \ldots \mid zA' \mid AA' \mid \ldots \mid ZA'$
$A' \to aA' \mid \ldots \mid ZA' \mid 0A' \mid \ldots \mid 9A' \mid \varepsilon$

(б) $S \to IA' \mid \ldots \mid NA'$
$A' \to aB' \mid \ldots \mid ZB' \mid 0B' \mid \ldots \mid 9B' \mid \varepsilon$
$B' \to \ldots$
$C' \to \ldots$
$D' \to \ldots$
$E' \to a \mid \ldots \mid Z \mid 0 \mid \ldots \mid 9 \mid \varepsilon$

(в) $S \to 0A' \mid 0D' \mid \ldots \mid 9D'$
$A' \to xB' \mid 0C' \mid \ldots \mid 7C'$
$B' \to 0B' \mid 1B' \mid \ldots \mid 9B' \mid aB' \mid \ldots \mid fB' \mid AB' \mid \ldots \mid FB' \mid$
$0z' \mid 1z' \mid \ldots \mid 9z' \mid az' \mid \ldots \mid fz' \mid Az' \mid \ldots \mid Fz'$
$C' \to OC' \mid \ldots \mid 7C' \mid z'$
$D' \to OD' \mid \ldots \mid 9D' \mid z'$
$z' \to u \mid l \mid ul \mid lu \mid U \mid L \mid UL \mid LU \mid \varepsilon$

(г) $S \to 0S \mid \ldots \mid 9S \mid .A \mid 0A \mid \ldots \mid 9A \mid 0D \mid \ldots \mid 9D$
$A \to 0A \mid \ldots \mid 9A \mid eD \mid e-D \mid e+D \mid 0D \mid \ldots \mid 9D$
$D \to 0D \mid \ldots \mid 9D \mid z'$
$z' \to l \mid L \mid \varepsilon$

(д) $A \to 0C \mid 1B \mid \varepsilon$
$B \to 0D \mid 1A$
$C \to 0A \mid 1D$
$D \to 0B \mid 1C$

4. Постройте КС-грамматики, порождающие

(а) все цепочки из нулей и единиц с одинаковым числом тех и других;
(б) $\{a_1a_2\dots a_na_n\dots a_2a_1 \mid a_i \in \{0,1\}, 1 \leqslant i \leqslant n\}$;
(в) $\{0^i1^j \mid i \neq j$ и $i,j \geqslant 0\}$;
(г) всевозможные последовательности правильно расставленных скобок.

Теория.

Грамматика G называется *контекстно-свободной*, если каждое правило из P имеет вид $A \to \alpha$, где $A \in N, \alpha \in (N \cup \Sigma)^*$.

Решение.

(а) $S \to 0AS \mid A0S \mid \varepsilon$
$A \to 0AA \mid AA0 \mid 1$

(б) $S \to 0S0 \mid 1S1 \mid \varepsilon$

(в) $S \to 0S1 \mid A$
$A \to 1B \mid 0C \mid 1 \mid 0$
$B \to 1B \mid 1$
$C \to 0C \mid 0$

(г) $S \to (S) \mid ()S \mid \varepsilon$

5. Следующая грамматика порождает язык регулярного выражения $0^*1(0+1)^*$.

$$S \to A1B$$
$$A \to 0A \mid \varepsilon$$
$$B \to 0B \mid 1B \mid \varepsilon$$

Запишите левый и правый вывод, а также нарисуйте деревья вывода следующих цепочек:

(а) 00101;
(б) 1001;
(в) 00011.

Теория.

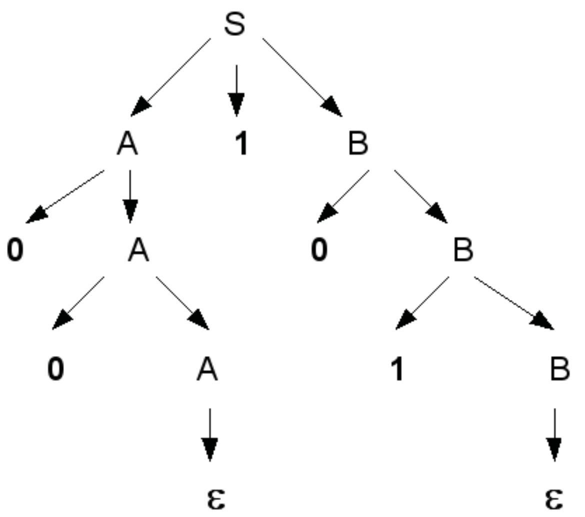
Определение. Вывод, при котором каждый раз правило применяется к самому левому нетерминалу, называется *левым выводом.* Аналогично определяется *правый вывод.*

Определение. Дерево вывода в грамматике $G = (N, \Sigma, P, S)$ — это помеченное упорядоченное дерево, каждая вершина которого помечена символом из множества $N \cup \Sigma \cup \{\varepsilon\}$. Если внутренняя вершина помечена символом A, а ее прямые потомки — символами $X_1X_2 \dots X_n$, то $A \to X_1X_2 \dots X_n$ — правило этой грамматики.

Решение.

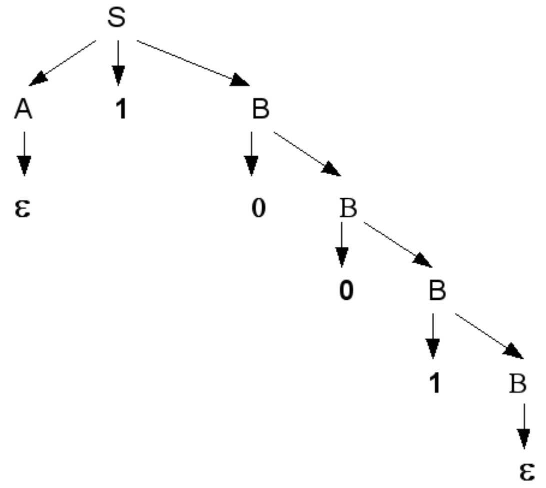
(а) левый вывод: $S \to A1B \to 0A1B \to 00A1B \to 001B \to 0010B \to 00101B \to 00101$

правый вывод: $S \to A1B \to A10B \to A101B \to A101 \to 0A101 \to 00A101 \to 00101$



(б) левый вывод: $S \to A1B \to 1B \to 10B \to 100B \to 1001B \to 1001$

правый вывод: $S \to A1B \to A10B \to A100B \to A1001B \to A1001 \to 1001$



(в) левый вывод: $S \to A1B \to 0A1B \to 00A1B \to 000A1B \to 0001B \to 00011B \to 00011$
правый вывод: $S \to A1B \to A11B \to A11 \to 0A11 \to 00A11 \to 000A11 \to 00011$