Лабораторная работа № 1. Распознавание типов формальных языков и грамматик

Цель:

- закрепить понятия «алфавит», «цепочка», «формальная грамматика» и «формальный язык», «выводимость цепочек», «эквивалентная грамматика»;
- сформировать умения и навыки распознавания типов формальных языков и грамматик по классификации Хомского, построения эквивалентных грамматик.

Теоретические основы

Определение 1. Алфавитом V называется конечное множество символов.
Определение 2. Цепочкой α в алфавите V называется любая конечная последовательность символов этого алфавита.
Определение 3. Цепочка, которая не содержит ни одного символа, называется пустой цепочкой и обозначается ε.
Определение 4. Формальное определение цепочки символов в алфавите:

1) ε - цепочка в алфавите V;

2) если α - цепочка в алфавите V и a – символ этого алфавита, то αa – цепочка в алфавите V;

3) β - цепочка в алфавите V тогда и только тогда, когда она является таковой в силу утверждений 1) и 2).

Определение 5. Длиной цепочки α называется число составляющих ее символов (обозначается | α |).
Обозначим через V^* множество, содержащее все цепочки в алфавите V, включая пустую цепочку ε, а через $V+$ - множество, содержащее все цепочки в алфавите V, исключая пустую цепочку ε.
Пример 1. Пусть $V = \{0, 1\}$, тогда $V^* =\{ ε,0,1,00,01,10,111, 000 , … \}$, а $V^+ = \{ 0,1,00,01,10,11, 000…\}$.

Определение 6. Формальной грамматикой называется четверка вида:

$$G = \{T,N,P,S\}, \text{ где}$$

N - конечное множество нетерминальных символов грамматики (обычно прописные латинские буквы);

T - множество терминальных символов грамматики (обычно строчные латинские буквы, цифры, и т.п.), $T \cap N = \varnothing$;

P – множество правил вывода грамматики, являющееся конечным подмножеством множества $(T \cup N)^+ \times (T \cup N)^*$; элемент (α, β) множества P называется правилом вывода и записывается в виде α→ β (читается: «из цепочки α выводится цепочка β»);

S - начальный символ грамматики, $S \in N$.

Для записи правил вывода с одинаковыми левыми частями вида $\alpha \to \beta_1, \alpha \to \beta_2, ..., \alpha \to \beta_n$ используется сокращенная форма записи $\alpha \to \beta_1 | \beta_2 | ... | \beta_n$.
Пример 2. Грамматика $G1=(\{0, 1\}, \{A, S\}, P_1, S)$, где множество P_1 состоит из правил вида: 1) $S \to 0A1$; 2) $0A \to 00A1$; 3) $A \to ε$.

Определение 7. Цепочка $\beta \in (T \cup N)^*$ непосредственно выводима из цепочки $\alpha \in (T \cup N)^+$ в грамматике G = {T,N,P,S} (обозначается: $\alpha \Rightarrow \beta$), если $\alpha = \xi_1 \gamma \xi_2$ и $\beta = \xi_1 \delta \xi_2$, где ξ_1 , ξ_2 , $\delta \in (T \cup N)^*$, $\gamma \in (T \cup N)^+$ и правило вывода $\gamma \rightarrow \delta$ содержится во множестве *P*.
Определение 8. Цепочка $\beta \in (T \cup N)^*$ выводима из цепочки $\alpha \in (T \cup N)^+$ в грамматике G = {T,N,P,S}, если существует последовательность цепочек $\gamma_0, \gamma_1, ..., \gamma_n$ $(n \geq 0)$ такая, что $\alpha \Rightarrow \gamma_0 \Rightarrow \gamma_1 \Rightarrow ... \Rightarrow \gamma_n = \beta$.
Пример 3. В грамматике G_1 $S \Rightarrow^* 000111$, т.к. существует вывод $S \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 000A111 \Rightarrow 000111$.

Определение 9. Языком, порожденным грамматикой G = {T,N,P,S} называется множество всех цепочек в алфавите *T*, которые выводимы из начального символа грамматики *S* с помощью правил множества *P*, т.е. множество $L(G) = \{\alpha \in T^* \mid S \Rightarrow {}^*\alpha\}$.
Пример 4. Для грамматики G_1 $L(G_1) = \{0^n 1^n \mid n > 0\}$.

Определение 10. Цепочка $\alpha \in (T \cup N)^*$, для которой существует вывод $S \Rightarrow^* \alpha$, называется сентенциальной формой в грамматике G = {T,N,P,S} .
Определение 11. Грамматики G_1 и G_2 называются эквивалентными, если $L(G_1) = L(G_2)$.
Пример 5. Для грамматики G_1 эквивалентной будет грамматика $G_2 = (\{0, 1\}, \{S\}, P_2, S)$, где множество правил вывода P_2 содержит правила вида $S \rightarrow 0S1 \mid 01$.

Классификация грамматик по Хомскому

Тип 0. Грамматика G = {T,N,P,S} называется грамматикой типа 0, если на ее правила вывода не наложено никаких ограничений, кроме тех, которые указаны в определении грамматики.
Тип 1. Грамматика G = {T,N,P,S} называется контекстно-зависимой грамматикой (КЗ-грамматикой), если каждое правило вывода из множества *P* имеет вид $\alpha \rightarrow \beta$, где $\alpha \in (T \cup N)^+$, $\beta \in (T \cup N)^*$ и $|\alpha| \leq |\beta|$.
Тип 2. Грамматика G = {T,N,P,S} называется контекстно-свободной грамматикой (КС-грамматикой), если ее правила вывода имеют вид: $A \rightarrow \beta$, где $A \in N$ и $\beta \in T^*$
Тип 3. Грамматика G = {T,N,P,S} называется регулярной грамматикой (Р-грамматикой) выровненной вправо, если ее правила вывода имеют вид $A \rightarrow aB \mid a$, где $a \in T$, $A, B \in N$.
Грамматика G = {T,N,P,S} называется регулярной грамматикой (Р-грамматикой) выровненной влево, если ее правила вывода имеют вид $A \rightarrow Ba \mid a$, где $a \in T$, $A, B \in N$.

a Ba A | → , где *N T V B A V a* □ □ , ; .
Определение 12. Язык $L(G)$ называется языком типа *k*, если его можно описать грамматикой типа *k*, где *k* – максимально возможный номер типа грамматики.

Пример 6. Примеры различных типов формальных языков и грамматик по классификации Хомского. Терминалы будем обозначать строчными символами, нетерминалы – прописными буквами, начальный символ грамматики – *S*.
а) Язык типа 0 $L(G) = \{a^n b^{n^2-1} \mid n \geq 1\}$ определяется грамматикой с правилами вывода:

1) $S \rightarrow aaCFD$;
2) $AD \rightarrow D$;
3) $F \rightarrow AFB \mid AB$;
4) $Cb \rightarrow bC$;
5) $AB \rightarrow bBA$;

6) $CB \to C$;
7) $Ab \to bA$;
8) $bCD \to \varepsilon$.

б) Контекстно-зависимый язык $L(G) = \{a^n b^n c^n \mid n \geq 1\}$ определяется грамматикой с правилами вывода:

1) $S \to aSBC \mid abc$;
2) $bC \to bc$;
3) CB → BC;
4) cC → cc;
5) $BB \to bb$.

в) Контекстно-свободный язык $L(G) = \{(ab)^n (cb)^n \mid n > 0\}$ определяется грамматикой с правилами вывода:

1) $S \to aQb \mid accb$;
2) $Q \to cSc$.

г) Регулярный язык $L(G) = \{w \perp \mid w \in (a,b)^+,$ где нет двух рядом стоящих $a\}$ определяется грамматикой с правилами вывода:

1) $S \to A\perp \mid B\perp$;
2) $A \to a \mid Ba$;
3) $B \to b \mid Bb \mid Ab$.

Постановка задачи к лабораторной работе № 1

При выполнении лабораторной работы следует реализовать следующие действия:

1) составить грамматику, порождающую формальный язык, заданный в соответствии с вариантом;
2) определить тип формальной грамматики и языка по классификации Хомского;
3) разработать программное обеспечение, распознающее принадлежность введенного слова составленной грамматике.

Варианты индивидуальных заданий:

1) $L(G)=\{a^n b^m c^k \mid n, m, k>0\}$
2) $L(G)=\{(ab)^n(cb)^m \mid n, m\geq 0\}$
3) $L(G)=\{0^n(10)^m \mid n, m\geq 0\}$
4) $L(G)=\{wcwcw \mid w\in \{a, b\}^+\}$
5) $L(G)=\{c^{2n}d^n \mid n>0\}$
6) $L(G)=\{l+l\text{-}l \mid l \in\{a, b\}^+\}$
7) $L(G)=\{(10)^{n-1}(01)^{n+1} \mid n>0\}$
8) $L(G)=\{(ac)^n \mid n>0, a\in\{b, d\}, c\in\{+, -\}\}$
9) $L(G)=$ *{100, 0100, 1100, 00100, 01100, 11100, 000100, 001100, 011100, 111100, 101100, 0000100, 0001100, 0011100, 0111100, 0101100, 1111100, 1101100, 1010100,…}*
10) $L(G)=\{a_1a_2\ldots a_na_n\ldots a_2a_1 \mid a_i\in\{0, 1\}\}$
11) $L(G)=\{a_1a_2\ldots a_na_1a_2\ldots a_n \mid a_i\in\{c, d\}\}$
12) $L(G)=\{ab.b \mid a_i\in\{+, -\}, b\in\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}^+\}$
13) *L(G) = {АБВВ, ААББВВВ, АААБББВВВВ,... }*
14) *L(G) = {ААБВ, АААББВВ, ААААБББВВВ, ...}*
15) *L(G) = {АББВ, ААББББВВ, АААББББББВВВ....}*

16) $L(G)$ = {*А, ААБ, АААБББ, ААААББББВВ*,...}
17) $L(G)$ = {*В, БВВ, АББВВВ, ААББББВВВВ*,**...**}
18) $L(G)$ = {*Б, АББ, ААББББ, АААББББВВ*,... }
19) $L(\mathrm{G})$ = {*01010, 0110110, 011101110,…*}
20) $L(\mathrm{G})$ = {*АБВАБ, АБАБВАБАБ, АБАБАБВАБАБАБ*,…}
21) $L(\mathrm{G})$ = { *1, 11, 01, 111, 101, 011, 1111, 1101, 0101, 0111, 11111, 11101, 10111, 10101, 01011, 011111, …*}
22) $L(\mathrm{G})$ = { *1, 01, 11, 001, 011, 101, 111,…*}
23) $L(\mathrm{G})$ = { *01, 001, 0001, 1001, 00001, 10001, 000001, 100001, 101001*,…}
24) $L(\mathrm{G})$ = {ε, *А, Б, В, АА, АБ, АВ, БА, ББ, БВ, ВА, ВБ, ВВ, ААА, ААБ, ААВ, АБА, АББ, АБВ, АВА, АВБ, АВВ, БАА, БАБ, БАВ, ББА, БББ, ББВ, БВА, БВБ, БВВ, ВАА, ВАБ, ВАВ,ВБА, ВББ, ВБВ, ВВА, ВВБ, ВВВ,…* }
25) $L(\mathrm{G})$ = {*АБ, АВ, АБАБ, АБАВ, АВАБ, АВАВ, АБАБАБ, АБАБАВ, АБАВАБ, АБАВАВ*, *АВАБАБ, АВАБАВ, АВАВАБ, АВАВАВ*,… }
26) $L(G)=\{a_1a_2\ldots a_na_n\ldots a_2a_1 \mid a_i\in\{0, 1\}\}$
27) $L(G)=\{a_1a_2\ldots a_na_1a_2\ldots a_n \mid a_i\in\{c, d\}\}$
28) $L(G)=\{ab.b \mid a_i\in\{+, -\}, b\in\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}^+\}$
29) $L(G)=\{a^nb^mc^k \mid n, m, k>0\}$
30) $L(G)=\{(ab)^n(cb)^m \mid n, m\geq 0\}$
31) $L(G)=\{0^n(10)^m \mid n, m\geq 0\}$
32) $L(G)=\{wcwcw \mid w\in\{a, b\}^+\}$