**Лабораторная работа №5**

Распознавание типов формальных языков и грамматик

1) Составить грамматику, порождающую формальный язык, заданный в соответствии с вариантом.

2) Определить тип формальной грамматики и языка по классификации Хомского.

3) Разработать программное средство, распознающее тип введенной пользователем грамматики по классификации Хомского.

Варианты индивидуальных заданий

|  |  |
| --- | --- |
| Вариант | Формальный язык |
| 1 | *L*(*G*)={*an bm  cm*| *n*, *m*, *k*>0} |
| 2 | *L*(*G*)={(*ab*)n (*cb*)m| *n*, *m*≥0} |
| 3 | *L*(*G*)={0n(10)m| *n*, *m*≥0} |
| 4 | *L*(*G*)={*wcwcw* | *w*∈{*a*, *b*}+} |
| 5 | *L*(*G*)={*c2n* *dn*| *n*>0} |
| 6 | (*G*)={*l*+*l*-*l* | *l* ∈{*a*, *b*}+} |
| 7 | *L*(*G*)={(10)n-1(01)*n*+1 | *n*>0} |
| 8 | *L*(*G*)={(*ac*)*n* | *n*>0, *a*∈{*b*, *d*}, *c*∈{+, -}} |
| 9 | *L*(*G*)={⊥(010)*n*⊥ | *n*>0} |
| 10 | *L*(*G*)={*a*1*a*2…*anan*…*a*2*a*1| *ai*∈{0, 1}} |
| 11 | *L*(*G*)={*a*1*a*2…*ana*1*a*2…*an*| *ai*∈{*c*, *d*}} |
| 12 | *L*(*G*)={*ab*.*b* | *ai*∈{+, -}, *b*∈{0, 1, 2, 3, 4, 5, 6, 7, 8, 9}+} |

***Необходимые теоретические сведения***

**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 *V*:

1) ε - цепочка в алфавите *V*;

2) если α - цепочка в алфавите *V* и *а* – символ этого алфавита, то α*а* – 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.

**Определение 6.** Формальной грамматикой называется четверка вида: *G* = (*VT* ,*VN* , *P*, *S*),

где *VN* - конечное множество нетерминальных символов грамматики (обычно прописные латинские буквы);

*VT* - множество терминальных символов грамматики (обычно строчные латинские буквы, цифры, и т.п.), *VT* ∩*VN =*∅;

*Р* – множество правил вывода грамматики, являющееся конечным подмножеством множества (*VT* ∪ *VN*)*+* × (*VT* ∪ *VN*)*\**; элемент

(α*,* β) множества *Р* называется правилом вывода и записывается в виде α→β (читается: «из цепочки α выводится цепочка β»);

*S -* начальный символ грамматики, *S* ∈*VN*.

**Определение 7.** Цепочка β ∈ (*VT* ∪ *VN*)*\** непосредственно выводима из цепочки α ∈(*VT* ∪*VN* )+ в грамматике *G* = (*VT* ,*VN* , *P*, *S*) (обозначается: α⇒β), если α =ξ1γξ2 и β=ξ1δξ2, где ξ1,ξ2,δ ∈ (*VT* ∪*VN* ) \* , γ∈(*VT* ∪*VN* )+ и правило вывода γ →δ содержится во множестве *Р*.

**Определение 8.** Цепочка β ∈ (*VT* ∪ *VN*)*\** выводима из цепочки α∈(*VT* ∪*VN* )+ в грамматике *G* = (*VT* ,*VN* , *P*, *S*) (обозначается α⇒\*β), если существует последовательность цепочек γ0 ,γ1 ,…,γ *n* (*n*≥*0*) такая, что α =γ0 ⇒γ1 ⇒ … ⇒γ*n* = β .

**Определение 9.** Языком, порожденным грамматикой *G* = (*VT* ,*VN* , *P*, *S*)**,** называется множество всех цепочек в алфавите *VT*, которые выводимы из начального символа грамматики *S* c помощью правил множества *Р*, т.е. множество *L*(*G*) = {α ∈*VT*\* | *S* ⇒ \*α}.

**Определение 10.** Цепочка α ∈(*VT* ∪*VN* )\* , для которой существует вывод *S*⇒\*α, называется сентенциальной формой в грамматике *G* = (*VT* ,*VN* , *P*, *S*).

**Определение 11.** Грамматики *G*1 и *G*2 называются эквивалентными, если *L*(*G*1) = *L*(*G*2).

***Классификация грамматик по Хомскому***

**Тип 0.** Грамматика *G* = (*VT* ,*VN* , *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* = (*VT* ,*VN* , *P*, *S*) называется контекстно-зависимой грамматикой (КЗ-грамматикой), если каждое правило вывода из множества *Р* имеет вид α→β, где α ∈ (*VT* ∪ *VN*)+, β ∈ (*VT* ∪ *VN*)\* и |α| ≤ |β|.

**Тип 2.** Грамматика *G* = (*VT* ,*VN* , *P*, *S*) называется контекстно-свободной

грамматикой (КС-грамматикой), если ее правила вывода имеют вид: *A* → β, где *A*∈*VN* и β ∈*V* \*.

**Тип 3.** Грамматика *G* = (*VT* ,*VN* , *P*, *S*) называется регулярной грамматикой (Р-грамматикой) выровненной вправо, если ее правила вывода имеют вид *A* → *aB* | *a* , где *a* ∈*VT*;*A*, *B*∈*VN*.

Грамматика *G* = (*VT* ,*VN* , *P*, *S*) называется регулярной грамматикой (Р-грамматикой) выровненной влево, если ее правила вывода имеют вид *A* → *Ba* | *a* , где *a* ∈*VT*;*A*, *B*∈*VN*.

**Определение 12.** Язык *L*(*G*) называется языком типа *k*, если его можно описать грамматикой типа *k*, где *k* – максимально возможный номер типа грамматики.

**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*)={*a2*≥1} определяется грамматикой с правилами вывода:

1) *S* → *aaCFD*; 2) *AD* → *D*;

3) *F* → *AFB | AB*; 4) *Cb* → *bC*;

5) *AB* → *bBA*; 6) *CB* → *C*;

7) *Ab* → *bA*; 8) *bCD* → ε.

б) Контекстно-зависимый язык *L*(*G*)={*an  bn  cn*| *n*≥1} определяется грамматикой с правилами вывода:

1) *S* → *aSBC | abc* ; 2) *bC* → *bc*;

3) CB → BC; 4) cC → cc;

5) *BB* → *bb*.

в) Контекстно-свободный язык *L*(*G*)={(*ab*)n (*cb*)n| *n*>0 } определяется грамматикой с правилами вывода:

1) *S* → *aQb | accb*;

2) *Q* → *cSc*.

г) Регулярный язык *L*(*G*)={ω⊥ | ω∈{*a, b*}+, где нет двух рядом стоящих *а*}

определяется грамматикой с правилами вывода:

1) *S* → *A*⊥ *| B*⊥;

2) *A* → *a | Ba*;

3) *B* → *b | Bb | Ab*.