**1 Лабораторная работа № 1.**

**Распознавание типов формальных языков и грамматик**

**Цель:** - закрепить понятия **«**алфавит», «цепочка», «формальная грамма-

тика» и «формальный язык», «выводимость цепочек», «эквивалентная грамматика»;

**-** сформировать умения и навыки распознавания типов формальных языков и грамматик по классификации Хомского, построения эквивалентных грамматик.

**Основы теории**

**Определение 1.1.** Алфавитом *V* называется конечное множество символов.

**Определение 1.2.** Цепочкой α в алфавите *V* называется любая конечная

последовательность символов этого алфавита.

Определение 1.3**. Цепочка, которая не содержит ни одного символа, называется пустой цепочкой и обозначается** ε**.**

**Определение 1.4.** 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 та-

ковой в силу утверждений 1) и 2).

**Определение 1.5.** 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 1.1.** Пусть *V* = {1, 0}, тогда *V*\* = {ε , 0,1, 00, 01,10,11, 000,K}, а

*V* + = {0,1, 00, 01,10,11, 000,K}.

**Определение 1.6.** Формальной грамматикой называется четверка вида:

*G* = (*VT* ,*VN* , *P*, *S*) , (1.1)

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 записывает-

5

ся в виде α→β (читается: «из цепочки α выводится цепочка

β»);

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

Для записи правил вывода с одинаковыми левыми частями вида

α →β1,α →β 2,K,α →β *n* используется сокращенная форма записи

α →β1 | β 2 |K| β *n* .

**Пример 1.2.** Грамматика *G*1*=*({0, 1}, {*A*, *S*}, *P*1, *S*), где множество *Р*1 со-

стоит из правил вида: 1) *S*→ 0*A*1; 2) 0*A*→ 00*A*1; 3) *A*→ε*.*

**Определение 1.7.** Цепочка β ∈ (*VT* ∪ *VN*)*\** непосредственно выводима из

цепочки α ∈(*VT* ∪*VN* )+ в грамматике *G* = (*VT* ,*VN* , *P*, *S*) (обозначается: α⇒β),

если α =ξ1γξ 2 и β =ξ1δξ 2, где \*

ξ1,ξ 2,δ ∈(*VT* ∪*VN* ) , ∈( ∪ )+ γ *VT VN* и правило

вывода γ →δ содержится во множестве *Р*.

Определение 1.8. Цепочка β ∈ (*VT* ∪ *VN*)*\** выводима из цепочки α ∈(*VT* ∪*VN* )+ в

грамматике *G* = (*VT* ,*VN* , *P*, *S*) (обозначается α⇒\*β), если существует последо-

вательность цепочек γ 0 ,γ 1 ,K,γ *n* (*n*≥*0*) такая, что α =γ 0 ⇒γ 1⇒K⇒γ *n* = β .

**Пример 1.3.** В грамматике *G*1 *S*⇒\*000111, т.к. существует вывод

*S* ⇒0*A*1⇒00*A*11⇒000*A*111⇒000111.

**Определение 1.9.** Языком, порожденным грамматикой *G* = (*VT* ,*VN* , *P*, *S*) **,**

называется множество всех цепочек в алфавите *VT*, которые выводимы из на-

чального символа грамматики *S* c помощью правил множества *Р*, т.е. множест-

во *L*(*G*) = {α ∈*VT*\* | *S* ⇒\*α}.

**Пример 1.4.** Для грамматики *G*1 *L*(*G*1)*=*{0*n*1*n* | *n>*0}.

Определение 1.10. Цепочка α ∈(*VT* ∪*VN* )\* , для которой существует вывод

*S*⇒\*α, называется сентенциальной формой в грамматике *G* = (*VT* ,*VN* , *P*, *S*) .

**Определение 1.11.** Грамматики *G*1 и *G*2 называются эквивалентными, ес-

ли *L*(*G*1) = *L*(*G*2 ) .

**Пример 1.5.** 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* → 0*S*1 | 01.

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

Пример 1.6. 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*)={ 2 1 | 1 2

*a bn* − *n* ≥ } определяется грамматикой с прави-

лами 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*)={*anbncn* | *n*≥1} определяется грамма-

тикой с правилами вывода:

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

Тип 0

КЗ

КС

Р

7

**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*.

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 средство, реализующее распознавание введенной цепочки и ее синтаксический разбор.

Варианты заданий:

1 *L*(*G*)={*anbmck* | *n*, *m*, *k*>0}

2 *L*(*G*)={(*ab*)*n*(*cb*)*m* | *n*, *m*≥0}

3 *L*(*G*)={0*n*(10)*m* | *n*, *m*≥0}

4 *L*(*G*)={*wcwcw* | *w*∈{*a*, *b*}+}

5 *L*(*G*)={*c*2*ndn* | *n*>0}

6 *L*(*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}+}