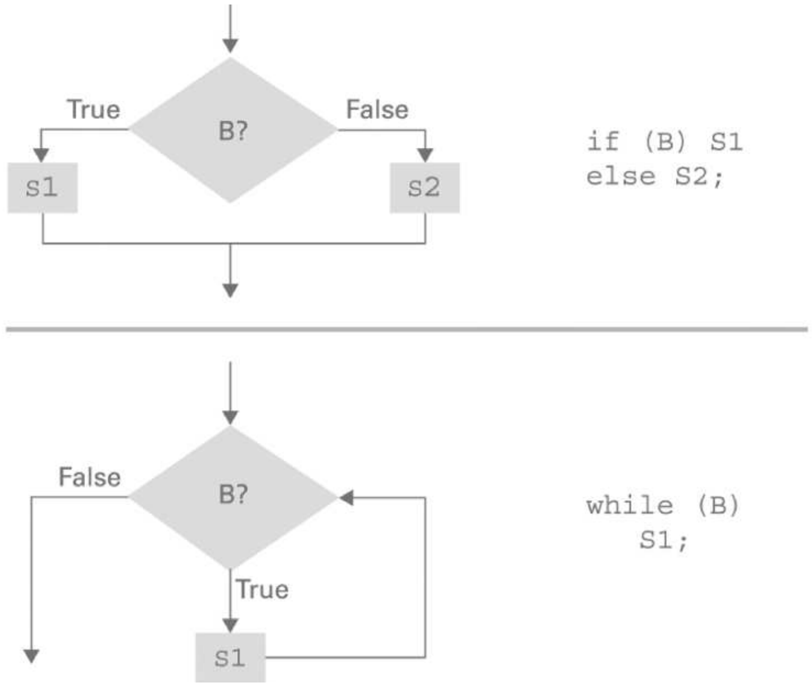
Д. Ж. Корзун

Практикум по формальным грамматикам и языкам



Федеральное агентство по образованию
Государственное образовательное учреждение
высшего профессионального образования
ПЕТРОЗАВОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Д. Ж. Корзун

Практикум по формальным грамматикам и языкам

Учебное пособие

Петрозаводск
Издательство ПетрГУ
2004

УДК 519.682
ББК 22.18
К666

Печатается по решению редакционно-издательского совета
Петрозаводского государственного университета

Корзун, Д. Ж.

К666 Практикум по формальным грамматикам и языкам: Учеб. пособие / Д. Ж. Корзун; ПетрГУ. — Петрозаводск: Изд-во ПетрГУ, 2004. — 46 с.

ISBN 5-8021-0424-4

Учебное пособие содержит краткие теоретические сведения и набор типовых упражнений для практических, лабораторных и самостоятельных занятий. Материал пособия позволяет закрепить базовый теоретический аппарат, вырабатывая практические навыки решения задач, возникающих при построении трансляторов языков программирования, и в других областях использования и развития теории формальных грамматик и языков.

Пособие адресовано студентам специальностей, учебные планы которых предполагают изучение каких-либо разделов «теории формальных грамматик, языков и автоматов», «языков программирования» или «методов трансляции», а также преподавателям, ведущим практические и лабораторные занятия по соответствующим дисциплинам.

УДК 519.682
ББК 22.18

ISBN 5-8021-0424-4

Оглавление

Введение

Теория формальных языков является ярким примером весьма практичной математической теории, знание которой чрезвычайно важно для специалистов в области информатики и прикладной математики, для системных и прикладных программистов. Важным достижением теории формальных языков явилась разработка методов эффективного решения задач лексического и синтаксического анализа текста. Последний представлен на языке, который задается формальной грамматикой.

Материал данного учебного пособия предназначен для закрепления базовых аспектов теории формальных грамматик и языков, для выработки практических навыков решения задач, возникающих как при построении трансляторов языков программирования, так и в других областях использования и развития теории формальных языков и методов трансляции.

Пособие написано по материалам занятий по дисциплинам «Языки программирования и методы трансляции», «Формальные языки и компиляторы», «Разработка компиляторов», читаемых автором на математическом факультете Петрозаводского государственного университета студентам специальностей «Прикладная математика и информатика» и «Информационные системы и технологии». Часть из представленных задач являются оригинальными, другие были отобраны из различных источников и переработаны автором.

§ 1. Синтаксис и семантика

Любые данные в ЭВМ (файл, таблица, программа, область памяти и т. п.) можно рассматривать как последовательность символов (бит, байт, записей, кластеров, слов и т. п.). С этой точки зрения любой объект при представлении в ЭВМ есть текст из символов некоторого алфавита.

Текст характеризуется синтаксисом и семантикой. Правила, определяющие способ записи текста как цепочку из некоторых символов, называются *синтаксисом*. Правила, определяющие смысловое содержание текста, называются *семантикой*. В случае явного использования синтаксиса говорят о *структурированных текстах*.

Синтаксис программы определяется правилами языка программирования. Семантика программы определяет результат трансляции, например представление исходной программы в машинном коде.

В упражнениях 1.1–1.10 приведены примеры программ на различных языках программирования (см. также www.fact-index.com/h/he/hello_world_program.html). Выполните анализ синтаксиса и предложите схему представления результата трансляции.

Упр. 1.1. BASIC.

```
10 PRINT "Hello World"
20 END
```

Упр. 1.2. Pascal.

```
program Hello(input, output)
begin
  writeln( 'Hello World' )
end.
```

Упр. 1.3. C.

```
#include <stdio.h>
int main(void) {
  printf("Hello, world!\n");
  return 0;
}
```

Упр. 1.4. C++.

```
#include <iostream>
using namespace std;
int main() {
  cout << "Hello, world!" << endl;
  return 0;
}
```

Упр. 1.5. Java.

```
public class HelloWorld {
  public static void main( String[] args ) {
    System.out.println( "Hello World!\n" );
  }
}
```

Упр. 1.6. Assembler (Intel x86, MS-DOS, TASM).

```
MODEL   SMALL
IDEAL
STACK   100H
DATASEG
        MSG DB 'Hello, world!$'
CODESEG
        MOV AX, @data
        MOV DS, AX
        MOV DX, OFFSET MSG
        MOV AH, 09H ; DOS: output ASCII$ string
        INT 21H
        MOV AX, 4C00H
        INT 21H
END
```

Упр. 1.7. Lisp.

```
(defun hello
   (print
      (cons 'Hello (list 'World))))
```

Упр. 1.8. SQL.

```
create table MESSAGE (TEXT char(15));
insert into MESSAGE (TEXT) values ('Hello, world!');
select TEXT from MESSAGE;
drop table MESSAGE;
```

Упр. 1.9. PostScript.

```
/font /Courier findfont 24 scalefont
font setfont
100 100 moveto
(Hello world!) show
showpage
```

Упр. 1.10. TeX.

```
\font\HW=cmr10 scaled 3000
\leftline{\HW Hello world}
\bye
```


Упр. 1.11. Опишите произвольным образом синтаксис следующих конструкций для некоторого языка программирования:

a) объявление переменных;
b) математич. выражение;
c) логическое выражение;
d) оператор присваивания;
e) ветвление (усл. оператор);
f) цикл («для», «пока», «до»);
g) объявление подпрограммы;
h) вызов подпрограммы.

Каким образом каждая конструкция может быть реализована на ЭВМ (результат трансляции)?

Упр. 1.12. Приведите примеры, когда некоторая синтаксическая конструкция может быть реализована (результат трансляции) несколькими различными способами.

Упр. 1.13. Приведите примеры из известных вам языков программирования, когда в рамках одного языка некоторое семантическое действие может быть реализовано несколькими синтаксически различными способами.

Упр. 1.14. Приведите примеры синтаксически правильных программ (или фрагментов программ) на некотором языке программирования, которые являются семантически некорректными.

Упр. 1.15 (А. Ахо, Дж. Ульман [1]**).** Приведите пример программы, которая синтаксически построена правильно, но не задает (всюду определенный) алгоритм.

§ 2. Грамматические правила, формы Бекуса — Наура

Часто структурированный текст можно представить как иерархическую конструкцию из заранее определенных *синтаксических групп*. Каждой группе в тексте соответствует подцепочка, например «подлежащее», «оператор присваивания», «тело подпрограммы» и т. п. Синтаксической группе наивысшего уровня соответствует весь текст. Следующий уровень определяет крупные синтаксические группы, на которые разбивается текст. В свою очередь каждая такая синтаксическая группа может быть по аналогии разделена на меньшие группы — следующие уровни иерархии, пока деление не достигнет наименьших возможных единиц, — *синтаксические группы нижнего уровня* или *лексемы*. Графически такую структуру для конкретного текста удобно оформлять в виде дерева, называемого *синтаксическим деревом*.

Правила, описывающие синтаксические группы и их использование для построения текстов, называются *грамматическими*. Грамматические правила — это один из возможных способов описания синтаксиса текста, восходящий к грамматикам естественных языков.

В качестве метаязыка для представления грамматических правил используют *формы Бекуса — Наура* (металингвистические формы, Backus—Naur forms, BNF или БНФ). В БНФ синтаксическим группам сопоставляются *синтаксические переменные*, которые обозначаются словами естественного языка, заключенными в угловые скобки $\langle\ldots\rangle$ и поясняющими смысл синтаксической группы. Лексемы в БНФ имеют смысл *синтаксических констант*.

Каждая БНФ описывает грамматическое правило построения некоторой синтаксической группы. Синтаксическая переменная записывается в левой части правила, далее идет связка $\to$, имеющая смысл глагола «быть», она соединяет левую и правую части правила[1]. Правая часть состоит из цепочки синтаксических переменных и констант, определяющей, как строится синтаксическая группа из левой части посредством синтаксических групп из правой части. Для более компактного представления БНФ с одинаковыми левыми частями используется связка $|$, имеющая значение «или».

Упр. 2.1. Постройте синтаксические деревья для предложений естественного языка и опишите соответствующие грамматические правила с помощью БНФ.

a) `Мама мыла мылом раму.`
b) `Чистота - залог здоровья, порядок - прежде всего.`
c) `Ехал Грека через реку, видит Грека - в реке рак.`
d) `Eat at pleasure, drink with measure.`
e) `East or West, home is best.`
f) `Исключить, нельзя поставить зачет.`
g) `Тише едешь - дальше будешь ...`
h) `Грация - затрата минимума действий на движение.`

Предлагаемый синтаксис должен допускать более одного предложения. Рассмотрите не менее двух конкретных предложений, отличных от заданного, и осуществите их построение.

[1] Часто для обозначения такой связки также используют `::=` .

Упр. 2.2 (Д. Грис [7])**.** Постройте синтаксические деревья для следующих предложений:

a) `John ate the big peanut.`
b) `John ate the big brown peanut.`
c) `John ate the salted big brown roasted peanut.`

Опишите соответствующие грамматические правила. Дайте только два грамматических правила для новой синтаксической группы ⟨группа существительного⟩, определяемой как ⟨существительное⟩, которому предшествует любое количество прилагательных (в том числе и ни одного).

Упр. 2.3. Какие предложения можно построить из ⟨предл⟩?

$$\begin{array}{rcl} \langle\text{предл}\rangle & \to & \langle\text{подпредл}\rangle\,\langle\text{знак}\rangle\,\langle\text{подпредл}\rangle \\ \langle\text{знак}\rangle & \to & \texttt{,} \\ \langle\text{подпредл}\rangle & \to & \langle\text{глаг}\rangle\,\langle\text{сущ}\rangle \\ \langle\text{сущ}\rangle & \to & \texttt{отлично} \mid \texttt{неудовл.} \\ \langle\text{глаг}\rangle & \to & \texttt{нельзя} \mid \texttt{поставить} \end{array}$$

Осуществите их разбор.

Упр. 2.4. Постройте синтаксические деревья для следующих арифметических выражений:

a) `2 + a * 7`
c) `a + 3*b - b/8`
b) `2 a 7 * +`
d) `a 3 b * + b 8 / -`

Упр. 2.5. Предложить общие грамматические правила для представления всех арифметических выражений, подобных приведенному в упр. 2.4. Лексемами считать: `VAR`, `CONST`, `+`, `-`, `*`, `/`, `(`, `)`. Осуществить построение выражений из упр. 2.4 с помощью полученных грамматических правил.

Упр. 2.6 (А. Ю. Молчанов [9])**.** Поездом называется произвольная последовательность локомотивов и вагонов, начинающаяся с локомотива. Предложите грамматические правила для синтаксической группы ⟨поезд⟩, считая, что ⟨локомотив⟩ и ⟨вагон⟩ являются группами нижнего уровня. Модернизируйте правила:

a) все локомотивы должны быть сосредоточены в начале поезда;
b) поезд начинается с локомотива и заканчивается локомотивом;
c) поезд не содержит два локомотива или два вагона подряд.

Упр. 2.7 (Т. Пратт, М. Зелковиц [11]). Рассмотрим БНФ:

$\langle$pop$\rangle \to$ [$\langle$bop$\rangle$, $\langle$pop$\rangle$] | $\langle$bop$\rangle$
$\langle$bop$\rangle \to \langle$boop$\rangle$ | ($\langle$pop$\rangle$) $\langle$boop$\rangle \to$ `x` | `y` | `z`

Для каждой из перечисленных ниже строк укажите все синтаксические группы, к которым они относятся (если такие существуют):

a) `c`   c) `[y]`   e) `[(x),y]`
b) `(x)`   d) `([x,y])`   f) `[(x),[y,x]]`

Упр. 2.8 (Т. Пратт, М. Зелковиц [11]). Синтаксис *обезьяньего* языка очень прост, но только обезьяны умеют говорить на нем, не делая ошибок. Алфавит состоит из символов $\{\mathtt{a}, \mathtt{b}, \mathtt{d}, \#\}$, причем # используется в качестве пробела. БНФ заданы следующим образом:

$\langle$остановка$\rangle \to$ `b` | `d`
$\langle$звук$\rangle \to \langle$остановка$\rangle$ `a`
$\langle$слог$\rangle \to \langle$звук$\rangle$ | $\langle$звук$\rangle$ $\langle$остановка$\rangle$ | `a` $\langle$звук$\rangle$ | `a` $\langle$остановка$\rangle$
$\langle$слово$\rangle \to \langle$слог$\rangle$ | $\langle$слог$\rangle$ $\langle$слово$\rangle$ $\langle$слог$\rangle$
$\langle$предложение$\rangle \to \langle$слово$\rangle$ | $\langle$предложение$\rangle$ # $\langle$слово$\rangle$

Какой из перечисленных ниже ораторов — секретный агент, замаскированный под обезьяну?

a) Горилла: `ba#ababadada#bad#dabbada`
b) Шимпанзе: `abdabaadab#ada`
c) Бабуин: `dad#ad#abaadad#badadbaad`

Упр. 2.9. Опишите синтаксис заданного ниже фрагмента программы, написанной на языке C.

a)
```
a[i][j] = 3.14 * a[j][i];
```
b)
```
k = 5;
for( i=0; i<5; i++ )
   j = k-i;
```
c)
```
i = j++ + (--a!=3);
```
d)
```
if( i > j )
   if( k == 0 )
      k = 1;
```

Упр. 2.10. Опишите синтаксис текстового представления следующих объектов:

a) Оператор цикла.
b) Логическое выражение.
c) Прототип функции языка C++.
d) Числовая вещественная константа языка C.
e) Комментарий языка C++.
f) Телефонный справочник.
g) Справочник деканата по студентам.
h) Гипертекстовый ресурс.

§ 3. Стадии трансляции

Транслятор — это программа, которая выполняет анализ исходного текста (структурированного) и генерирует выходной текст. Для трансляторов языков программирования исходным текстом является программа на языке программирования. Если транслятор является *компилятором*, то выходным текстом выступает объектная программа на языке машинных кодов или ассемблера. Если транслятор — *интерпретатор*, то явного выходного текста не порождается, его «виртуальным» аналогом выступает последовательность выполненных действий.

Процесс трансляции разбивают на несколько стадий:

1. Лексический анализ (сканирование). Выделение в тексте лексем и отсев лишней с точки зрения трансляции информации.
2. Синтаксический анализ. Определение синтаксической структуры текста, как правило, в виде синтаксического дерева.
3. Семантический анализ. Выявление смысла исходного текста и его внутреннее представление.

Эти три стадии образуют *этап анализа*. Выделяют также *этап синтеза* (генерация выходного текста). Например, для языков программирования этап синтеза может включать подготовку к генерации кода, собственно генерацию кода, оптимизацию кода и т. п. Часто именно этап анализа является наиболее сложной частью транслятора, а этап синтеза сводится к простому преобразованию внутреннего представления смысла исходного текста в выходной текст.

Для взаимодействия вышеперечисленных стадий используются информационные таблицы, в которых накапливается извлеченная на каждой стадии информация об исходном тексте. Например, таблицы имен переменных, функций, заголовков for-циклов и т. п. Эти таблицы являются частью внутреннего представления структуры и смысла исходного текста.

Упр. 3.1. Предложить схему транслятора арифметических выражений в псевдомашинный код (выражение состоит из констант и переменных, знаков операций `+`, `-`, `*`, `/`, `^`, порядок операций — обычный и может быть изменен при помощи скобок). Описать работу каждой стадии транслятора на примерах:

a) `pi * Height^2 * (Radius - Height/3)`
b) `Sum/2 + (percent/100.0 - 0.13)*PrivateDebet`
c) `(Volume1 - Volume2)^3  +  1/2 * PI^2 * Volume3^3`
d) `2*A1*A3 + 1/2*PI/A2`   e) `(a+b)*3 - (a+b)`

Предложите варианты оптимизации генерируемого кода.

Упр. 3.2. Предложите схему транслятора арифметических выражений в псевдомашинный код. Выражения записываются в:

a) прямой польской записи: $\langle\textsf{операция}\rangle$ $\langle\textsf{операнд}_1\rangle$ $\langle\textsf{операнд}_2\rangle$, описать работу транслятора на примере выражения
$$- * + \mathtt{a\ b\ 3} + \mathtt{a\ b}$$

b) обратной польской записи: $\langle\textsf{операнд}_1\rangle$ $\langle\textsf{операнд}_2\rangle$ $\langle\textsf{операция}\rangle$, описать работу транслятора на примере выражения
$$\mathtt{a\ b} + 3 * \mathtt{a\ b} + -$$

Упр. 3.3. Аналогично упр. 3.1 и 3.2, но при трансляции арифметических выражений происходит *не генерация* программы вычисления выражения, *а перевод* из инфиксной в: a) прямую и b) обратную польскую запись.

Упр. 3.4. Предложить схему транслятора логических выражений (в стиле языка C) в псевдомашинный код. Опишите работу транслятора на конкретном примере. Предложите варианты оптимизации генерируемого кода.

Упр. 3.5. Предложите схему транслятора цикла некоторого языка программирования в псевдомашинный код. Опишите работу транслятора на конкретном примере.

a) Цикл «для»;

b) цикл «пока»;

c) цикл «до»;

d) цикл с возможностями `break` и `continue`, как в языке C.

Упр. 3.6. Предложите схему трансляции макроопределений, используемых в языке C и аналогичных следующему:

```
. . .
#define dsum(a,b) (2*((a)+(b)))
. . .
c = dsum(i++,2*i+j);
. . .
```

Разберите конкретные примеры. Укажите основные проблемы трансляции макроопределений.

Упр. 3.7. Предложите схему трансляции директив включения файлов, аналогичных используемым в языке C:

```
#include <stdio.h>
#include "consts.h"
```


Разберите конкретные примеры. Укажите основные проблемы трансляции директив включения.

Упр. 3.8. Приведите пример задачи на обработку структурированного текста. В чем заключается структура? Опишите этапы лексического, синтаксического и семантического анализа для этого примера.

Упр. 3.9. Рассмотрите следующие типы информационных таблиц:

1) имена переменных;
2) имена констант;
3) имена функций;
4) имена меток;
5) имена макроопределений;
6) имена счетчиков вложенных for-циклов;
7) имена вложенных файлов.

Предложите и обоснуйте способы решения следующих задач:

a) доступ к элементу таблицы (линейный, бинарный, хеширование и др.);

b) область видимости (локальные и глобальные имена);

c) использование объекта до его определения в программе;

d) ограничения на длину имен;

e) иерархическое пространство имен (`ancestor.parent.sun.descendant`);

f) объединение некоторых из представленных выше информационных таблиц в одну.

Упр. 3.10. Предложите схему распределения памяти ЭВМ под переменные и константы при трансляции объявлений переменных некоторого языка программирования. Как учитывается в схеме область видимости переменных? Опишите работу транслятора на конкретном примере.

Упр. 3.11. Предложите алгоритм распределения a) регистров и b) оперативной памяти для хранения промежуточных результатов вычисления математических выражений. Приведите конкретные примеры работы алгоритма.

Упр. 3.12. Предложите схему работы с динамической памятью в процессе выполнения программы. Рассмотрите задачи выделения (`malloc()`), возврата (`free()`) и перераспределения (`realloc()`), используемые, в частности, в языке C.

Упр. 3.13. Предложите схему работы с массивами произвольной размерности: a) размещение в памяти, b) доступ к элементам по индексам, c) набор операций.

Упр. 3.14. Предложите схему управления подпрограммами: a) вызов, b) передача параметров, c) вложенные вызовы и рекурсия, d) реентерабельность.

§ 4. Цепочки и формальные языки

Конечное множество символов называется *алфавитом*. *Цепочка* в алфавите $\Sigma = \{a_1, \dots, a_t\}$ есть любая конечная последовательность символов этого алфавита: $\alpha = a_{i_1} a_{i_2} \cdots a_{i_n}$, при этом $n = |\alpha|$ — *длина* цепочки. Пустая цепочка обозначается ε, при этом $|\varepsilon| = 0$. Множество всех цепочек в алфавите Σ обозначается как Σ^*. Множество всех цепочек без ε — как Σ^+.

Пусть $\alpha = a_1 \cdots a_n$ и $\beta = b_1 \cdots b_m$ — цепочки. Их *конкатенацией* называется цепочка $\alpha \cdot \beta = \alpha\beta = a_1 \cdots a_n b_1 \cdots b_m$, *обращением* α — цепочка $\alpha^R = a_n \cdots a_1$, префиксом α — цепочка вида $a_1 \dots a_k$ для $0 \leqslant k \leqslant n$, суффиксом α — цепочка вида $a_k \dots a_n$ для $1 \leqslant k \leqslant n+1$, подцепочкой α — цепочка вида $a_k \dots a_l$ для $1 \leqslant k \leqslant l \leqslant n$ или ε.

Произвольное множество цепочек называется *формальным языком*. Пусть L, L_1 и L_2 — формальные языки. Тогда $L_1 \cdot L_2 = L_1 L_2 = \{\alpha\beta \mid \alpha \in L_1, \beta \in L_2\}$, $L^0 = \{\varepsilon\}$, $L^n = L \cdot L^{n-1}$. Языки $L^+ = \bigcup_{n=1}^{\infty} L^n$ и $L^* = \bigcup_{n=0}^{\infty} L^n$ называются *положительной* и *неотрицательной итерацией* языка L.

Упр. 4.1 (Д. Грис [7])**.** Пусть в алфавите $\Sigma = \{\mathtt{0}, \mathtt{1}, \mathtt{2}\}$ цепочки $x = \mathtt{01}$, $y = \mathtt{2}$, $z = \mathtt{011}$. Выпишите следующие цепочки: xy, yz, xyz, x^4, x^3y^2, $(xy)^2$, $(yxx)^3$. Определите их длины.

Упр. 4.2. Заданы цепочки $x = \mathtt{01}$, $y = \mathtt{10}$ и $z = \mathtt{0110}$ в двоичном алфавите $\Sigma = \{\mathtt{0}, \mathtt{1}\}$. Выпишите цепочки $(xy)^2z$, $zxzy$ и $xzyz$. Определите их длины.

Упр. 4.3. Заданы цепочки $x = \mathtt{10}$, $y = \mathtt{11}$ и $z = \mathtt{0101}$ в двоичном алфавите $\Sigma = \{\mathtt{0}, \mathtt{1}\}$. Выпишите цепочки $y(xy^3z)^2x$, $x^2(yz)^3$ и $xzzyz$. Определите их длины.

Упр. 4.4. Дать и сравнить итеративное и рекурсивное определение цепочки в алфавите Σ.

Упр. 4.5. Докажите или опровергните, что для любых цепочек α, β и $n \in \mathbb{Z}_+$ выполняется равенство:

a) $|\alpha\beta| = |\beta\alpha|$
b) $|\alpha\beta| = |\alpha||\beta|$
c) $|\alpha\beta| = |\alpha| + |\beta|$
d) $|\alpha^n| = |\alpha|^n$
e) $|\alpha^n| = n|\alpha|$
f) $|\alpha^R| = |\alpha|$
g) $(\alpha^n\beta^n)^R = (\beta^R\alpha^R)^n$
h) $(\alpha^n\beta^n)^R = (\beta^R)^n(\alpha^R)^n$

Упр. 4.6. Докажите, что операция конкатенации цепочек является ассоциативной: $\alpha \cdot (\beta \cdot \gamma) = (\alpha \cdot \beta) \cdot \gamma$ для любых цепочек α, β и γ.

Упр. 4.7. Докажите, что операция конкатенации цепочек не является коммутативной: вообще говоря, $\alpha\beta \neq \beta\alpha$.

Упр. 4.8 (А. В. Гладкий [6])**.** Доказать, что для произвольных цепочек α и β тогда и только тогда $\alpha\beta = \beta\alpha$, когда существует цепочка γ и числа $m, n \in \mathbb{Z}_+$ такие, что $\alpha = \gamma^m$, $\beta = \gamma^n$.

Упр. 4.9 (Д. Грис [7])**.** Пусть $\Sigma = \{\mathtt{0}, \mathtt{1}, \mathtt{2}\}$. Найдите 7 самых коротких цепочек в множестве: a) Σ^+, b) Σ^*.

Упр. 4.10. Покажите, что относительно операции конкатенации цепочек a) Σ^+ есть полугруппа, b) Σ^* — моноид.

Упр. 4.11. Выпишите все суффиксы и префиксы для цепочек:
a) `911`, `UNIX`, `ИМО`; b) `215`, `LaTeX`, `TCP`; c) `AlphA`, `beta`, `gaMMa`.

Упр. 4.12. Сколько существует: a) префиксов, b) суффиксов, c) подцепочек для цепочки длины n?

Упр. 4.13. Пусть $L_1 = \{\mathtt{a}, \mathtt{aa}, \mathtt{0a}\}$ и $L_2 = \{\mathtt{1}, \mathtt{aa}\}$.
Вычислите: a) $L_1 \cdot L_2$, b) $L_1^2 \cup L_2^0$.

Упр. 4.14. Пусть $L_1 = \{\mathtt{ba}, \mathtt{ab}, \mathtt{1baa}\}$ и $L_2 = \{\mathtt{0}, \mathtt{aba}\}$.
Вычислите: a) $L_1(L_1 \cup L_2)$, b) $L_1 \setminus L_2^2$.

Упр. 4.15. Вычислите L^2, L^3 и L^4, если
a) $L = \{\mathtt{aba}\}$;
b) $L = \{\mathtt{aba}, \mathtt{a}\}$;
c) $L = \{\mathtt{ab}, \mathtt{bab}\}$;
d) $L = \{\mathtt{aa}, \mathtt{b}\}$;
e) $L = \{\mathtt{aa}, \mathtt{aaa}\}$;
f) $L = \{\mathtt{aa}, \mathtt{aaaa}\}$;
g) $L = \{\mathtt{a}, \mathtt{b}, \mathtt{ab}\}$;
h) $L = \{\mathtt{0a}, \mathtt{11a}, \mathtt{010}\}$.

Упр. 4.16. Пусть $L_1 = \{\mathtt{a}, \mathtt{aba}, \mathtt{0a}\}$ и $L_2 = \{\mathtt{1}, \mathtt{baa}\}$. Вычислите:
a) L_1L_2; c) L_2L_1; b) $L_1^2 \cup L_2^0$; d) $L_1^0 \cup L_2 \cup L_2^3$.

Упр. 4.17. Пусть $L_1 = \{\mathtt{a}, \mathtt{aba}, \mathtt{1baa}\}$ и $L_2 = \{\mathtt{1}, \mathtt{baa}\}$. Вычислите:
a) $L_1(L_1 \cup L_2)$; c) $L_1(L_2 \cup L_1)$; b) $L_1 \setminus L_2^2$; d) $L_1 L_2^0 \setminus L_2 L_1^0$.

Упр. 4.18. Пусть $L_1 = \{\mathtt{a}\}$, $L_2 = \{\mathtt{aa}\}$, $L_3 = \{\mathtt{1}, \mathtt{11}\}$, $L_4 = \{\mathtt{aba}, \mathtt{a}\}$, $L_5 = \{\mathtt{aba}, \mathtt{a}, \mathtt{b}\}$, $L_6 = \{\mathtt{aa}, \mathtt{bbb}, \mathtt{ab}\}$. Выпишите 10 самых коротких цепочек множеств L_i^+ и L_i^*, $i = 1, \ldots, 6$.

Упр. 4.19. Пусть $L_1 = \varnothing$ и $L_2 = \{\varepsilon\}$. Отличаются ли эти языки?

Упр. 4.20. a) Может ли язык L^* или L^+ быть пустым? b) При каких условиях L^* и L^+ конечны?

Упр. 4.21. Докажите, что относительно операции конкатенации a) L^+ есть полугруппа, b) L^* — моноид для любого языка L.

Упр. 4.22 (Дж. Хопкрофт и др. [14])**.** Докажите или опровергните заданные равенства для произвольных формальных языков.

a) $L_1(L_2 \cup L_3) = L_1L_2 \cup L_1L_3 \qquad \forall\, L_1, L_2, L_3$;
b) $L_1 \cdot (L_2 \cdot L_3) = (L_1 \cdot L_2) \cdot L_3 \qquad \forall\, L_1, L_2, L_3$;
c) $L^n \cdot L^m = L^{n+m} \qquad \forall\, L$ и $\forall\, n, m \in \{0, 1, 2, \ldots\}$;
d) $L^* \setminus L = \{\varepsilon\} \cup LL^+ \qquad \forall\, L$;
e) $(L^+)^n = L^+$ и $(L^*)^n = L^* \qquad \forall\, L$ и $\forall\, n \in \{1, 2, \ldots\}$;
f) $(L^+)^n = (L^n)^+$ и $(L^*)^n = (L^n)^* \qquad \forall\, L$ и $\forall\, n \in \{1, 2, \ldots\}$;
g) $(L^+)^+ = L^+$ и $(L^*)^* = L^* \qquad \forall\, L$;
h) $L^+ = LL^* = L^*L \qquad \forall\, L$;
i) $L^+ \cup L^* = L^* \qquad \forall\, L$;
j) $(L_1^* L_2^*)^* = (L_1 \cup L_2)^* \qquad \forall\, L_1, L_2$;
k) $(L_1 \cup L_2)^* = L_1^* \cup L_2^* \qquad \forall\, L_1, L_2$;
l) $(L_1L_2 \cup L_1)^* L_1 = L_1(L_2L_1 \cup L_1)^* \qquad \forall\, L_1, L_2$;
m) $(L_1 \cup L_2)^* L_2 = (L_1^* L_2)^* \qquad \forall\, L_1, L_2$;
n) $L_2(L_1L_2 \cup L_2)^* L_1 = L_1 L_1^* L_2 (L_1 L_1^* L_2)^* \qquad \forall\, L_1, L_2$;

Упр. 4.23. Пусть $h : \Sigma_1 \to \Sigma_2^*$. Определим $h(a\alpha) = h(a)h(\alpha)$, $h^{-1}(\beta) = \{\alpha \in \Sigma_1 | h(\alpha) = \beta \in \Sigma_2\}$, $h(\varepsilon) = \varepsilon$.

a) Пусть $\Sigma_1 = \{\mathtt{0}, \mathtt{1}, \mathtt{2}\}$, $\Sigma_2 = \{\mathtt{a}, \mathtt{b}\}$, $h(\mathtt{0}) = \mathtt{a}$, $h(\mathtt{1}) = \mathtt{bb}$, $h(\mathtt{2}) = \varepsilon$.

Найти:
1) $h(\mathtt{11020})$; 4) $h^{-1}(\mathtt{bab})$;
2) $h(\mathtt{21201222})$; 5) $h(L)$, где $L = \{\mathtt{0}, \mathtt{1}, \mathtt{2}\}^*$;
3) $h^{-1}(\mathtt{abb})$; 6) $h^{-1}(L)$, где $L = \{\mathtt{a}^n(\mathtt{bb})^m\}$.

b) Докажите или опровергните:
1) $h^{-1}(h(L)) = L$; 2) $h(h^{-1}(L)) = L$.

Упр. 4.24 (W. Rytter [16])**.**
a) Пусть $f_1 = \mathtt{b}$, $f_2 = \mathtt{a}$, $f_{n+2} = f_{n+1}f_n$. Цепочки f_n называются словами Фибоначчи. Докажите, что $c(f_{n-1}fn) = f_nf_{n-1}$, где $c(\alpha)$ меняет местами последние два символа в цепочке α.
b) Пусть $L \subseteq \{\mathtt{ab}, \mathtt{ba}\}^*$. Докажите, что если $x \in L$, то $\mathtt{a}x\mathtt{a}$ и $\mathtt{b}x\mathtt{b}$ не принадлежат L.

§ 5. Бинарные отношения

Задано множество A. *Бинарным отношением* R над A называется любое подмножество декартова произведения $A \times A$. Факт, что два элемента $a, b \in A$ состоят в отношении R, обозначается либо как $(a, b) \in R$, либо как $a\ R\ b$.

Отношение R является:
(1) *рефлексивным*, если $a\ R\ a\ \ \forall a \in A$,
(2) *транзитивным*, если из $a\ R\ b$ и $b\ R\ c$ всегда следует, что $a\ R\ c$,
(3) *рефлексивно-транзитивным*, если оно рефлексивно и транзитивно.

Пусть R, R_1 и R_2 — произвольные бинарные отношения. Определим сумму бинарных отношений $R_1 + R_2$ как объединение множеств $R_1 \cup R_2$, произведение как $R_2 \cdot R_2 = \{(a, b) \mid \exists c : a\ R_1\ c$ и $c\ R_2\ b\}$, степень как $R^0 = \{(a, a) \mid a \in \Sigma\}$, $R^n = R \cdot R^{n-1}$, транзитивное и рефлексивно-транзитивное замыкания — соответственно как $R^+ = \sum_{i=1}^{\infty} R^i$ и $R^* = \sum_{i=0}^{\infty} R^i$.

Если бинарное отношение R определено на конечном множестве $A = \{a_1, \ldots, a_n\}$, то отношению можно сопоставить булеву $(n \times n)$-матрицу B, где $B_{ij} = 1 \Leftrightarrow a_i\ R\ a_j$ и $B_{ij} = 0$ иначе. Сложение булевых матриц $B = B' + B''$ определяется как $B_{ij} = 0$, если $B'_{ij} = B''_{ij} = 0$, иначе $B_{ij} = 1$ (булева сумма элементов). Умножение $B = B' \cdot B''$ — как $B_{ij} = \sum_{k=1}^{n} B'_{ik} B''_{kj}$, где суммирование также булево.

Булевы матрицы сводят вычисление транзитивного (рефлексивно-транзитивного) замыкания к нахождению $B^+ = \sum_{i=1}^{n} B^i$ ($B^0 = \sum_{i=0}^{n} B^i$). Более эффективный способ дает алгоритм Уоршалла (см. напр. [7, 13]).

Упр. 5.1. Являются ли следующие отношения рефлексивными? Как выглядит транзитивное и рефлексивно-транзитивное замыкание каждого из них?

a) Отношение НижеЧем на множестве людей.

b) Отношение КручеЧем на множестве новых русских.

c) Отношение $>$ на множестве действительных чисел.

d) Отношение R на множестве действительных чисел,
$a\,R\,b \overset{\text{def}}{\iff} a = -b.$

e) Отношение R на множестве действительных чисел,
$a\,R\,b \overset{\text{def}}{\iff} a + b \geqslant 0.$

f) Отношение R на множестве целых чисел,
$a\,R\,b \overset{\text{def}}{\iff} a \equiv b \pmod 7.$

g) Отношение R на множестве целых чисел,
$a\,R\,b \overset{\text{def}}{\iff}$ a делит b.

h) Отношение R на множестве чисел $\{7, 4, 3, 12\}$,
$a\,R\,b \overset{\text{def}}{\iff} \left[\begin{array}{l} a < b/2, \\ b \text{ делит } a. \end{array}\right.$

i) Отношение R на множестве чисел $\{1, 2, 3, 4, 5\}$,
$a\,R\,b \overset{\text{def}}{\iff} \left[\begin{array}{l} a + b > 4, \\ b + a \leqslant 3. \end{array}\right.$

j) Отношение R на множестве чисел $\{1, 2, 3, 4, 5, 6\}$,
$a\,R\,b \overset{\text{def}}{\iff} \left[\begin{array}{l} a + b \text{ — четное}, \\ a + b < 5. \end{array}\right.$

k) Отношение R на множестве чисел $\{-2, -1, 0, 1, 2, 9\}$,
$a\,R\,b \overset{\text{def}}{\iff} ab > 0.$

l) Отношение R на множестве целых чисел,
$a\,R\,b \overset{\text{def}}{\iff} 2a + b = 0.$

m) Отношение R на множестве цепочек Σ^* (отношение префикса),
$\alpha\,R\,\beta \overset{\text{def}}{\iff} \beta = \alpha\gamma.$

n) Отношение R на множестве цепочек Σ^* (отношение суффикса),
$\alpha\,R\,\beta \overset{\text{def}}{\iff} \beta = \gamma\alpha.$

o) Отношение R на множестве цепочек Σ^* (отношение подцепочки),
$\alpha\,R\,\beta \overset{\text{def}}{\iff} \beta = \gamma'\alpha\gamma''.$

p) Отношение R на множестве цепочек Σ^* (отношение обращения),
$\alpha\,R\,\beta \overset{\text{def}}{\iff} \begin{array}{l} \alpha = a_1 \ldots a_n, \\ \beta = a_n \ldots a_1. \end{array}$

Упр. 5.2. Задано отношение R. Найдите отношение R^+, используя 1) графовую интерпретацию бинарных отношений; 2) сумму степеней булевых матриц; 3) алгоритм Уоршалла.

a) $R = \{(a, a), (b, a), (d, c), (a, d), (c, c)\}$;

b) $R = \{(a, b), (b, c), (c, a), (a, a), (d, e), (e, e), (e, f)\}$;

c) $R = \{(b, a), (c, c), (e, a), (f, a), (d, a), (e, e), (a, f)\}$;

d) $R = \{(a, b), (b, a), (c, c), (b, d), (c, b), (c, a), (e, c)\}$;

e) $R = \{(d, b), (e, a), (c, a), (b, d), (c, b), (d, a), (c, e), (a, d)\}$;

f) $R = \{(a, b), (b, d), (c, b), (c, a), (d, c), (d, a), (d, d), (b, b)\}$.

Упр. 5.3 (А. Ахо, Дж. Ульман [1]). Приведите пример отношения, которое: a) рефлексивно, но не транзитивно; b) транзитивно, но не рефлексивно.

Упр. 5.4 (Дж. Рейуорд-Смит [12]). Пусть R_1 и R_2 — бинарные отношения. Докажите, что:

a) если R_1 рефлексивно, то $R_1 \cup R_2$ тоже рефлексивно;

b) если R_1 и R_2 рефлексивны, то $R_1 \cap R_2$ также рефлексивно;

c) если R_1 и R_2 рефлексивны, то $R_1 \cdot R_2$ также рефлексивно.

Упр. 5.5. Докажите, что:

a) транзитивное замыкание R^+ является транзитивным для любого отношения R;

b) рефлексивно-транзитивное замыкание R^* является рефлексивным и транзитивным для любого отношения R;

c) отношение R^+ является наименьшим транзитивным отношением, включающим R;

d) отношение R^* является наименьшим рефлексивно-транзитивным отношением, включающим R;

e) $R = R^+$ тогда и только тогда, когда R транзитивно;

f) $R = R^*$ тогда и только тогда, когда R рефлексивно и транзитивно.

Упр. 5.6. Пусть R — бинарное отношение на множестве A. Доказать, что R рефлексивно тогда и только тогда, когда рефлексивно отношение $R \cap R^{-1}$. Здесь $R^{-1} = \{(a, b) \mid (b, a) \in R\}$.

Упр. 5.7. Докажите, что умножение бинарных отношений ассоциативно (т. е. $R \cdot P \cdot Q = R \cdot (P \cdot Q) = (R \cdot P) \cdot Q$), но не коммутативно (т. е., вообще говоря, $R \cdot P \neq P \cdot R$).

Упр. 5.8. Докажите, что для любых бинарных отношений R_1, R_2 и R_3 справедливо равенство $R_1 \cdot (R_2 \cup R_3) = R_1 \cdot R_2 \cup R_1 \cdot R_3$.

Упр. 5.9. Докажите, что равенство $R^n \cdot R^m = R^{n+m}$ справедливо для любого бинарного отношения R и любых неотрицательных целых n и m.

Упр. 5.10. Пусть в $|A| = n$ и R — произвольное отношение в A. Тогда если для $a, b \in A$ имеет место $a\ R^+\ b$, то существует положительное целое $k \leqslant n$ такое, что $a\ R^k\ b$. Докажите этот факт.

Упр. 5.11. Докажите, что для двух отношений, заданных в одном и том же множестве $A = \{a_1, a_2, \ldots, a_n\}$:

a) сумме отношений соответствует сумма их булевых матриц;

b) произведению отношений соответствует произведение их булевых матриц.

Упр. 5.12. Пусть B — булева матрица порядка, представляющая отношение R, заданное в A, где $|A| = n$. Докажите, что матрица $B^+ = B+B^2+B^3+\ldots+B^n$ соответствует транзитивному замыканию R^+ отношения R.

Упр. 5.13. Докажите правильность алгоритма Уоршалла, осуществляющего транзитивное замыкание любого отношения на конечном множестве из n элементов.

Вход: булева $(n \times n)$-матрица `B` для отношения R.

Выход: булева $(n \times n)$-матрица `B+` для отношения R^+.

```
B+ = B
For i = 1 to n
  For j = 1 to n
    If( B+[j,i] == 1)
      then
        For k = 1 to n
          B+[j,k] = B+[j,k] + B+[i,k]
        Next k
    EndIf
  Next j
Next i
```


§ 6. Формальные грамматики

Формальной *грамматикой* называется $G = (N, \Sigma, P, S)$, где $N = \{A_1, \ldots, A_n\}$ — нетерминальный алфавит (нетерминальные символы или нетерминалы), $\Sigma = \{a_1, \ldots, a_t\}$ — терминальный алфавит (терминальные символы или терминалы), $P = \{r_1, \ldots, r_m\}$ — правила грамматики, каждое из которых имеет вид $r = (\alpha \to \beta)$ для $\alpha \in (N \cup \Sigma)^* N (N \cup \Sigma)^*$ и $\beta \in (N \cup \Sigma)^*$, символ $S \in N$ — выделенный нетерминал, называемый *начальным*. Правила грамматики обобщают понятие БНФ, где синтаксическим переменным соответствуют нетерминалы, а синтаксическим константам — терминалы.

Выводом цепочки w в грамматике G называется последовательность цепочек, начинающаяся с начального нетерминала $S = \alpha_0$ и завершающаяся $w = \alpha_l$: $\alpha_0 \Rightarrow \alpha_1 \Rightarrow \alpha_2 \Rightarrow \ldots \Rightarrow \alpha_l$, где очередная цепочка α_i получается из предыдущей α_{i-1} с помощью некоторого правила $\alpha \to \beta$ грамматики (*непосредственный вывод* $\alpha_{i-1} \Rightarrow \alpha_i$): $\alpha_{i-1} = \alpha'_{i-1}\alpha\alpha''_{i-1} \Rightarrow \alpha'_{i-1}\beta\alpha''_{i-1} = \alpha_i$. Непосредственный вывод определяет отношение "$\Rightarrow$" на множестве $(N \cup \Sigma)^*$.

Цепочка, выводимая в грамматике, называется *сентенциальной формой*. *Предложением* называется сентенциальная форма, состоящая только из терминалов. Множество всех предложений определяет *язык* грамматики G, обозначаемый как $L[G]$. Грамматики называются *эквивалентными, если их языки совпадают*.

Выделяют 4 типа грамматик (иерархия Хомского) в зависимости от дополнительных ограничений на вид правил.

- Грамматика *типа 0* или *грамматика с фразовой структурой*. Нет ограничений на правила.
- Грамматики *типа 1* или *контекстно-зависимые грамматики* (КЗ-грамматики[1]). Правила имеют вид: $\alpha A \beta \to \alpha\gamma\beta$, где $A \in N$, $\alpha\beta \in (N \cup \Sigma)^*$, $\gamma \in (N \cup \Sigma)^+$.
- Грамматики *типа 2* или *контекстно-свободные грамматики* (КС-грамматики). Правила имеют вид: $A \to \alpha$, где $A \in N$, $\alpha \in (N \cup \Sigma)^*$.
- Грамматики *типа 3* или *регулярные грамматики* (Р-грамматики). Правила имеют вид[2]: $A \to a$ или $A \to aB$, где $A, B \in N$, $a \in \Sigma$.

Вывод цепочки в грамматике удобно представлять в виде *синтаксического графа*. Для КС-грамматик — это *синтаксическое дерево* (дерево разбора).

[1] Также их называют *грамматики непосредственных составляющих* [3].

[2] В данном случае используется определение праволинейной автоматной грамматики. Возможны и другие эквивалентные определения для Р-грамматик:
- праволинейная: $A \to xB$ или $A \to x$, где $A, B \in N$, $x \in \Sigma^*$;
- леволинейная: $A \to Bx$ или $A \to x$, где $A, B \in N$, $x \in \Sigma^*$;
- леволинейная автоматная: $A \to Ba$ или $A \to a$, где $A, B \in N$, $a \in \Sigma$.

В следующих упражнениях определите, какой тип иерархии Хомского имеет грамматика. Постройте несколько выводов и синтаксических графов для подтверждения правильности построенной грамматики.

Упр. 6.1 (Д. Грис [7]). В грамматике $G = (\{S\}, \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{0}, \mathtt{1}\}, P, S)$ множество правил P определено следующим образом: $S \to \mathtt{a} \mid \mathtt{b} \mid \mathtt{c} \mid S\mathtt{a} \mid S\mathtt{c} \mid S\mathtt{0} \mid S\mathtt{1}$. Выведите, когда это возможно, цепочки: a, ab0, a0c01, 0a, 11, aaa.

Упр. 6.2. Дана грамматика. Выведите, когда это возможно, заданную цепочку.

a) $x = \mathtt{abbb}$,
$S \to AB \mid BA$
$A \to AB \mid AA \mid \mathtt{a}$
$B \to \mathtt{b}$

b) $x = \mathtt{bbaa}$,
$S \to \mathtt{a}B \mid \mathtt{b}A$
$A \to \mathtt{a} \mid \mathtt{a}S \mid \mathtt{b}AA$
$B \to \mathtt{b} \mid \mathtt{b}S \mid \mathtt{a}BB$

c) $x = \mathtt{aabbbbbbbb}$,
$S \to \mathtt{aa}CFD, \quad F \to AFB \mid AB$
$AB \to \mathtt{b}BA, \quad A\mathtt{b} \to \mathtt{b}A$
$AD \to D, \quad C\mathtt{b} \to \mathtt{b}C$
$CB \to C, \quad \mathtt{b}CD \to \varepsilon$

d) $x = \mathtt{aaabbbccc}$,
$S \to \mathtt{a}SBC \mid \mathtt{ab}C$
$CB \to BC, \quad \mathtt{b}B \to \mathtt{bb}$
$\mathtt{b}C \to \mathtt{bc}, \quad \mathtt{c}C \to \mathtt{cc}$

Упр. 6.3. Построить все сентенциальные формы для грамматики:

a) $S \to A{+}B \mid B{+}A$
$A \to \mathtt{a}, \quad B \to \mathtt{b}$

b) $S \to A \mid BA$
$A \to A \mid \mathtt{1}, \quad B \to \mathtt{0} \mid \mathtt{00}$

c) $S \to APA \mid AP(APA)$
$P \to + \mid -, \quad A \to \mathtt{a} \mid \mathtt{b}$

d) $S \to \mathtt{1}A, \quad A \to \mathtt{0}B\mathtt{0} \mid C$
$B \to \mathtt{11} \mid \mathtt{1} \mid \varepsilon, \quad C \to A$

Упр. 6.4 (Д. Грис [7]). Предложите грамматику для языка:

a) десятичных записей неотрицательных целых чисел;
b) десятичных записей четных целых чисел;
c) десятичных записей четных целых чисел, не содержащих нуль в начале;
d) записей вещественных чисел с десятичной точкой;
e) десятичных записей вещественных чисел с плавающей точкой;
f) десятичных записей комплексных чисел с плавающей точкой.

Упр. 6.5. Предложите грамматику, описывающую все правильные идентификаторы произвольной длины некоторого языка программирования (состоят из букв латинского алфавита, цифр и знака подчеркивания, но не начинаются с цифры).

Упр. 6.6. Аналогично упр. 6.5, но длина идентификаторов ограничена тремя символами.

Упр. 6.7. Пусть язык L состоит из конечного числа цепочек. Предложите a) КС-грамматику, b) Р-грамматику для этого языка. Как выглядит грамматика для случая $L = \varnothing$?

Упр. 6.8. Предложите грамматику для языка:

a) $L = \{\, \mathtt{0}^n \mathtt{1}^m \mid n, m \in \mathbb{N}\}$
b) $L = \{\, \mathtt{0}^n \mathtt{1}^n \mid \in \mathbb{N}\}$
c) $L = \{\, \mathtt{0}^n \mathtt{1}^{n+m} \mathtt{2}^m \mid n, m \in \mathbb{N}\}$
d) $L = \{\, \mathtt{0}^n \mathtt{1}^m \mathtt{2}^k \mid n, m, k \in \mathbb{N}\}$
e) $L = \{\, \mathtt{0}^n \mathtt{1}^n \mathtt{2}^n \mid n \in \mathbb{N}\}$
f) $L = \{\, \mathtt{0}^n \mathtt{1}^n \mathtt{2}^m \mid n, m \in \mathbb{N}\}$
g) $L = \{\, \mathtt{0}^{n^2} \mid n \in \mathbb{N}\}$
h) $L = \{(\mathtt{0}^{2m} \mathtt{1}^m)^n \mid m \in \mathbb{N}, n \in \mathbb{Z}_+\}$
i) $L = \{\mathtt{0}^{3^n+1}! \mid n \in \mathbb{N}\}$
j) $L = \{\mathtt{0}^{n^3+1} \mid n \in \mathbb{N}\}$
k) $L = \{\, xx \mid x \in \{\mathtt{0}, \mathtt{1}\}^+\}$
l) $L = \{\, x!x \mid x \in \{\mathtt{0}, \mathtt{1}\}^+\}$
m) $L = \{\, xx^R \mid x \in \{\mathtt{0}, \mathtt{1}\}^+\}$
n) $L = \{\, x!x^R \mid x \in \{\mathtt{0}, \mathtt{1}\}^+\}$
o) $L = \left\{ x!y \;\middle|\; \begin{array}{l} x \neq y, \\ x, y \in \{\mathtt{0}, \mathtt{1}\}^+ \end{array} \right\}$
p) $L = \{x!y!z \mid x, y, z \in \{\mathtt{0}, \mathtt{1}\}^+\}$

Упр. 6.9 (W. Rytter [16]). a) Построить КС-грамматику, порождающую язык $\{x!y \mid x, y \in \{\mathtt{a}, \mathtt{b}\}^+, x \neq y\}$.

b) Построить КС-грамматику, имеющую всего один нетерминал и порождающую все цепочки над алфавитом $\{\mathtt{a}, \mathtt{b}\}$ с равным числом вхождений терминалов a и b в каждую цепочку.

Упр. 6.10. Предложите Р-грамматику, язык которой состоит из всех возможных цепочек из 0 и 1, в которых:

a) четное число 0 и 1;
b) нечетное число 0 и четное 1;
c) четное число 0 и нечетное 1;
d) нечетное число 0 и 1;
e) равное число 0 и 1;
f) неравное число 0 и 1;
g) между любыми двумя 0 находится хотя бы одна 1;
h) между любыми двумя 0 находится более одной 1.

Упр. 6.11. Предложите грамматику, порождающую те и только те цепочки из $\{0, 1, !\}^*$, в которых между любым символом 0 и ближайшим к нему символом 1 (слева или справа) находится непустая цепочка из символов !. Такими цепочками, например, являются 0!!1, !0!1, !!!, !1! и 1!!1!0. Осуществите разбор любых двух цепочек в полученной грамматике.

Упр. 6.12. Предложите грамматику, порождающую те и только те цепочки из $\{(,)\}^+$, которые являются правильной скобочной записью. Например, такими цепочками являются: (), ()(), (()), (()()). Примерами неправильных скобочных записей являются цепочки:)(, (()(), ())(().

Упр. 6.13. Предложите грамматику, порождающую те и только те цепочки над алфавитом {0, 1, +}, удовлетворяющие условию: между 0 и ближайшей к нему 1 в цепочке должна быть непустая цепочка из +. Например: 0++1, +0+1, +++, +1+, 1++1+0.

Упр. 6.14. Предложите грамматику, которая порождает те и только те цепочки из $\{\,\mathtt{a}\,\{\mathtt{a},\mathtt{b}\}^+\,\mathtt{b}\,\}$, в которых после любого символа a следует либо bb, либо ba. Например, следующие цепочки допускаются: abb, ababb, ababbbbbababb, abbbabb, а такие цепочки не допускаются: aba, abbab, baba, bbabb.

Упр. 6.15. Предложите грамматику, которая порождает те и только те цепочки из символов 0 и 1, у которых в любом префиксе символов 0 не меньше, чем 1.

Упр. 6.16. Пусть правила грамматики G заданы как $A \to \mathtt{1}\,A \mid \mathtt{00}$. Докажите, что множество $\{\mathtt{00}, \mathtt{100}, \mathtt{1100}, \ldots\}$ совпадает с $L[G]$.

Упр. 6.17. Пусть грамматика $G[S]$ задана правилами $S \to \mathtt{b}A \mid \mathtt{cc}$, $A \to \mathtt{b}A \mid \mathtt{b}$. Докажите, что $L[G] = \{\mathtt{b}^n\mathtt{cc} \mid n = 0, 2, 3, 4, \ldots\}$.

Упр. 6.18 (И. А. Волкова и Т. В. Руденко [5])**.** Построить КС-грамматику, эквивалентную грамматике:

a) $S \to \mathtt{a}A\mathtt{b}$
$\mathtt{a}A \to \mathtt{aa}A\mathtt{b}, \quad A \to \varepsilon$

b) $S \to AB \mid ABS$
$AB \to BA, \quad BA \to AB$
$A \to \mathtt{a}, \quad B \to \mathtt{b}$

Упр. 6.19 (И. А. Волкова и Т. В. Руденко [5], А. В. Гладкий [6], Ф. Л. Бауэр и Г. Гооз [2])**.** Какой язык порождает грамматика?

a) $S \to \mathtt{a} \mid B\mathtt{a}, \quad B \to B\mathtt{a} \mid \mathtt{b}$

b) $S \to A\mathtt{b}, \quad A \to A\mathtt{a} \mid \mathtt{ba}$

c) $S \to \mathtt{a}SB\mathtt{a}, \quad S \to \mathtt{aba}$
$\mathtt{a}B \to B\mathtt{a}, \quad \mathtt{b}B \to \mathtt{ba}$

d) $S \to \mathtt{0}A\mathtt{1} \mid \mathtt{01}$
$\mathtt{0}A \to \mathtt{00}A\mathtt{1}, \quad A \to \mathtt{01}$

e) $S \to AB, \quad AB \to BA$
$A \to \mathtt{a}, \quad B \to \mathtt{b}$

f) $S \to SS \mid A, \quad A \to \mathtt{a} \mid \mathtt{bb}$

g) $S \to A \mid B, \quad A \to \mathtt{a}A\mathtt{b} \mid \mathtt{0}$
$B \to \mathtt{a}B\mathtt{bb} \mid \mathtt{1}$

h) $S \to \mathtt{a}S\mathtt{b} \mid \mathtt{a} \mid \mathtt{a}A\mathtt{b} \mid \mathtt{ab}$
$A \to \mathtt{a}A\mathtt{b} \mid \mathtt{ab}$

i) $S \to BAB, \quad BA \to BC$
$CA \to AAC, \quad CB \to AAB$
$B \to \varepsilon, \quad A \to \mathtt{1}$

j) $S \to \mathtt{0}SA \mid \mathtt{0}B, \quad \mathtt{0}A \to A\mathtt{0}$
$BA \to \mathtt{1}B\mathtt{0}, \quad B \to \mathtt{10}$

k) $S \to \mathtt{0}S \mid S\mathtt{0} \mid D$
$D \to DD \mid \mathtt{1}A \mid \varepsilon$
$A \to \mathtt{0}B \mid \varepsilon, \quad B \to \mathtt{0}A \mid \mathtt{0}$

l) $S \to \mathtt{0}A \mid \mathtt{1}S \mid \varepsilon$
$A \to \mathtt{1}A \mid \mathtt{0}B, \quad B \to \mathtt{0}S \mid \mathtt{1}B$

m) $S \to \mathtt{a}BA \mid \varepsilon$
$B \to \mathtt{b}SA, \quad AA \to \mathtt{c}$

n) $S \to A\mathtt{b} \mid \mathtt{c}$
$A \to B\mathtt{a}, \quad B \to \mathtt{c}S$

Упр. 6.20. Восстановите грамматику по следующим двум выводам и постройте для них синтаксические деревья:

a) $S \Rightarrow A \Rightarrow AD \Rightarrow A\mathtt{3} \Rightarrow AD\mathtt{3} \Rightarrow A\mathtt{23} \Rightarrow D\mathtt{23} \Rightarrow \mathtt{123}$

b) $S \Rightarrow A \Rightarrow AD \Rightarrow D\mathtt{3} \Rightarrow AD\mathtt{3} \Rightarrow DD\mathtt{3} \Rightarrow \mathtt{1}D\mathtt{3} \Rightarrow \mathtt{123}$

Упр. 6.21. Рассмотрим грамматику $G[S]$:

$$\begin{array}{lcl} S & \to & A \mid B \\ A & \to & B\,\mathtt{0} \mid B\,\mathtt{1} \mid \mathtt{0} \mid \mathtt{1} \\ B & \to & A\,\mathtt{+} \mid B\,\mathtt{+} \mid \mathtt{+} \end{array}$$

Найти для любого $n \in \mathbb{N}$ величину $\alpha_n = |\{x \mid x \in L[G], |x| = n\}|$.

Упр. 6.22. Для языков, рассмотренных в упражнениях 6.10–6.14, определите для любого $n \in \mathbb{Z}_+$ сколько существует цепочек длины n, принадлежащих соответствующему языку.

Упр. 6.23 (W. Rytter [16])**.** Пусть в грамматике G все правила имеют вид $xA \to yB$ или $A \to \varepsilon$, где x и y — цепочки из терминалов, A, B, C — нетерминалы. Докажите, что G эквивалентна некоторой Р-грамматике.

§7. Конечные автоматы, регулярные языки и сканеры

Недетерминированный конечный автомат (НКА) есть $A = (Q, \Sigma, M, q_0, F)$, где Q — алфавит *состояний*, Σ — алфавит *входных символов*, $q_0 \in Q$ — *начальное состояние*, $F \subseteq Q$ — множество *заключительных* (допускающих, финальных) состояний, $M : Q \times \Sigma \to 2^Q$ — *функция переходов*. Детерминированный конечный автомат (ДКА) есть НКА с функцией переходов вида $M : Q \times \Sigma \to Q \cup \varnothing$, т. е. новое состояние при переходе определяется однозначно. Конечный автомат с ε-переходами (ε-КА) есть обобщение НКА с функцией переходов $M : Q \times (\Sigma \cup \varepsilon) \to 2^Q$, т. е. возможны переходы без обработки текущего входного символа.

Конфигурацией КА называется пара $(q, \alpha) \in Q \times \Sigma^*$. КА начинает работу из начальной конфигурации (q_0, x), где x — входная цепочка. Тактом НКА называется переход $(q, \mathtt{a}y) \vdash (p, y)$, где $p \in M(q, a)$. Для НКА $p = M(q, a)$. Для ε-КА возможно $(q, y) \vdash (p, y)$, если $p \in M(q, \varepsilon)$.

Такт $\vdash$ является бинарным отношением на множестве конфигураций. Транзитивное замыкание $\vdash^+$ позволяет обобщить функцию переходов до $M : Q \times \Sigma^+ \to 2^Q$ в случае НКА, до $M : Q \times \Sigma^+ \to Q \cup \varnothing$ в случае ДКА и до $M : Q \times \Sigma^* \to 2^Q$ в случае ε-КА. При этом $M(q, ay) = M(M(q, a), y)$ и $M(P, y) = \bigcup\limits_{p \in P} M(p, y)$.

КА *допускает* входную цепочку $x \in \Sigma^+$, если $M(q_0, x) \cap F \neq \varnothing$. Пустая цепочка допускается КА, если $q_0 \in F$. Для ДКА условие допустимости упрощается до $M(q_0, x) = p \in F$. В терминах тактов допустимость означает, что $\exists\, p \in F$ такое, что $(q_0, x) \vdash^* (p, \varepsilon)$.

Языком $L[A] \subset \Sigma^*$ конечного автомата A называется множество всех цепочек, допускаемых этим КА. Два КА называются *эквивалентными*, если их языки совпадают.

Известно, что: 1) для любого НКА и ε-КА возможно построить эквивалентный ДКА; 2) для любого КА можно построить эквивалентную Р-грамматику или регулярное выражение и наоборот; 3) для любого КА можно построить эквивалентный КА, но с минимально возможным числом состояний; 4) КА, Р-грамматики и регулярные выражения порождают один и тот же класс формальных языков — *регулярные языки*.

Упр. 7.1. ДКА задан таблицей переходов. 1) Постройте диаграмму переходов. 2) Опишите в виде Р-грамматики и регулярного выражения язык, который допускает данный КА.

a) S — начальное состояние, Z — заключительное,

M	0	1
S	S	Z
Z	Z	S

b) S — начальное состояние, S и Z — заключительные,

M	0	1
S	Z	S
Z	A	S
A	A	B
B	B	S

c) S_1, S_2 — начальные состояния, Z — заключительное,

M	a	b	c
S_1	$\varnothing$	S_2	A
S_2	Z	A	$\varnothing$
A	A	$\varnothing$	$\varnothing$
Z	Z	A	S_1

d) Все состояния — начальные, B — заключительное,

M	a	b	c	d
S	A	A	$\varnothing$	S
A	$\varnothing$	$\varnothing$	$\varnothing$	B
B	A	A	A	$\varnothing$

Упр. 7.2. Постройте КА, который допускает все указанные ниже цепочки и только их. Построен НКА, ДКА или ε-КА?

a) любое слово из латинских букв, начинающееся с `un` и заканчивающееся на `d`;

b) все цепочки над алфавитом $\Sigma = \{0, 1, 2\}$, включая и ε;

c) все цепочки из $\{\mathtt{a}, \mathtt{b}\}^*$, кроме начинающихся на `a`;

d) $\mathtt{a}^n \mathtt{b}^m$, где $n, m \in \mathbb{N}$;

e) $\mathtt{a}^n \mathtt{b}^m \mathtt{c}^k$, где $n, m, k \in \mathbb{N}$;

f) цепочки из `0` и `1`, префиксом которых является цепочка `1`, а суффиксом — цепочка `11`;

g) цепочки из `0` и `1`, в которых за каждой `1` непосредственно следует `0`;

h) непустые цепочки из десятичных цифр, начинающиеся с `1`, `3` или `9`, заканчивающиеся на `2`, `3` или `0` и обязательно содержащие хотя бы одну цифру `1`;

i) цепочки из $\{0, 1, \ldots, 9\}^+$, в которых последняя цифра встречается еще и где-то внутри самой цепочки;

j) цепочки из $\{0, 1, \ldots, 9\}^+$, в которых последняя цифра не встречается больше нигде в цепочке;

k) цепочки из десятичных цифр, начинающиеся с `1`, `3` или `9`, заканчивающиеся на `2`, `3` или `0` и содержащие хотя бы одну цифру `1`;

l) цепочки из десятичных цифр, начинающиеся с `1`, `3` или `4`, заканчивающиеся на `2`, `4` или `0` и содержащие не более двух вхождений `1`;

m) цепочки `booba`, `toll`, `today`, `toggle`, `toga`;

n) цепочки `вечер`, `вечно`, `вчера`, `сегодня`, `завтра`, `сейчас`;

o) цепочки из `0` и `1`, которые содержат подцепочку `01`;

p) цепочки из `0` и `1`, в которых число символов `1` делится на два, а число `0` — на три;

q) цепочки из `0` и `1`, которые начинаются на цепочку `01`, а число `0` в цепочке четное;

r) цепочки из $\{0,1\}^+$, в которых за каждым символом `a` где-то далее следует `b`;

s) цепочки из $\{0,1\}^+$, в которых любая подцепочка длины 4 содержит не более двух `0`.

t) цепочки из латинских букв, в которых на 3-й позиции справа стоит `z`.

Упр. 7.3. Построить эквивалентный ДКА по заданному НКА.

a) $A = (\{S, X, Z\}, \{\mathtt{a}, \mathtt{b}\}, M, \{S\}, \{Z\})$,

$$M(S,\mathtt{a}) = \{S, Z\},\quad M(X,\mathtt{a}) = \{S, X\},\quad M(Z,\mathtt{a}) = \{S, Z\},$$
$$M(S,\mathtt{b}) = \{S\},\quad M(X,\mathtt{b}) = \varnothing,\quad M(Z,\mathtt{b}) = \{S, X\}.$$

b) $A = \big(\{X, Y, Z\}, \{\mathtt{0}, \mathtt{1}\}, M, \{X\}, \{Z\}\big)$,

$$M(X,\mathtt{0}) = \{Z\},\quad M(Y,\mathtt{0}) = \{X, Y\},\quad M(Z,\mathtt{0}) = \{X, Z\},$$
$$M(X,\mathtt{1}) = \{X\},\quad M(Y,\mathtt{1}) = \varnothing,\quad M(Z,\mathtt{1}) = \{Y\}.$$

c) $A = (Q, \Sigma, M, S, F)$, где

$$\begin{array}{lll}
Q = \{S_1, S_2, A, F_1, F_2\}, & M(S_1, 0) = \{S_2, F_1\}, & M(S_1, 1) = \{A\}, \\
\Sigma = \{0, 1\}, & M(S_2, 0) = \{S_1, F_2, A\}, & M(S_2, 1) = \{A, S_2\}, \\
S = \{S_1, S_2\}, & M(A, 0) = \{A, F_1, F_2\}, & M(A, 1) = \{S_1\}, \\
F = \{F_1, F_2\}, & M(F_1, 0) = \varnothing, & M(F_1, 1) = \{F_2\}, \\
 & M(F_2, 0) = \{F_1\}, & M(F_2, 1) = \varnothing.
\end{array}$$

d) $A = (\{S, A, B, Z\}, \{0, 1\}, M, \{S\}, \{Z\})$,

M	0	1
S	$\{S, A\}$	$\{S\}$
A	$\{B\}$	$\{B, A\}$
B	$\{Z\}$	$\varnothing$
Z	$\{Z\}$	$\{Z\}$

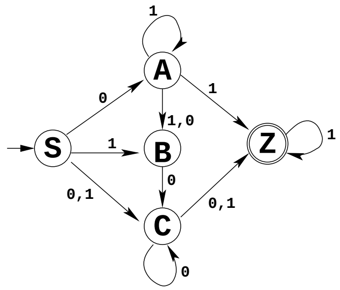
e) $A = (\{S, A, B, Z\}, \{\mathtt{0}, \mathtt{1}, \mathtt{3}\}, M, \{S, A\}, \{B, Z\})$,

M	0	1	3
S	$\{A, B\}$	$\varnothing$	S
A	$\varnothing$	A	$\{S, A, B\}$
B	A	$\{A, B\}$	$\varnothing$
Z	$\{S, B\}$	$\{A, Z\}$	$\varnothing$

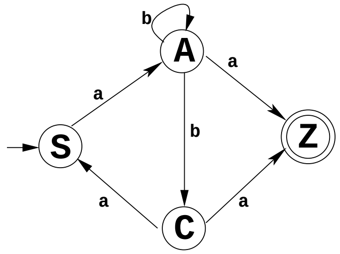
f) $A = (\{S_1, S_2, A, B, Z\}, \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}, M, \{S_1, S_2\}, \{B, Z\})$,

M	0	1	3
S_1	$\{A, Z\}$	$\varnothing$	$\{S_2, A, Z\}$
S_2	$\{A, S_1\}$	$\varnothing$	$\varnothing$
A	$\varnothing$	A	$\{S_2, A, Z\}$
B	A	$\{A, B\}$	$\varnothing$
Z	$\{S_2, A, B\}$	$\{S_1, Z\}$	B

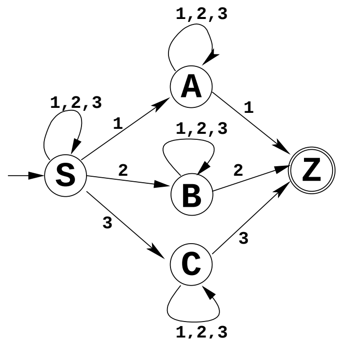
Упр. 7.4. Аналогично предыдущему упражнению, но НКА задается диаграммой переходов.



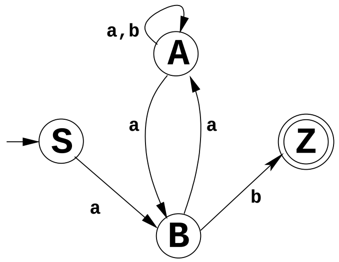
a)



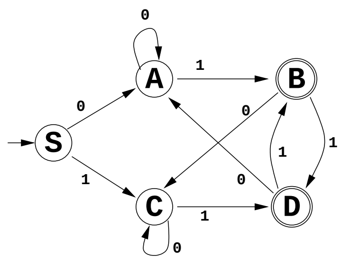
c)



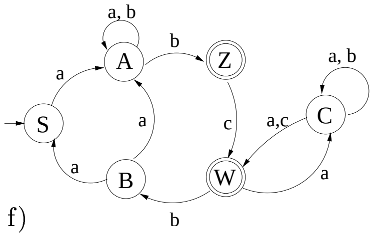
b)



d)



e)



f)

Упр. 7.5 (W. Rytter [16] (string-matching machine))**.** Для заданной цепочки w ненулевой длины постройте ДКА с не более чем $n+1$ состояниями, который допускает все цепочки, содержащие w как подцепочку, и только их.

Упр. 7.6. Рассмотрите КА, построенные в упр. 7.2. Постройте эквивалентные ДКА.

Упр. 7.7. Постройте ДКА, который допускает только идентификаторы (например, `A1`, `PI`, `Summa` и т. п.), целые десятичные константы (например, `10`, `-123`, `007` и т. п.) и целые двоичные константы (например, `10B`, `-000101B`, `1111110B` и т. п.).

Упр. 7.8. Докажите теорему о приведении любого НКА к эквивалентному ДКА.

Упр. 7.9. Пусть задана формальная грамматика. Постройте эквивалентные грамматике: 1) ДКА, 2) регулярное выражение.

a) $S \to \mathtt{0}A,\quad A \to \mathtt{0}A \mid \mathtt{1}S \mid \mathtt{0}$

b) $S \to \mathtt{b}A \mid \mathtt{b}B$
$A \to \mathtt{a}S \mid \mathtt{b}B \mid \mathtt{a}$
$B \to \mathtt{a}B \mid S \mid \mathtt{a}S \mid \mathtt{a}$

c) $S \to \mathtt{b}A \mid \mathtt{b}B \mid \mathtt{a}$
$A \to \mathtt{a}S \mid B \mid \mathtt{a}A$
$B \to \mathtt{a}B \mid \mathtt{a}S \mid \mathtt{b} \mid \mathtt{a}$

d) $S \to A \mid \mathtt{b}B \mid \mathtt{a}$
$A \to \mathtt{a}S \mid \mathtt{a}B \mid \mathtt{a}A$
$B \to \mathtt{a}B \mid \mathtt{b}S$

e) $S \to \mathtt{b}A \mid \mathtt{b}S \mid \mathtt{a}S$
$A \to \mathtt{a}S \mid \mathtt{b} \mid \mathtt{a}A$
$B \to \mathtt{a}A \mid \mathtt{a}S \mid \mathtt{b}B \mid \mathtt{a}$

f) $S \to \mathtt{a}S \mid \mathtt{a}A \mid \mathtt{b}$
$A \to \mathtt{a}S \mid \mathtt{b}B,\quad C \to \mathtt{ab}$
$B \to \mathtt{b}A \mid \mathtt{b}C \mid \mathtt{a}C$

g) $S \to S\mathtt{ab} \mid A\mathtt{a} \mid \mathtt{b},\quad A \to B\mathtt{b}$
$B \to A\mathtt{a} \mid C\mathtt{b} \mid C\mathtt{a},\quad C \to \mathtt{a}$

h) $S \to C\mathtt{a} \mid A\mathtt{a} \mid \mathtt{a},\quad A \to B\mathtt{b}$
$B \to A\mathtt{a} \mid C\mathtt{b} \mid B\mathtt{a},\quad C \to \mathtt{a} \mid C\mathtt{b}$

i) $S \to S\mathtt{0} \mid S\mathtt{1} \mid P\mathtt{0} \mid P\mathtt{1}$
$P \to N,\quad N \to \mathtt{0} \mid \mathtt{1} \mid N\mathtt{0} \mid N\mathtt{1}$

j) $S \to \mathtt{0}D \mid \mathtt{1}B$
$B \to \mathtt{0}C \mid \mathtt{1}C$
$C \to \mathtt{0}D \mid \mathtt{1}D \mid \varepsilon$
$D \to \mathtt{0}D \mid \mathtt{1}D$

k) $S \to \mathtt{0}C \mid \mathtt{1}B,\quad C \to \mathtt{0}C \mid \mathtt{1}C$
$B \to \mathtt{0}B \mid \mathtt{1}C \mid \varepsilon$

l) $S \to \mathtt{0}S \mid \mathtt{0}B$
$B \to \mathtt{0}B \mid \mathtt{1}C,\quad C \to \mathtt{1}C \mid \varepsilon$

m) $S \to \mathtt{a}A \mid \mathtt{a}B \mid \mathtt{b}A$
$A \to \mathtt{b}S,\quad B \to \mathtt{a}S \mid \mathtt{b}B \mid \varepsilon$

Упр. 7.10. Задано регулярное выражение. Постройте эквивалентные: 1) ДКА, 2) Р-грамматику.

a) $(\mathtt{a} \mid \mathtt{b}) \mid (\mathtt{a} \mid \mathtt{c})^* \mid \mathtt{dem}$

b) $(\,(\mathtt{0} \mid \mathtt{1})^* \mid (\mathtt{11})\,)^*$

c) $(\mathtt{a} \mid \mathtt{b})^+\mathtt{a}^2(\mathtt{ab})^* \mid \mathtt{1245}$

d) $\big(\mathtt{ab} \mid (\mathtt{0} \mid \mathtt{1})^*\big)^+ \mid \mathtt{a}\{0,3\}$

e) $\mathtt{a}^*((\mathtt{ba} \mid \mathtt{b})^*\mathtt{c})^+\mathtt{b}$

f) $(\mathtt{0} \mid \mathtt{1})^+ \mid (\mathtt{0} \mid \mathtt{1})^*.(\mathtt{0} \mid \mathtt{1})^+$

g) $(\mathtt{a} \mid \mathtt{ab})\mathtt{a}^* \mid \mathtt{b}^+$

h) $\big(\mathtt{ab} \mid (\mathtt{0} \mid \mathtt{1})^*\big)^+ \mid \big(\mathtt{a}\{0,3\} \mid \mathtt{b}^3\big)^+$

i) $(\mathtt{L}|\mathtt{l})\mathtt{ex}\big(\,(\mathtt{0}|\mathtt{1}|\mathtt{2}|\mathtt{3}|\mathtt{4}|\mathtt{5}|\mathtt{6}|\mathtt{7}|\mathtt{8}|\mathtt{9})^2 \mid (\mathtt{0}|\mathtt{1}|\mathtt{2}|\mathtt{3}|\mathtt{4}|\mathtt{5}|\mathtt{6}|\mathtt{7}|\mathtt{8}|\mathtt{9})^3\big)$

Упр. 7.11. Дана праволинейная грамматика. Постройте эквивалентные леволинейную и праволинейную автоматную грамматики.

a) $A \to \mathtt{011}A$
$A \to \mathtt{10}A \mid \mathtt{1}S \mid \mathtt{00}$

b) $S \to \mathtt{b}A \mid \mathtt{ab}B \mid \mathtt{abb}$
$A \to \mathtt{aa}S \mid B \mid \mathtt{a}A$
$B \to \mathtt{a}B \mid \mathtt{ab}S \mid \mathtt{b} \mid \mathtt{a}$

c) $S \to A \mid \mathtt{baa}B \mid \mathtt{aba}$
$A \to \mathtt{a}S \mid B \mid \mathtt{ab}A$
$B \to \mathtt{a}B \mid \mathtt{baa}S$

d) $S \to \mathtt{b}A \mid \mathtt{b}S$
$A \to \mathtt{aa}S \mid \mathtt{bb}A \mid \mathtt{aa}$

e) $S \to \mathtt{b}A \mid \mathtt{b}S \mid \mathtt{ab}S$
$A \to S \mid \mathtt{b} \mid \mathtt{a}A$
$B \to \mathtt{aa}A \mid \mathtt{a}S \mid \mathtt{bb}B \mid \mathtt{a}$

f) $S \to \mathtt{aa}S \mid \mathtt{a}A \mid \mathtt{aba}$
$A \to \mathtt{aaa}S \mid \mathtt{b}B,\quad C \to \mathtt{bbbb}$
$B \to \mathtt{b}A \mid \mathtt{ba}C \mid C$

Упр. 7.12. Докажите теорему об эквивалентности праволинейных, леволинейных и автоматных грамматик.

Упр. 7.13. Пусть $K_1 = (Q_1, \Sigma, M_1, S_1, Z_1)$ и $K_2 = (Q_2, \Sigma, M_2, S_2, Z_2)$ допускают языки L_1 и L_2 соответственно. Постройте КА для языка:

a) $L_1 \cup L_2$;
b) $L_1 \cap L_2$;
c) L_1^*;
d) L_1^+;
e) $L = \Sigma^* \setminus L_1$;
f) $L_1^* \setminus L_2^*$;
g) $L_1^+ \cup L_2^*$;
h) $L_1^* \cap L_2^+$;
i) $L_1^* \cap L_2$.

Упр. 7.14 (В. Дж. Рейуорд-Смит [12]). Пусть задан произвольный НКА $= (Q, \Sigma, M, S, Z)$, допускающий язык L. Введем множество

$$\mathrm{Pref}(L) = \{x \mid x = yz,\, z \in L,\, y \in \Sigma^*\}\ .$$

Постройте НКА, язык которого есть $\mathrm{Pref}(L)$.

Упр. 7.15 (лемма о разрастании для регулярных языков). Покажите, что для любого регулярного языка L существует константа n, такая, что любая цепочка $x \in L$, $|x| \geqslant n$ может быть представлена в виде $x = uvw$, где $|uv| \leqslant n$ и для любого $k \geqslant 0$ верно, что $uv^kw \in L$.

Упр. 7.16. Докажите, что следующие языки не являются регулярными (используйте лемму о разрастании):

a) $\{\mathtt{a}^n\mathtt{b}^n\}$;
b) $\{\mathtt{a}^p \mid p$ — простое$\}$;
c) $\{\mathtt{a}^{n^2}\}$;
d) $\{\mathtt{a}^{3^n}\}$;
e) $\{x \mid x$ — палиндром $(x = x^R)\}$;
f) скобочные выражения;
g) арифметические выражения;
h) логические выражения.

Упр. 7.17 (W. Rytter [16]). a) Докажите, что для любого формального языка L (не обязательно регулярного) над одноэлементным алфавитом язык L^* есть регулярный.

b) Докажите, что если L — регулярный язык, то язык $\{x \mid xx^R\}$ также регулярный.

c) Докажите, что любой контекстно-свободный язык над одноэлементным алфавитом является регулярным.

Упр. 7.18. Минимизируйте заданный КА.

a) $A = (\{S, A, B, Z\}, \{\mathtt{0}, \mathtt{1}, \mathtt{3}\}, M, \{S, A\}, \{B, Z\})$,

M	0	1	3
S	$\{A, B\}$	$\varnothing$	s
A	$\varnothing$	A	$\{S, A, B\}$
B	A	$\{A, B\}$	$\varnothing$
Z	$\{S, B\}$	$\{A, Z\}$	$\varnothing$

b) $A = (\{A, B, C, D, E, F, G\}, \{\mathtt{0}, \mathtt{1}\}, M, \{A\}, \{D, E\})$,

M	A	B	C	D	E	F	G
0	B	$\varnothing$	$\varnothing$	C	B	D	F
1	C	D	E	E	D	G	F

Упр. 7.19. Постройте в виде ДКА сканер, который:

a) исключает из текста программы комментарии языка C++;
b) добавляет бит четности/нечетности к цепочке из `0` и `1`;
c) распознает целочисленные константы языка C;
d) подсчитывает количество парных комбинаций `<>`, `<=` и `>=` ;
e) распознает лексемы простейшего языка программирования: 1) идентификаторы, 2) ключевые слова, 3) комментарии, 4) числовые и строковые константы;
f) подсчитывает в тексте программы: 1) число строк, содержащих операторы, 2) число операторов, 3) число идентификаторов, 4) долю комментариев в программе;
g) преобразует код Морзе (последовательность «точек», «тире» и «пауз») в обычный текст;
h) исключает из текста слова из заранее заданного набора;
i) аналогичен предыдущему сканеру, но слова заданного набора шифруются.

§ 8. Контекстно-свободные грамматики

Цепочка w называется *неоднозначной* в КС-грамматике, если для w существует более одного дерева разбора, в противном случае — *однозначной*. Если каждая цепочка имеет не более одного дерева разбора, то грамматика называется *однозначной*, в противном случае — *неоднозначной*. Язык, все порождающие КС-грамматики которого неоднозначны, есть *существенно неоднозначный*.

Нетерминал A называется *рекурсивным*, если $A \Rightarrow^+ \alpha A \beta$ для некоторых α и β. В случае $\alpha = \varepsilon$ ($\beta = \varepsilon$) нетерминал называется *леворекурсивным* (*праворекурсивным*). Если $\alpha\beta = \varepsilon$, то нетерминал называется *циклическим*.

Для КС-грамматики определяют следующие отношения на множестве всех символов (нетерминалы и терминалы) грамматики:

U FIRST V $\overset{\mathrm{def}}{\Longleftrightarrow}$ $\exists$ правило $U \to V\alpha$ для некоторой цепочки α;
U LAST V $\overset{\mathrm{def}}{\Longleftrightarrow}$ $\exists$ правило $U \to \alpha V$ для некоторой цепочки α;
U WITHIN V $\overset{\mathrm{def}}{\Longleftrightarrow}$ $\exists$ правило $U \to \alpha V \beta$ для некоторых цепочек α и β;
U SYMB V $\overset{\mathrm{def}}{\Longleftrightarrow}$ $\exists$ правило $U \to V$.

Нетерминал A называется *производящим*, если существует вывод $A \Rightarrow^+ x$ для некоторой терминальной цепочки x, в противном случае нетерминал — *бесплодный*. Символ U называется *выводимым*, если существует вывод $S \Rightarrow^* \alpha U \beta$, где S — начальный символ, α и β — произвольные цепочки, в противном случае символ — *недостижимый*. КС-грамматика называется *приведенной*, если она не содержит недостижимых и бесплодных символов. Для любой КС-грамматики может быть построена эквивалентная приведенная КС-грамматика.

Правила КС-грамматики называются *ε-правилами*, если они имеют вид $A \to \varepsilon$. КС-грамматика называется *грамматикой без ε-правил*, если она либо не содержит ε-правил, либо $\varepsilon \in L[G]$ и существует только одно ε-правило $S \to \varepsilon$, но при этом начальный нетерминал S не встречается в правой части ни одного правила. Для любой КС-грамматики может быть построена эквивалентная грамматика без ε-правил.

Циклом называется вывод вида $A \Rightarrow^+ A$ $(A \in N)$. Циклы возможны лишь при наличии *цепных правил* $A \to B$ $(A, B \in N)$. Для любой КС-грамматики может быть построена эквивалентная грамматика без цепных правил.

Упр. 8.1. Является ли предложение 000 однозначным в следующей КС-грамматике $S \to SS \mid 0$? Приведите примеры однозначных и неоднозначных предложений для этой грамматики.

Упр. 8.2. Покажите, что следующая КС-грамматика неоднозначна (неоднозначная грамматика арифметических выражений). Приведите примеры однозначных и неоднозначных предложений.

$$\begin{array}{lcl} E & \to & \mathtt{i} \mid (\,S\,) \mid E\,O\,E \\ O & \to & + \mid * \end{array}$$

Упр. 8.3. Покажите, что заданные предложения однозначны в КС-грамматике (однозначная грамматика арифметических выражений):

$$E \to T \mid E{+}T \mid E{-}T\ , \qquad T \to F \mid T{*}F \mid T/F\ , \qquad F \to (E) \mid \mathtt{a}$$

a) `a+a+a` c) `(a+a*a)-a/a` e) `(a+a)*(a-a)`
b) `a/a+a` d) `a+a*a-a` f) `a/a*(a+a+a)`

Упр. 8.4. Показать, что следующая КС-грамматика, определяющая синтаксис условного оператора, неоднозначна.

$$\begin{array}{lll} \langle\text{оператор}\rangle \to & \mathtt{IF}\ \text{усл}\ \mathtt{THEN}\ \langle\text{оператор}\rangle\ \mathtt{ELSE}\ \langle\text{оператор}\rangle \\ & \mid\ \mathtt{IF}\ \text{усл}\ \mathtt{THEN}\ \langle\text{оператор}\rangle\ \mid\ \text{дейст} \end{array}$$

Упр. 8.5. Задана КС-грамматика. Какой язык она порождает? Является ли она однозначной? Если нет, то предложите эквивалентную однозначную грамматику.

a) $S \to \mathtt{a}S \mid \mathtt{a}S\mathtt{b}S \mid \varepsilon$

b) $S \to SS \mid (S) \mid ()$

c) $S \to S\mathtt{b}S \mid \mathtt{b}S \mid \mathtt{a} \mid \mathtt{c}$

d) $S \to \mathtt{b}A \mid \mathtt{ba}B$
$A \to \mathtt{a}A \mid \mathtt{ab}, \quad B \to \mathtt{a}B \mid \mathtt{b}$

e) $S \to A, \quad A \to B$
$A \to \mathtt{1}, \quad B \to A$

f) $S \to S\mathtt{a}T \mid S\mathtt{b}T \mid T$
$T \to \mathtt{c}T \mid \mathtt{d}$

g) $S \to AB \mid \mathtt{a} \mid \mathtt{b} \mid \varepsilon$
$A \to \mathtt{a}A \mid \varepsilon, \quad B \to \mathtt{b}B \mid \varepsilon$

h) $S \to A\mathtt{1}B, \quad A \to \mathtt{0}A \mid \varepsilon$
$B \to \mathtt{0}B \mid \mathtt{1}B \mid \varepsilon$

i) $S \to A{+}A \mid A{+}B$
$A \to \mathtt{a}A \mid \mathtt{a}, \quad B \to A \mid \mathtt{a}$

Упр. 8.6 (Дж. Хопкрофт и др. [14]**).** Для КС-грамматик из упр. 8.5 некоторые предложения являются однозначными. Предложите алгоритм проверки, является ли заданное предложение однозначным.

Упр. 8.7. Предложите однозначную грамматику для порождения правильных `if–then–else` выражений (см. упр. 8.4).

Упр. 8.8 (И. А. Волкова и Т. В. Руденко [5]**).** Покажите, что наличие в КС-грамматике правил вида:

a) $A \to AA \mid \alpha$ *b)* $A \to A\alpha A \mid \beta$ *c)* $A \to \alpha A \mid A\beta \mid \gamma$

где $\alpha, \beta, \gamma \in (N \cup \Sigma)^*$, $A \in N$ делает ее неоднозначной.

Упр. 8.9 (S. Sippu и E. Soisalon-Soininen [17]**).** Покажите, что КС-грамматика без бесполезных символов неоднозначна, если она содержит нетерминал, который одновременно лево- и праворекурсивный.

Упр. 8.10. Найти отношения FIRST$^+$, LAST$^+$, WITHIN$^+$ и SYMB$^+$ для заданной КС-грамматики. Определите с их помощью, какие нетерминалы являются: 1) рекурсивными, 2) леворекурсивными, 3) праворекурсивными, 4) циклическими:

a) $S \to SA \mid \mathtt{a}A$
$A \to \mathtt{a}B \mid \mathtt{a}$
$B \to \mathtt{b}B \mid A$

b) $S \to \mathtt{a}SAS \mid BBB \mid \mathtt{aab}$
$A \to SS\mathtt{ab} \mid AA$
$B \to \mathtt{a}CAC \mid AC \mid CC$
$C \to A\mathtt{b}AB\mathtt{aaa}$

c) $$\begin{array}{l} S \to \mathtt{a}SS \mid AB \mid \mathtt{b} \\ A \to S\mathtt{a}, \quad C \to AS\mathtt{a}S \\ B \to CA \mid AC \mid CC \end{array}$$

d) $$\begin{array}{l} S \to \mathtt{b}SB \mid B \mid \mathtt{b} \\ A \to S\mathtt{a} \mid A, \quad B \to A\mathtt{a}A \\ C \to AC \mid A\mathtt{a}C \mid BA \end{array}$$

e) $$\begin{array}{l} S \to AB\mathtt{a}S \mid BA \mid \mathtt{b} \\ A \to A\mathtt{a} \mid SA, \quad C \to S\mathtt{a}B \mid \mathtt{aba} \\ B \to \mathtt{b}A \mid \mathtt{a}A\mathtt{c} \mid AB \end{array}$$

f) $$\begin{array}{l} S \to \mathtt{a}A \mid CCC \mid \mathtt{abbb} \\ A \to A\mathtt{ab}B \mid CS \\ B \to A \mid B \mid C, \quad C \to C\mathtt{a}S \mid S \end{array}$$

Упр. 8.11. Докажите, что КС-грамматика без рекурсивных нетерминалов порождает конечный язык.

Упр. 8.12. Пусть $G = (N, \Sigma, P, S)$ произвольная КС-грамматика, а множество $M = \{U \in N \mid (S, U) \notin \mathrm{WITHIN}^*\}$. Докажите, что грамматика $G'[S] = (N', \Sigma, P', S)$ порождает тот же язык, что и грамматика G, если $N' = N \setminus M$, а P' содержит только те правила из P, в левой и правой части которых нет нетерминалов из M.

Упр. 8.13. Пусть $G = (N, \Sigma, P, S)$ произвольная КС-грамматика, а множество $M = \big\{U \in N \mid (U, U) \in \mathrm{SYMB}^+\big\}$. Каким свойством обладают нетерминалы из M с точки зрения вывода? Можно ли изменить правила грамматики $G[S]$ так, чтобы новая грамматика порождала тот же язык, а отношение $\mathrm{SYMB}^+ = \varnothing$?

Упр. 8.14. Пусть $G = (N, \Sigma, P, S)$ произвольная КС-грамматика, а множество $M = \{U \in N \mid U \Rightarrow^+ x\,,\ x \in \Sigma^*\}$. Предложите, как можно использовать M для проверки языка грамматики на пустоту. Предложите алгоритм построения M.

Упр. 8.15. Постройте эквивалентную приведенную грамматику.

a) $$\begin{array}{l} S \to AB \mid CA, \quad A \to \mathtt{a} \\ B \to BC \mid AB, \quad C \to \mathtt{a}B \mid \mathtt{b} \end{array}$$

b) $$\begin{array}{l} S \to A \mid B, \quad A \to \mathtt{a}B \mid \mathtt{b}S \mid \mathtt{b} \\ B \to AB \mid B\mathtt{a}, \quad B \to AS \mid \mathtt{b} \\ C \to CD \mid \mathtt{c}, \quad D \to C \mid A \end{array}$$

c) $$\begin{array}{l} S \to \mathtt{a}ABS \mid \mathtt{b}CAC\mathtt{d} \\ A \to \mathtt{b}AB \mid \mathtt{c}SA \mid \mathtt{c}CC \\ B \to \mathtt{b}AB \mid \mathtt{c}SB, \quad C \to \mathtt{c}S \mid \mathtt{c} \end{array}$$

d) $$\begin{array}{l} S \to AB \mid D, \quad B \to ED \mid \mathtt{b} \\ A \to BD \mid EF \mid \mathtt{a}, \quad C \to AB \\ E \to B, \quad D \to F \end{array}$$

e) $$\begin{array}{l} S \to \mathtt{a}AB \mid E, \quad A \to \mathtt{a}A \mid \mathtt{b}B \\ B \to AC\mathtt{b} \mid \mathtt{b}, \quad D \to \mathtt{a} \mid \mathtt{c} \mid F\mathtt{b} \\ C \to A \mid \mathtt{b}A \mid \mathtt{c}C \mid \mathtt{a}E \\ E \to A \mid \mathtt{c}E \mid \mathtt{a}E \mid ED \mid FG \\ F \to BC \mid EC \mid AC, \quad G \to G\mathtt{a} \mid G\mathtt{b} \end{array}$$

f) $$\begin{array}{l} S \to \mathtt{b}ASA \mid A\,\mathtt{b}\,A \mid C \\ A \to \mathtt{a}S \mid \mathtt{ba} \mid \mathtt{aa}, \quad B \to \mathtt{a}B \mid A \mid \mathtt{a} \\ C \to \mathtt{a}CSS \mid SC \end{array}$$

Упр. 8.16. Постройте эквивалентную грамматику без ε-правил.

a) $$\begin{array}{l} S \to AB, \quad A \to \mathtt{a}AA \mid \varepsilon \\ B \to \mathtt{b}BB \mid \varepsilon \end{array}$$

b) $$\begin{array}{l} S \to A\mathtt{a}B \mid \mathtt{a}B \mid \mathtt{c}C \\ A \to AB \mid \mathtt{a} \mid \mathtt{c} \mid B \\ B \to AB \mid B\mathtt{a} \mid \varepsilon, \quad C \to AB \mid \mathtt{c} \end{array}$$

c) $$\begin{array}{l} S \to ABC, \quad A \to BB \mid \varepsilon \\ B \to CC \mid \mathtt{a}, \quad C \to AA \mid \mathtt{b} \end{array}$$

Упр. 8.17. Преобразуйте в грамматику без цепных правил:

a) $$\begin{array}{l} E \to E{+}T \mid E{-}T \mid T \\ T \to T{*}F \mid T/F \mid F, \\ F \to (E) \mid \mathtt{a} \mid \mathtt{b} \end{array}$$

b) $$\begin{array}{l} S \to AB, \quad A \to C, \quad A \to \mathtt{a} \\ B \to \mathtt{b}, \quad C \to A \end{array}$$

Упр. 8.18 (Дж. Хопкрофт и др. [14])**.** Проделайте с заданной грамматикой преобразования, удалив:

1) ε-правила;
2) цепные правила;
3) бесплодные символы;
4) недостижимые символы.

a) $$\begin{array}{l} S \to ASB \mid \varepsilon, \quad A \to \mathtt{a}AS \mid \mathtt{a} \\ B \to A\mathtt{b}S \mid A \mid \mathtt{bb} \end{array}$$

b) $$\begin{array}{l} S \to \mathtt{0}A\mathtt{0} \mid \mathtt{1}B\mathtt{1} \mid BB \\ A \to C, \quad B \to S \mid A \\ C \to S \mid \varepsilon \end{array}$$

c) $$\begin{array}{l} S \to AAA \mid B \\ A \to \mathtt{a}A \mid B, \quad B \to \varepsilon \end{array}$$

d) $$\begin{array}{l} S \to \mathtt{a}A\mathtt{a} \mid \mathtt{b}B\mathtt{b} \mid \varepsilon \\ A \to C \mid \mathtt{a}, \quad C \to CDE \mid \varepsilon \\ B \to C \mid \mathtt{b}, \quad D \to A \mid B \mid \mathtt{ab} \end{array}$$

e) $$\begin{array}{l} S \to E{+}T, \quad E \to E \mid Z{+}F \mid T \\ F \to F \mid FP \mid P, \quad P \to G \mid \varepsilon \\ G \to G \mid GG \mid F, \quad T \to T{*}\mathtt{a} \mid \mathtt{a} \\ Q \to E \mid E{+}F \mid T \mid Z, \quad Z \to \mathtt{a} \end{array}$$

f) $$\begin{array}{l} S \to A \mid B, \quad A \to C \mid D \\ B \to D \mid E, \quad C \to S \mid \mathtt{a} \mid \varepsilon \\ D \to S \mid \mathtt{b}, \quad E \to S \mid \mathtt{c} \mid \varepsilon \end{array}$$

g) $$\begin{array}{l} S \to \mathtt{a}AB \mid E, \quad A \to \mathtt{d}DA \mid \varepsilon \\ C \to \mathtt{c}AB \mid \mathtt{d}SD \mid \mathtt{a}, \quad D \to \mathtt{e}A \\ B \to \mathtt{b}E \mid \mathtt{f}, \quad E \to \mathtt{f}A \mid \mathtt{g} \end{array}$$

Упр. 8.19 (Дж. Хопкрофт и др. [14])**.** Пусть G — грамматика без ε-правил. Доказать, что если $x \in L[G]$, $|x| = n$ и x порождается за m шагов, то для x существует синтаксическое дерево с $n + m$ узлами.

Упр. 8.20. Пусть задана грамматика $G = (N, \Sigma, P, S)$, где множество правил P имеет вид: $P = \{U_i \to \alpha_i\}_{i=1}^{m}$ ($U_i \in N$, $\alpha_i \in (N \cup \Sigma)^*$). Что выполняет следующий алгоритм (из каких нетерминалов состоит множество V после окончания его работы)? Ответ обосновать.

```
V' := ∅;
repeat
   V := V';
   for i=1 to m do
      if( (αi = ε) или (αi = X1X2...Xn и все Xj ∈ V) )
         then
            V' := V ∪ {Ui};
      endif
   next i
until (V' = V);
```

§ 9. Задача грамматического разбора

Пусть задана КС-грамматика и текст. Задача грамматического разбора текста заключается в построении синтаксической структуры (дерево разбора) текста в соответствии с языком, порождаемым данной грамматикой. Если текст не принадлежит языку, то задача не имеет решения.

Грамматический разбор называется *нисходящим*, если синтаксическое дерево строится начиная с начального нетерминала (корня), постепенно спускаясь до уровня терминалов, составляющих текст (крона). Разбор называется *восходящим*, если синтаксическая структура строится начиная с исходной терминальной цепочки (крона), постепенно поднимаясь до начального нетерминала (корня).

В общем случае для КС-грамматик алгоритмы нисходящего и восходящего разбора требуют перебора вариантов (алгоритм с возвратами), что приводит к экспоненциальному от длины текста времени работы. Существуют табличные методы разбора, требующие полиномиального времени [1, 15]. При построении синтаксических анализаторов для трансляторов языков программирования, как правило, требуется линейная сложность грамматического разбора (детерминированный разбор). Построение таких алгоритмов возможно лишь для некоторых подклассов КС-грамматик.

Упр. 9.1. Осуществите восходящий и нисходящий разбор (с возвратами) заданной цепочки в заданной КС-грамматике.

a) `abbb`
$S \to AB \mid BA$
$A \to AB \mid AA \mid \mathtt{a}$
$B \to \mathtt{b}$

b) `bbaa`
$S \to \mathtt{a}B \mid \mathtt{b}A$
$A \to \mathtt{a} \mid \mathtt{a}S \mid \mathtt{b}AA$
$B \to \mathtt{b} \mid \mathtt{b}S \mid \mathtt{a}BB$

c) `(+(i(`
$S \to V_1, \quad V_1 \to V_2 \mid V_1\mathtt{i}V_2$
$V_2 \to V_3 \mid V_2\mathtt{+}V_3 \mid \mathtt{i}V_3$
$V_3 \to \mathtt{)}V_1\mathtt{)} \mid \mathtt{(}$

d) `561, 0, 0012`
$\langle\text{число}\rangle \to \langle\text{чс}\rangle$
$\langle\text{чс}\rangle \to \langle\text{чс}\rangle\,\langle\text{циф}\rangle \mid \langle\text{циф}\rangle$
$\langle\text{циф}\rangle \to \mathtt{0} \mid \mathtt{1} \mid \mathtt{2} \mid \ldots \mid \mathtt{8} \mid \mathtt{9}$

e) `10.1001`
$S \to S\mathtt{0} \mid S\mathtt{1} \mid D\mathtt{0} \mid D\mathtt{1}$
$D \to H.$
$H \to \mathtt{0} \mid \mathtt{1} \mid H\mathtt{0} \mid H\mathtt{1}$

f) `if a then b=a+b+b`
$S \to \mathtt{if}\ B\ \mathtt{then}\ S \;\mid\; B = E$
$E \to B \mid B\mathtt{+}E, \quad B \to \mathtt{a} \mid \mathtt{b}$

Упр. 9.2. Предложите алгоритм разбора любой цепочки в заданной КС-грамматике. Какая у него временна́я сложность?

a) $S \to \mathtt{a}ASB \mid \mathtt{c}$
$A \to \mathtt{aa}, \; B \to \mathtt{b} \mid B\mathtt{c}$

b) $S \to E, \; A \to \mathtt{a} \mid \mathtt{b}$
$E \to () \mid (E,E) \mid A$

c) $S \to \mathtt{aa}S \mid \mathtt{bc}$

d) $S \to \mathtt{a}S\mathtt{b} \mid !$

e) $S \to SAS \mid !A!$
$A \to \mathtt{101}$

f) $S \to (S) \mid SS \mid ()$

g) $S \to \mathtt{aa}A \mid \mathtt{b}S \mid \mathtt{c}$

h) $S \to AA!B \mid B!A$
$A \to \mathtt{0}S \mid \mathtt{0}, \; B \to S\mathtt{1} \mid \mathtt{1}$

Упр. 9.3. Предложите алгоритм, позволяющий по произвольным КС-грамматике и цепочке (входные данные) определить, выводима ли эта цепочка в заданной КС-грамматике или нет. Выполните пошагово алгоритм на нескольких примерах. Опишите его недостатки.

Упр. 9.4. Метод рекурсивного спуска реализует нисходящий разбор с возвратами. Для каждого нетерминала грамматики создается своя процедура, задача которой начиная с указанного места исходной цепочки найти подцепочку, выводимую из этого нетерминала. Опишите алгоритм рекурсивного спуска и выполните его на нескольких примерах. Для каких КС-грамматик этот алгоритм неприменим?

Упр. 9.5 (недетерминированный разбор на основе автомата с магазинной памятью)**.** Предложите алгоритм, использующий стек для выполнения a) нисходящего, b) восходящего разбора с возвратами. Выполните пошагово алгоритм на нескольких примерах. Опишите его недостатки.

§ 10. Темы для теоретических курсовых проектов

10.1. Математический аппарат

1.1. Алгоритм Флойда — Уоршалла построения транзитивного замыкания бинарного отношения на конечном множестве [7, гл. 2, п. 2.7], [13, гл. 7, пп. 7.2, 7.6].

1.2. Формальные языки как бесконечные множества [15, ch. 2, sect. 2.1].

10.2. Арифметические выражения

2.1. Перевод арифметических выражений из инфиксной записи в обратную польскую запись. Метод стека с приоритетами (алгоритм Э. Дейкстры) [7, гл. 11, 12].

2.2. Генерация автокода гипотетической ЭВМ по арифметическому выражению [1, т. 2, гл. 11, п. 2].

2.3. Оптимизация кода гипотетической ЭВМ для вычисления арифметического выражения [1, т. 2, гл. 11].

2.4. Стековый алгоритм вычисления арифметических выражений, представленных в обратной польской записи [10, гл. 1].

10.3. Регулярные языки и конечные автоматы

3.1. Минимизация конечных автоматов [1, т. 1, гл. 2, п. 2.3.1], [14, гл. 4, п. 4.4].

3.2. Лемма о разрастании для регулярных языков [1, т. 1, гл. 2, п. 2.3.2], [14, гл. 4, п. 4.1].

3.3. Уравнения над регулярными множествами [1, т. 1, гл. 2, п. 2.2.1].

3.4. Свойства замкнутости регулярных множеств [1, т. 1, гл. 2, п. 2.3.3], [14, гл. 4, п. 4.2].

3.5. ДКА для распознавания ключевых слов в тексте [14, гл. 2, п. 2.4.3].

10.4. Формальные грамматики

4.1. Нормальная форма Хомского для КС-грамматик [1, т. 1, гл. 2, п. 2.4.3], [14, гл. 7, пп. 7.1.5, 7.4.2], [15, ch. 4, sect. 4.2.3].

4.2. Нормальная форма Грейбах для КС-грамматик [1, т. 1, гл. 2, пп. 2.4.4, 2.4.5].

4.3. Лемма Огдена для КС-грамматик (лемма о разрастании) [1, т. 1, гл. 2, п. 2.6], [14, гл. 7, п. 7.2], [15, ch. 2, sect. 2.7.1].

4.4. Расширенные формы записи КС-грамматики [7, гл. 2, п. 2.9], [15, ch. 2, sect. 2.3.2.3].

4.5. Неукорачивающие грамматики и КЗ-грамматики: алгоритм распознавания принадлежности [3, гл. I, § 2].

4.6. Укорачивающие КС-грамматики и почти эквивалентные КС-грамматики [3, гл. I, § 3], [14, гл. 7, п. 7.1.3], [15, гл. 4, sect. 4.2.3.1].

4.7. VW-грамматики [3, гл. III], [15, гл. 2, sect. 2.4].

4.8. VW-грамматики (грамматики Вайнгаардена) [3, гл. III], [15, гл. 2, sect. 2.4].

4.9. Устранение левой рекурсии в КС-грамматиках [1, т. 1, гл. 2, п. 2.4.4], [15, ch. 6, sect. 6.4].

10.5. Методы грамматического разбора

5.1. Машины Тьюринга как распознающие устройства [3, гл. II, § 1], [14, гл. 8].

5.2. Грамматический разбор КС-языков методом рекурсивного спуска (недетерминированный нисходящий разбор) [7, гл. 4].

5.3. Варианты МП-автоматов [1, т. 1, гл. 2, п. 2.5.2], [3, гл. II, § 3, 4], [14, гл. 6].

5.4. Детерминированные МП-автоматы и КС-языки [1, т. 1, гл. 2, пп. 2.5.4, 2.6.4], [14, гл. 6, п. 6.4].

5.5. Грамматики предшествования операторов [7, гл. 6, пп. 7.2, 7.6], [1, т. 1, гл. 5, п. 5.4.3], [15, ch. 9, sect. 9.2].

5.6. Алгоритм Кока — Янгера — Касами [1, т. 1, гл. 4, п. 4.2.1], [15, ch. 4, sect. 4.2].

5.7. Алгоритм Эрли [1, т. 1, гл. 4, п. 4.2.2].

5.8. Глобальный анализатор. Алгоритм Унгера [3, гл. III, § 3], [15, ch. 4, sect. 4.1].

5.9. $LL(1)$-грамматики [1, т. 1, гл. 5, п. 5.1], [15, ch. 8, sect. 8.2].

10.6. Трансляция

6.1. Проектирование сканера некоторого языка программирования [7, гл. 3, пп. 3.3, 3.4].

6.2. Атрибутные грамматики. Семантические подпрограммы [7, гл. 7, 12, 13], [4, лекция 8].

6.3. Организация информационных таблиц транслятора [1, т. 2, гл. 10, п. 10.1], [7, гл. 9, 10], [4, лекция 9].

6.4. Проверка типов в трансляторе [4, лекция 10].

6.5. Организация памяти во время выполнения программы [7, гл. 8, 14], [4, лекция 11].

6.6. Оптимизация кода: линейные участки программы [1, т. 2, гл. 11, п. 11.1], [7, гл. 18, п. 18.1].

6.7. Оптимизация кода: циклы в программе [1, т. 2, гл. 11, п. 11.3], [7, гл. 18, пп. 18.2, 18.3].

6.8. Нейтрализация ошибок [7, гл. 15], [15, ch. 10].

10.7. Приложения методов трансляции

7.1. Алгоритмы поиска подстроки. Использование конечных автоматов и регулярных выражений [4, лекция 3], [14, гл. 2, п. 2.4; гл. 3, п. 3.3], [15, sect. 5.1.3, 5.3.4].

7.2. Приложение формальных грамматик к языку Пролог. Definite Clause Grammars [15, ch. 6, sect. 6.7].

7.3. Использование формальных грамматик для языков описания документов [14, гл. 5, пп. 5.3.3, 5.3.4], [15, sect. 5.1.3, 5.3.4].

10.8. Проблемы разрешимости в теории формальных языков

8.1. Проблема соответствий Поста [14, гл. 9, п. 9.4], [1, т. 1, гл. 0, п. 0.4.6].

8.2. Разрешимые проблемы для регулярных множеств: принадлежность, пустота и эквивалентность [14, гл. 4, п. 4.3], [1, т. 1, гл. 2, п. 2.3.4].

8.3. Разрешимые проблемы для КС-языков: принадлежность и пустота [14, гл. 7, п. 7.4], [1, т. 1, гл. 2, пп. 2.4.2, 2.6.3].

§ 11. Темы для программных курсовых проектов

11.1. Калькуляторы

1.1. Калькулятор математических выражений над вещественными числами.

1.2. Калькулятор математических выражений над комплексными числами.

1.3. Калькулятор арифметических выражений над кватернионами.

1.4. Калькулятор матричных выражений.

1.5. Калькулятор арифметических выражений над полиномами.

1.6. Калькулятор графов.

1.7. Калькулятор случайных величин.

1.8. Калькулятор перестановок.

1.9. Калькулятор геометрических фигур на плоскости.

1.10. Дифференциатор математических выражений.

11.2. Интерпретаторы

2.1. Интерпретатор черепашьей графики.

2.2. Интерпретатор языка ассемблер.

2.3. Интерпретатор языка BASIC.

2.4. Интерпретатор языка Pascal.

2.5. Интерпретатор языка C.

2.6. Интерпретатор языка SQL.

2.7. Эмуляция интерпретатора командной строки операционной системы (например, COMMAND.COM, UNIX-shell).

11.3. Конверторы

3.1. Конвертор программ с языка Pascal на язык C.

3.2. Конвертор программ с языка Pascal на язык C++.

3.3. Конвертор программ с языка BASIC на язык Pascal.

3.4. Конвертор программ с языка BASIC на язык C.

3.5. Конвертор формул из формата $\TeX$ в текстовый формат.

3.6. Конвертор HTML-документов в текстовый формат.

3.7. Транслятор арифметических выражений в автокод гипотетической ЭВМ.

11.4. Средства проверки лексики, синтаксиса и семантики

4.1. Проверка текстов английского языка.

4.2. Проверка текстов русского языка.

4.3. Анализатор программ на языке C.

4.4. Анализатор программ на языке BASIC.

4.5. Анализатор программ на языке Pascal.

Библиографический список использованной литературы

1. Ахо А. *Теория синтаксического анализа, перевода и компиляции*: Пер. с англ. / А. Ахо., Д. Ульман. М.: Мир, 1978. Т. 1: Синтаксический анализ. 612 с. Т. 2: Компиляция. 487 с.
2. Бауэр Ф. Л. *Информатика. Вводный курс*: В 2 ч. / Ф. Л. Бауэр, Г. Гооз. Ч. 2: Пер. с нем. М.: Мир, 1990. 423 с.
3. Братчиков И. Л. *Синтаксис языков программирования* / И. Л. Братчиков. М.: Наука, 1975. 232 с. (Библиотечка программиста).
4. Варсанофьев Д. В. *Основы компиляции*: Лекции, МГУ, ф-т ВМиК, 1991–93: [Электронный ресурс] / Д. В. Варсанофьев, А. Г. Дымченко. Режим доступа: `http://kit.kulichki.net/library/compilat.zip`, свободный.
5. Волкова И. А. *Формальные грамматики и языки. Элементы теории трансляции*: Учеб. пос. / И. А. Волкова, Т. В. Руденко. 2-е изд., перераб. и доп. М.: МГУ, 1999. 62 с.
6. Гладкий А. В. *Формальные грамматики и языки* / А. В. Гладкий. М: Наука, 1973. 368 с.
7. Грис Д. *Конструирование компиляторов для цифровых вычислительных машин*: Пер. с англ. / Д. Грис; Под ред. Ю. М. Баяковского, В. С. Штаркмана. М.: Мир, 1975. 544 с.
8. Льюис Ф. *Теоретические основы проектирования компиляторов*: Пер. с англ. / Ф. Льюис Ф., Д. Розенкранц, Р. Стирнз. М.: Мир, 1979. 654 с.

9. Молчанов А. Ю. *Системное программное обеспечение*: Учебник для вузов / А. Ю. Молчанов. СПб.: Питер, 2003. 396 с.
10. Нивергельт Ю. *Машинный подход к решению математических задач*: Пер. с англ. / Ю. Нивергельт, Дж. Фаррар, Э. Рейнголд. М.: Мир, 1977. 351 с.
11. Пратт Т. *Языки программирования: разработка и реализация* / Т. Пратт, М. Зелковиц; Под общей ред. А. Матросова. 4-е изд. СПб.: Питер, 2002. 688 с.
12. Рейуорд-Смит В. Дж. *Теория формальных языков. Вводный курс*: Пер. с англ. / В. Дж. Рейуорд-Смит. М.: Радио и Связь, 1988. 128 с.
13. Романовский И. В. *Дискретный анализ*: Уч. пос. для студентов, специализирующихся по прикладной математике и информатике / И. В. Романовский. 2-е изд., испр. СПб.: Невский Диалект, 2000. 240 с.
14. Хопкрофт Д. *Введение в теорию автоматов, языков и вычислений*, 2-е изд.: Пер. с англ. / Д. Хопкрофт, Р. Мотвани, Д. Ульман; Под ред. А. Б. Ставровского. М.: Издательский дом «Вильямс», 2002. 528 с.
15. Grune D. *Parsing Techniques: A Practical Guide*: [Электронный ресурс] / D. Grune, C. Jacobs. Department of Mathematics and Computer Science, Vrije Universiteit, Amsterdam, The Netherlands. 1998. 321 p. Режим доступа: `http://kit.kulichki.net/library/parsing.zip`, свободный.
16. Rytter W. *100 Exercises in the theory of automata and formal languages* / W. Rytter. Research report 99, April 1987. Department of Computer Science, University of Warwick, Conventry, UK. 12 p.
17. Sippu S. *Parsing theory* / S. Sippu, E. Soisalon-Soininen. Vol. 1: Languages and parsing. Berlin: Springer-Verlag, 1988. 228 p.

Учебное издание

КОРЗУН Дмитрий Жоржевич

ПРАКТИКУМ ПО ФОРМАЛЬНЫМ ГРАММАТИКАМ И ЯЗЫКАМ

Редактор *И. И. Куроптева*

Оригинал-макет и компьютерная верстка — *Д. Ж. Корзуна*

Подписано в печать 01.11.2004. Формат 60 × 84 1/16.
Бумага офсетная. Печать офсетная.
Уч.-изд. л. 2.2. Тираж 150 экз. Изд. № 163.

Государственное образовательное учреждение
высшего профессионального образования
Петрозаводский государственный университет
Типография Издательства ПетрГУ
185910, Петрозаводск, пр. Ленина, 33