

ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО  
ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ  
УФИМСКИЙ ГОСУДАРСТВЕННЫЙ АВИАЦИОННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Кафедра ТК

курс лекций по дисциплине

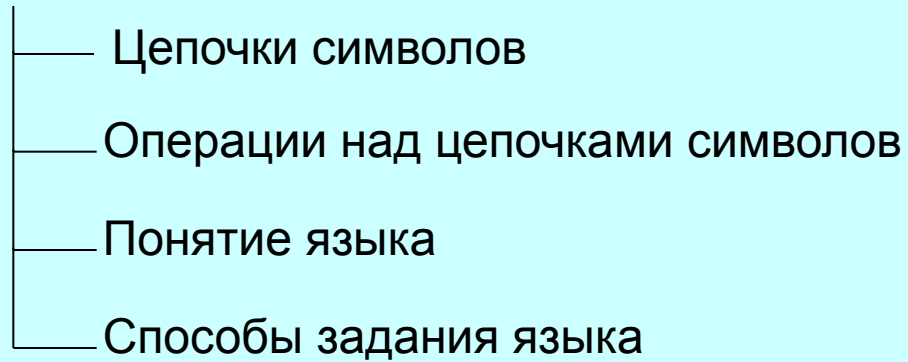
# Методы построения трансляторов

Тема: Языки и цепочки символов

Преподаватель: к.т.н., доцент Карамзина А.Г.

## Тема № 3

### Языки и цепочки символов



## Цепочки символов

**Цепочка символов (строка)** – произвольная последовательность символов, записанных один за другим.

Цепочки символов обозначаются греческими буквами:  $\alpha$ ,  $\beta$ ,  $\gamma$ .

Цепочка – это необязательно набор символов.

Последовательность «*aaaabbbb*» – это набор символов.

*Для цепочки символов важен состав и количество символов в ней, а также порядок символов в цепочке.*

Цепочки «*a*» и «*aa*», а также «*ab*» и «*ba*» – это различные цепочки символов.

Цепочки символов  $\alpha$  и  $\beta$  **равны (совпадают)**,  $\alpha=\beta$ , если они имеют один и тот же состав символов, одно и то же их количество и одинаковый порядок следования символов в цепочке.

Количество символов в цепочке называют **длиной цепочки**.

Длина цепочки символов  $\alpha$  обозначается как  $|\alpha|$  (если  $\alpha=\beta$ , то и  $|\alpha|=|\beta|$ ).

# Операции над цепочками символов

- **конкатенация** (сложение, объединение) двух цепочек символов – это дописывание второй цепочки в конец первой.

Конкатенация цепочек  $\alpha$  и  $\beta$  обозначается  $\alpha\beta$

операция конкатенации не обладает свойством КОММУТАТИВНОСТИ  
 $\exists \alpha$  и  $\beta$ , такие, что  $\alpha\beta \neq \beta\alpha$

- **обращение**  
Обращение

справедливо следующее равенство  $\forall \alpha, \beta: (\alpha\beta)^R = \beta^R\alpha^R$

- **итерация** (повторение) цепочки  $n$  раз где  $n \in \mathbb{N}, n > 0$

это

справедливы следующие равенства  
 $\forall \alpha: \alpha^1 = \alpha, \alpha^2 = \alpha\alpha, \alpha^3 = \alpha\alpha\alpha, \dots$  и т. д.

Итерация цепочки  $\alpha$   $n$  раз обозначается  $\alpha^n$

Для пустой цепочки справедливы следующие равенства:

**Пустая цепочка**  $\lambda$

Пустую цепочку обозначают  $\lambda$

1.  $|\lambda| = 0$ ;
2.  $\forall \alpha: \lambda\alpha = \alpha\lambda = \alpha$ ;
3.  $\lambda^R = \lambda$ ;
4.  $\forall n \geq 0: \lambda^n = \lambda$ ;
5.  $\forall \alpha: \alpha^0 = \lambda$ .

## Понятие языка

**Язык** – это заданный набор символов и правил, устанавливающих способы комбинации этих символов между собой для записи осмысленных текстов.

**Алфавит** – это счетное множество допустимых символов языка

обозначается символом  $V$ .

Цепочка символов  $\alpha$  является цепочкой над алфавитом  $V$ :  $\alpha(V)$ , если в нее входят только символы, принадлежащие множеству символов  $V$ .

Для любого алфавита  $V$  пустая цепочка  $\lambda$  может как являться, так и не являться цепочкой  $\lambda(V)$ .

**Язык**  $L(V)$  – это множество всех конечных цепочек над алфавитом  $V$ , принадлежащих языку  $L(V)$ .

Если  $V$  – некоторый алфавит, то:  
 $V^+$  – множество всех цепочек над алфавитом  $V$  без  $\lambda$ ;  
 $V^*$  – множество всех цепочек над алфавитом  $V$ , включая  $\lambda$ .  
 Цепочку символов, принадлежащую данному  $V^*$  языку, часто называют предложением языка, а множество цепочек символов некоторого языка  $L(V)$  – множеством предложений этого языка.

## Способы задания языка

- перечислением всех допустимых цепочек языка – *является чисто формальным и на практике не применяется;*
- указанием способа порождения цепочек языка (*заданием грамматики языка*) *предусматривает некоторое описание правил, с помощью которых строятся цепочки языка;*
- определением метода распознавания цепочек языка – *предусматривает построение некоторого логического устройства (распознавателя) – автомата, который на входе получает цепочку символов, на выходе выдает ответ: принадлежит или нет эта цепочка заданному языку.*

## Способы задания языка

**Лексика** языка – это совокупность слов (словарный запас) языка.

Сло  
СОСТ  
КОНС

**Лексическими единицами русского языка являются слова русского языка, а знаки препинания и пробелы представляют собой разделители, не образующие лексем.**

**Син**  
КОНС

**Лексическими единицами алгебры являются числа, знаки математических операций, обозначения функций и неизвестных величин.**

Синт  
КОТО

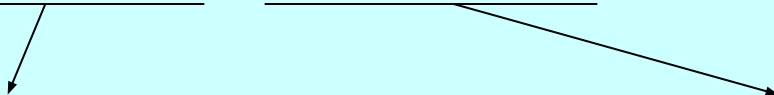
**Лексическими единицами в языках программирования являются ключевые слова, идентификаторы, константы, метки, знаки операций; в них также существуют и разделители (запятые, скобки, точки с запятой и т. д.).**

**Сем**  
язы

Семантика определяет «содержание языка» – задает значение для всех допустимых цепочек языка.

## Способы задания языка

Языки программирования занимают некоторое промежуточное положение между формальными и естественными языками.



с формальными языками их объединяют строгие синтаксические правила, на основе которых строятся предложения языка.

также из алгебры языки программирования переняли основные обозначения математических операций, что также делает их более понятными человеку.

от языков естественного общения в языки программирования перешли лексические единицы, представляющие основные ключевые слова



## Способы задания языка

Для задания языка программирования необходимо решить следующие задачи:

- *определить множество допустимых символов языка – решается легко – при определении алфавита языка, автоматически определяется множество допустимых символов (для языков программирования алфавит – это чаще всего тот набор символов, который можно ввести с клавиатуры);*

***Для решения используются другие подходы:***

- ***изложить смысл программы, написанной на языке программирования, на другом языке, более понятном тому, кому адресована программа;***
- ***использовать для проверки смысла некоторую «идеальную машину», которая предназначена для выполнения программ, написанных на данном языке.***

сится к теории формальных языков.