

ОСНОВЫ ПРОГРАММНОГО КОНСТРУИРОВАНИЯ

Лекция № 2
12 сентября 2016 г.



ИНЬ И ЯН ПРОГРАММИРОВАНИЯ



- Алгоритмы.
- Структуры данных.
- Алгоритмы + Структуры данных = Программы.

ДАННЫЕ (ИНФОРМАЦИЯ)

- Данные для человека:
 - Числа (целые, вещественные, комплексные)...
 - Наборы чисел (ряды, векторы, матрицы).
 - Текст (символы).
 - Изображения (видео).
 - Звук.
 - «Записи».
- Компьютер может:
00011100111010101

МАШИННОЕ ПРЕДСТАВЛЕНИЕ ДАННЫХ

- Целые числа — двоичная система счисления, дополнительный код.
- Вещественные числа — представление с плавающей точкой.
- Символы — числа (кодировка).
- Набор чисел, символов — массив.
- Картинка — массив пикселов с цветом (RGB).
- Звук — набор отсчетов (квантованная амплитуда).

АЛГОРИТМ

- Решает определенную задачу.
- Конечная упорядоченная последовательность действий.
- Обычно имеет входные параметры и выходные результаты.
- Пример: алгоритм Евклида поиска НОД двух чисел:
 - $\text{НОД}(a, b) = \text{НОД}(a-b, b)$ если $a > b$

ЯЗЫКИ ПРОГРАММИРОВАНИЯ

- Структуры данных:
 - Типы данных (числа, символы, строки, ...).
 - Переменные.
- Алгоритмы:
 - Операции над данными, в определенной последовательности — согласно синтаксису языка.

СИМВОЛЫ

- Кодировка: соответствие символа коду (и наоборот).
- ASCII: основа всего.
 - 32 управляющих символа.
 - 96 информационных СИМВОЛОВ.

! " # \$ % & ' () * + , - . /
0 1 2 3 4 5 6 7 8 9 : ; < = > ?
@ A B C D E F G H I J K L M N O
P Q R S T U V W X Y Z [\] ^ _
` a b c d e f g h i j k l m n o
p q r s t u v w x y z { | } ~

"ОРК @ AFTI"

~ [79, 80, 75, 32, 64, 32, 65, 70, 84, 73]

УПРАВЛЯЮЩИЕ СИМВОЛЫ ASCII

- Null, 0x00, \0 — пустой символ.
- Bell, 0x07, \a — звуковой сигнал.
- Backspace, 0x08, \b — возврат на шаг.
- Character Tabulation, 0x09, \t — горизонтальная табуляция.
- Line Feed, 0x0A, \n — перевод строки.
- Line Tabulation, 0x0B, \v — вертикальная табуляция.
- Form Feed, 0x0C, \f — смена страницы.
- Carriage Return, 0x0D, \r — возврат каретки.

НАЦИОНАЛЬНЫЕ КОДИРОВКИ

- КОИ-8 (код обмена информацией, 8 бит).
 - "ОПК АФТИ" без старшего бита: "opk afti"
 - псевдографика: ||—||—||—||—||
- Windows-1251.
 - "я": 0xFF
 - нет псевдографики :(

```
{C:\Work\cypok\JET\repo} - Far 3.0.4510 x64
C:\Work\cypok\HackDay
n      Name
..
2012 - HackDay I - BS M}
2013 - HackDay II - NXT}
2014 - HackDay III - JE}
2015 - HackDay IV - LED}

Notifier with help of xcr Folder 20.01.16 13:08
Bytes: 0, files: 0, folders: 4

C:\Work\cypok\JET\repo
n      Name
..
.git
.idea
.svn
annot-java
arm
bench
buildsystem
comp
core
depwizard
doc
importresolver
internal-doc
java-api
java-libs
jet-distro
jet-profile
jreversions
jvmsupport
mualyzer
plugins
r-n-d
readme
RP
runtime
samples
scala-libs
tomcat
tools









Up 09.09.16 22:34
Bytes: 13 210, files: 10, folders: 35

Far Manager, version 3.0 (build 4510) x64
Copyright © 1996-2000 Eugene Roshal, Copyright © 2000-2016 Far Group

C:\Work\cypok\JET\repo>
1Help 2UserMn 3View 4Edit 5Copy 6RenMov 7MkFold 8Delete 9ConfMn 10Quit 11Plugin 12Screen
```

Far Manager, сборка от 15.01.2016

UNICODE

- Решение проблем с 8-битными кодировками.
- Комбинированные символы (состоят из code point).
 - $\text{ë (U+0451)} = \text{e (U+0435)} + \text{ö (U+0308)}$
- Много кодов: 1 112 064.
 - В версии 9.0, июнь 2016 г., используется лишь 128 237.
-         ...

КОДИРОВАНИЕ UNICODE

- UTF-32: каждый code point – 4 байта.
 - Linux.
- UTF-16: либо 2 байта, либо 4 байта (суррогатные пары).
 - Java, Windows.
- UTF-8: 1-4 байт (хитрое кодирование).
 - Web, Linux, macOS.

СТРОКИ

- Строка — массив символов, но может иметь переменную длину!
- Нужно как-то хранить длину, как?

ДЛИНА СТРОКИ

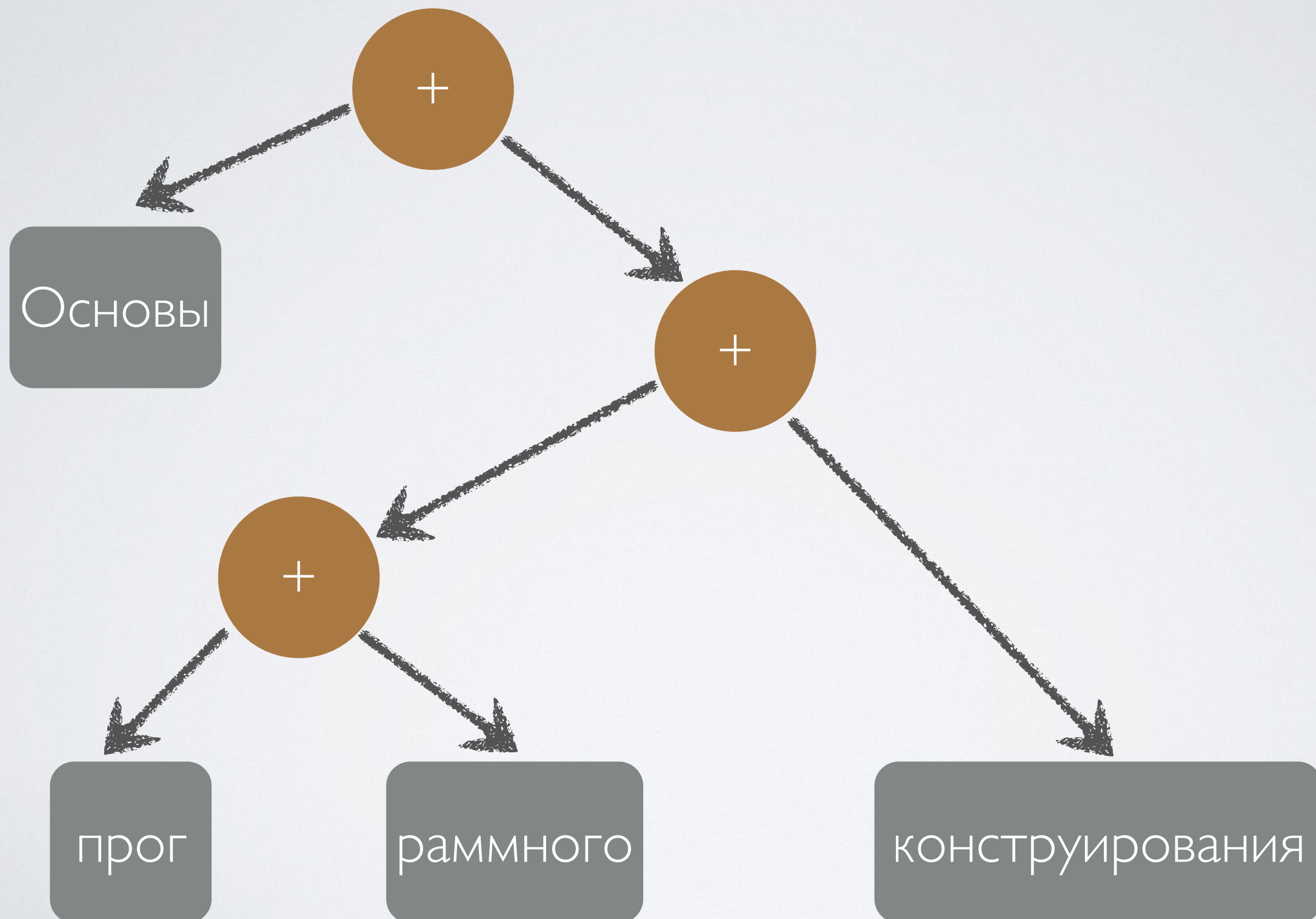
- Два стандартных приема:
 - Хранение длины в начале строки (первые 1/2/4 байта).

6	s	t	r	i	n	g
---	---	---	---	---	---	---

- Специальный символ. В языке C этим символом является символ с кодом 0.

s	t	r	i	n	g	\0
---	---	---	---	---	---	----

«ВЕРЕВКИ»



ОПЕРАЦИИ СО СТРОКАМИ

- Конкатенация.
- Вычисление длины.
- Выделение подстроки.
- Взятие символа по индексу i .
- Удаление/замена подстроки.
- Поиск символов.
- Поиск подстроки.
- Сравнение двух строк.
- Копирование.
- Разбиение на подстроки (например, на слова).
- Преобразование в число и наоборот.

СЕРИАЛИЗАЦИЯ

- Строка в широком смысле — последовательность байт известного размера.
- Сериализация: превращение любой структуры данных в строку (десериализация — восстановление структуры данных).
- Пример сериализации:
массив целых чисел (*int*) переменной длины (*N*).

СЕРИАЛИЗАЦИЯ МАССИВА

- $N+1$ целых чисел: длина + значения самого массива.
- Big-endian и little-endian:
 - $12345678_{16} \Rightarrow 12\ 34\ 56\ 78$ (big-endian, «от старшего к младшему»).
 - $12345678_{16} \Rightarrow 78\ 56\ 34\ 12$ (little-endian, «от младшего к старшему»).
- Сериализация матрицы? Произвольного множества векторов?



КОНЕЦ ВТОРОЙ ЛЕКЦИИ