Tipos de dados

ALGORITMOS E ESTRUTURAS DE DADOS III

Prof. Marcos André S. Kutova

Tipos primitivos de dados

TIPOS INTEIROS

- byte Número inteiro de 8 bits com sinal
 - -128 a +127
- short Número inteiro de 16 bits com sinal
 - -32.768 a +32.767
- int Número inteiro de 32 bits com sinal
 - -2.147.483.648 a 2.147.483.647
- long Número inteiro de 64 bits com sinal
 - -9.223.372.036.854.775.808 a
 - +9.223.372.036.854.775.807

Tipos primitivos de dados

TIPOS REAIS

 float Número de ponto flutuante de 32 bits com sinal ±1,40129846e-45 a

±3,40282347e+38

double Número de ponto flutuante de 64 bits com sinal

±4,94065645841246544e-324 a

±1,79769313486231570e+308

Bits	Sinal	Expoente	Fração
Precisão simples (float)	1	8	23
Precisão dupla (double)	1	11	52

Tipos primitivos de dados

OUTROS TIPOS

• char Carácter Unicode de 16 bits, sem sinal,

com valor de 0 a 65.535 (em Java).

• boolean Variável lógica que indica falso ou verdadeiro

Campos numéricos

Campos numéricos

- Decimal 11
 - Formato usado na comunicação humana. Sempre será o formato a ser usado nas interfaces com o usuário. O computador, no entanto, não armazena dados no formato decimal.
- Binário 0b00001011
 - Formato usado nas operações dos componentes eletrônicos (CPU, memória, etc.).
- Hexadecimal 0x0B
 - Formato usado nas representações de bytes
- String "11"
 - Formato usado na representações textuais. A string usada no exemplo é armazenada por meio dos bytes 0x31 0x31. Cada dígito é, na verdade, um caráter com valor 49 (0x31 em hexadecimal).

Exemplos

- Quantidade de produtos em um cupom fiscal de uma compra de supermercado:
 - short (até 32.767, pois 127 pode ser pouco)
- Contador de registros de operações bancárias
 - long (são milhões de registros por dia...)
- Preço de um livro
 - float (precisão suficiente para a faixa de valores)

Datas e horas

Datas e horas

- Em Java, uma data é um campo numérico que indica a quantidade de milissegundos passadas desde as 00:00 de 01/01/1970.
 - 1.439.500.953.888 = 13/08/2015 21:22:33.888
- No Excel, uma data possui uma parte inteira e uma parte decimal. A parte inteira, indica o número de dias passados desde 01/01/1900. A parte decimal indica a fração de um dia.
 - 42229,89067 = 13/08/2015 21:22:33.888
- A vantagem de se armazenar datas e horas como números é a facilidade de se fazer operações matemáticas (e a economia de bytes)

Os "falsos" campos numéricos

Os "falsos" campos numéricos

Exemplos

• CEP: 30535-901

• CPF: 072.424.817-08

• CNPJ: 12.454.755/0002-26

• Telefone: (31) 3319-4444

• Exercício:

 Apresente uma alternativa para o armazenamento de placas de carro.

Identificadores

Identificadores

- Identificadores devem ser números sequenciais exclusivos que não carregam significado
 - A alteração de um identificador compromete todas as referências à entidade
 - O formato numérico assegura o melhor aproveitamento da capacidade de armazenamento

• Exercício:

A PUC Minas recebe cerca de 5.000 novos alunos por semestre.
 Durante quanto tempo identificadores de matrículas dos tipos short, int e long serão válidos sem o reaproveitamento de números?

Sistemas de codificação

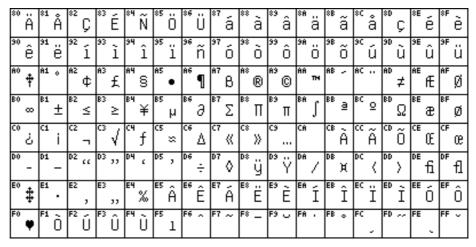
Sistemas de codificação

- Representação binária (ou hexa) de cada símbolo
- ASCII (anos 1960)

 American Standard Code for Information Interchange
 - Representação de 128 símbolos (7 bits)
 - 100 0001 = 65 = 0x41 = 'A'
 - O 8º bit (mais significativo) era usado como bit de paridade nas comunicações
 - Faltavam os caracteres latinos acentuados, os chineses, os cirílicos (russos), ...

Decimal	Hexadecimal	Binary	0ctal	Char	Decimal	Hexadecimal	Binary	0ctal	Char	Decimal	Hexadecimal	Binary	0ctal	Char
0	0	0	0	[NULL]	48	30	110000	60	0	96	60	1100000		*
1	1	1	1	[START OF HEADING]	49	31	110001		1	97	61	1100001		a
2	2	10	2	[START OF TEXT]	50	32	110010		2	98	62	1100010		b
3	3	11	3	[END OF TEXT]	51	33	110011		3	99	63	1100011		c
4	4	100	4	[END OF TRANSMISSION]	52	34	110100		4	100	64	1100100		d
5	5	101	5	[ENOUIRY]	53	35	110101		5	101	65	1100101		e
6	6	110	6	[ACKNOWLEDGE]	54	36	110110		6	102	66	1100110		f
7	7	111	7	[BELL]	55	37	110111		7	103	67	1100111		g g
8	8	1000	10	[BACKSPACE]	56	38	111000		8	104	68	1101000		h
9	9	1001	11	[HORIZONTAL TAB]	57	39	111001		9	105	69	1101001		ï
10	Á	1010	12	[LINE FEED]	58	3A	111010			106	6A	1101010		
11	В	1011	13	[VERTICAL TAB]	59	3B	111011		;	107	6B	1101011		k
12	Č	1100	14	[FORM FEED]	60	3C	111100		, <	108	6C	1101100		î
13	D	1101	15	[CARRIAGE RETURN]	61	3D	111101		=	109	6D	1101101		m .
14	E	1110	16	[SHIFT OUT]	62	3E	111110		>	110	6E	1101110		n
15	F	1111	17	[SHIFT IN]	63	3F	111111		?	111	6F	1101111		0
16	10	10000	20	[DATA LINK ESCAPE]	64	40	1000000		@	112	70	1110000		р
17	11	10000	21	[DEVICE CONTROL 1]	65	41	10000001		A	113	71	1110000		q
18	12	10001	22	[DEVICE CONTROL 2]	66	42	1000001		В	114	72	1110001		r
19	13		23	[DEVICE CONTROL 3]	67	43	1000011		Č	115	73	1110010		5
20	14	10100	24	[DEVICE CONTROL 3]	68	44	1000011		D	116	74	1110101		t
21	15	10100	25	[NEGATIVE ACKNOWLEDGE]	69	45	1000100		E	117	75	1110100		u
22	16	10111	26	1	70	46	1000101		F	118	76	1110101		v
23	17	10111	27	[SYNCHRONOUS IDLE] [ENG OF TRANS. BLOCK]	71	47	1000111		G	119	77	1110110		w
24	18	11000	30	[CANCEL]	72	48	1001000		н	120	78	1111000		X
25	19	11000	31	[END OF MEDIUM]	73	49	1001000		î .	121	79	1111000		
26	1A	11010	32	[SUBSTITUTE]	74	4A			i	122	79 7A	1111001		y z
27	1B	11011	33	[ESCAPE]	75	4B	1001010		K	123	7B	1111010		{
28	1C	11100	34	[FILE SEPARATOR]	76	4C	1001011		Ĺ	124	7C	11111011		1
29	1D	11101	35	[GROUP SEPARATOR]	77	4D	1001101		М	125	7D	1111100		}
30	1E		36	[RECORD SEPARATOR]	78	4E	1001101		N	126	7E	11111101		~
31	1F	11111		[UNIT SEPARATOR]	79	4F	1001111		0	127	7F	1111111		[DEL]
32	20	100000		[SPACE]	80	50	1010000		P	127	7.	1111111	1//	[DEL]
33	21	100000		!	81	51	1010000		Q	l				
34	22	100001			82	52	1010001		R	l				
35	23	100010		#	83	53	1010010		S	l				
36	24	100011		** \$	84	54	1010100		T	l				
37	25	100100		%	85	55	1010101		ΰ	l				
	26				86	56			v	l				
38 39	27	100110 100111		&	87	57	1010110		w	l				
40	28			1	88	58	1010111			l				
		101000		1			1011000		X	l				
41	29	101001		*	89	59	1011001		Y					
42	2A	101010			90	5A	1011010		Z					
43	2B	101011		+	91	5B	1011011		Ĺ					
44	2C	101100		1	92	5C	1011100		\					
45	2D	101101		•	93	5D	1011101		,					
46 47	2E 2F	101110		;	94 95	5E 5F	10111110							
47	21	101111	37	1	95	DF.	1011111	. 13/	-	I				

Codificações de 8 bits



Mac Roman

ISO-8859-5 (Cyrillic)

ISO-8859-1 (Latin-1)



Unicode

Unicode

- Sistema de codificação capaz de representar texto de qualquer sistema de escrita existente (mais de 100 mil símbolos)
 - Sistemas latino, arábico, cirílico, chinês, hebráico, ...
 - Símbolos matemáticos, geométricos, musicais, setas, ícones, emojis,
 ...
 - Escrita cuneiforme, Braille, runas, élfico, ...
 - Próximos: hieróglifos egípcios, alfabeto babilônico, ...
- Referências:
 - http://unicode.org/
 - http://www.unicodetables.com/



Unicode

- **UTF-8** (8-bit Unicode Transformation Format)
 - Codificação Unicode de comprimento variável (1 a 4 bytes), que pode representar qualquer caráter (code point) do padrão.
 - Padrão usado na Internet e na Web
- UTF-16
 - Codificação de 2 ou 4 bytes, usada principalmente para escrita em idiomas dos países asiáticos
- UTF-32
 - Codificação de comprimento fixo de 4 bytes

UTF-8

Bits do code point	Primeiro code point	Último code point	Bytes	Byte 1	Byte 2	Byte 3	Byte 4
7	U+0000	U+007F	1	0xxxxxxx			
11	U+0080	U+07FF	2	110xxxxx	10xxxxxx		
16	U+0800	U+FFFF	3	1110xxxx	10xxxxxx	10xxxxxx	
21	U+10000	U+1FFFFF	4	11110xxx	10xxxxxx	10xxxxxx	10xxxxxx

- 1 byte Tabela ASCII
- 2 bytes Caracteres latinos, hebraicos, gregos, ...
- 3 bytes Caracteres chineses, japoneses, coreanos, ...
- 4 bytes Alguns outros caracteres e símbolos

UTF-8

Representação de strings

С	0	n	С	е	i	Ç	ã	0
43	6F	6E	63	65	69	C3A7	СЗАЗ	6F

9 caracteres – 11 bytes

Strings

Strings

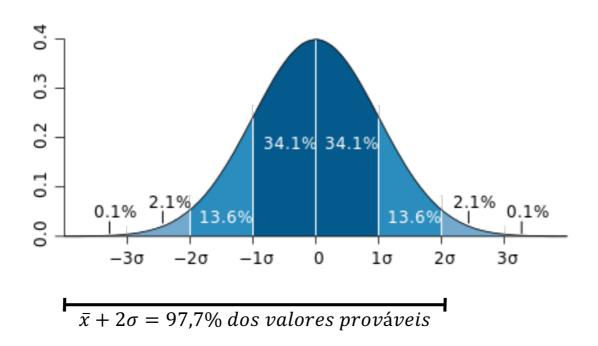
- Cadeias de caracteres usadas para representar um palavra, uma informação ou um texto
- Valor em bytes de cada caráter depende do sistema de codificação
 - ISO-8859-1: cada caráter usa 1 byte
 - UTF-8: cada caráter por usar de 1 a 4 bytes
- Campos que contém strings podem ser de tamanho fixo ou variável

Strings de tamanho fixo

- Como determinar o tamanho do campo
 - Campos de tamanho naturalmente fixo
 - CPF, CEP, UF, Telefone, Conta Corrente, EAN-13, ...
 - Campos limitados externamente
 - tamanho do relatório, da etiqueta, da tela, ...
 - Probabilidade estatística

Strings de tamanho fixo

Distribuição normal



Exemplo

- Email
 - Especificação (nome@dominio.com)
 - Nome até 64 caracteres
 - Domínio até 254 caracteres
 - Total: 64 + 1 + 254 = 319 caracteres
 - Estatística
 - Análise de 323 emails
 - Mínimo = 12
 - Máximo = 47
 - Média = 23,0
 - Desvio padrão = 5,2

Faixas:

$$0 a (\bar{x} + 1\sigma) = 84,1\% = 28,2$$

$$0 a (\bar{x} + 2\sigma) = 97,7\% = 33,4$$

$$0 \ a \ (\bar{x} + 3\sigma) = 99,8\% = 38,6$$

Strings de tamanho variável

Com indicador de tamanho

IND.	ГАМ.	С	0	n c		е	i	(3		О	
00	ОВ	43	6F	6E	63	65	69	C3	A7	C3	А3	6F

Com delimitador

С	0	n	С	е	i	Ç	3	6	ă	0	\n
43	6F	6E	63	65	69	С3	A7	С3	А3	6F	OA

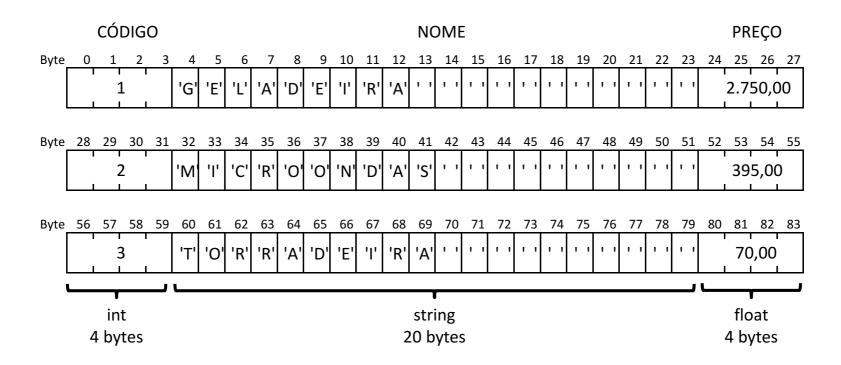
• Delimitadores tradicionais: \n \0 ; | (e outros)

Registros

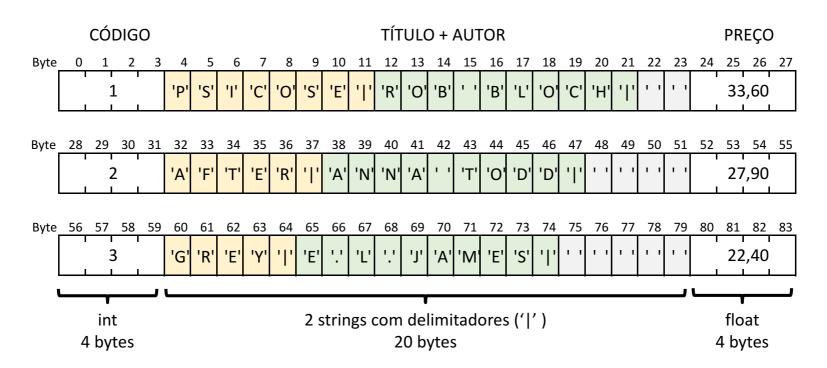
• Conjunto de dados (ou campos) sobre uma mesma entidade

Código	Nome	Capaci- dade	Altura (cm)	Largura (cm)	Profundi- dade (cm)	Cor	Preço
1981091	Refrigerador Electrolux Frost Free DW52X com dispenser externo de água 456 L	456 litros	193	71	76	Inox	R\$ 2.744,10
3608237	Forno de Microondas LG Grill Easy Clean MH7053R Espelhado - 30 L	30 litros	29	51	45	Branco	R\$ 395,43
2170868	Torradeira Britânia Tosta Pane	2 torradas	18	17	29	Inox - Preto	R\$ 71,91

 Registros de tamanho fixo com campos de tamanho fixo



 Registros de tamanho fixo com campos de tamanho variável

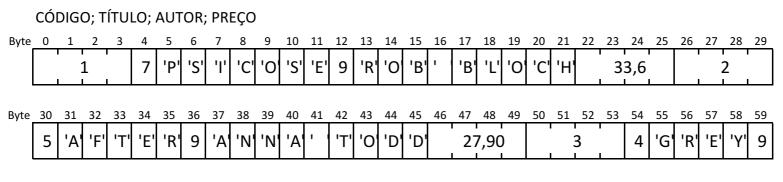


• Exercício:

- Compare os tamanhos de um registro de tamanho fixo, (1) usando campos de tamanho fixo e (2) usando campos de tamanho variável, que descreva um endereço, com os seguintes campos:
 - Rua, número, complemento, bairro, CEP, cidade, estado

- Na maioria das situações, a quantidade de campos será fixa e sempre na mesma ordem
 - Alternativas para processamento:
 - Leitura/escrita dos campos na ordem predeterminada
 - Uso de indicadores de tamanho
 - Uso de delimitadores de registro
- Casos especiais
 - Campos com múltiplos valores
 - Campos selecionados: uso de identificadores de campos

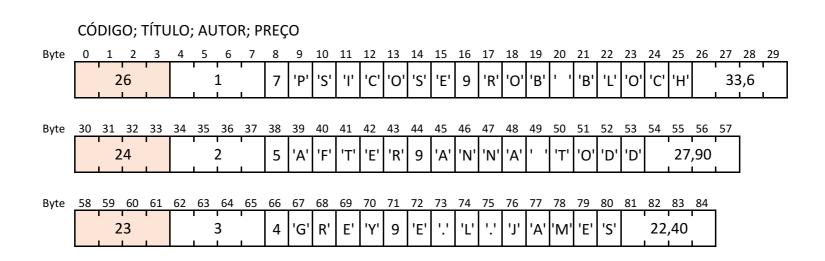
- Campos fixos
 - Os campos de cada registro são sempre os mesmos (mesmo se vazios)
 - Os campos de tamanho variável devem adotar indicador de tamanho ou delimitador



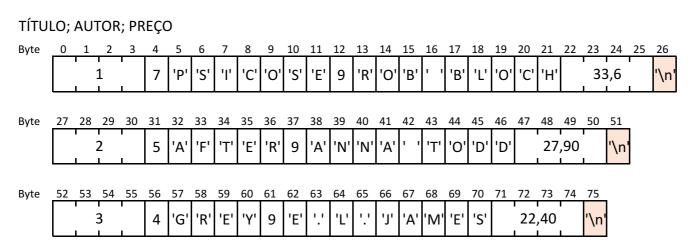
Byte										70	71	72
	'E'	:	'L'	1:	'J'	'A'	'M	' 'E'	'S'	22	2,40)

Nesse exemplo, apenas o acesso sequencial aos registros é possível. No entanto, o acesso aleatório pode ser feito por meio do uso de índices.

- Indicadores de tamanho de registro
 - Úteis para evitar o processamento individual dos campos, quando o registro não é o desejado
 - ao se saber o tamanho do registro, pode-se saltar todos os seus bytes



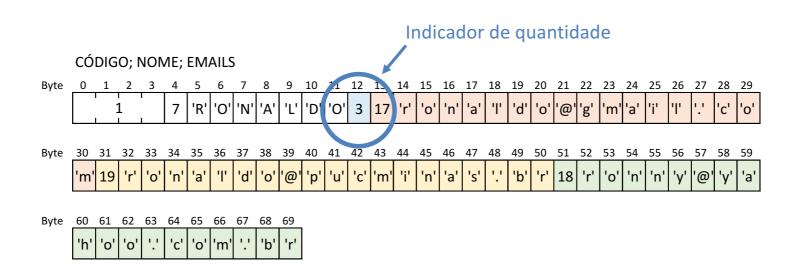
- Delimitadores de registro
 - Algumas linguagens de programação oferecem o método readLine()
 que facilita a leitura de um registro inteiro
 - Um método como o split() também pode ser usado na separação de vários registros



Outros tipos de campos

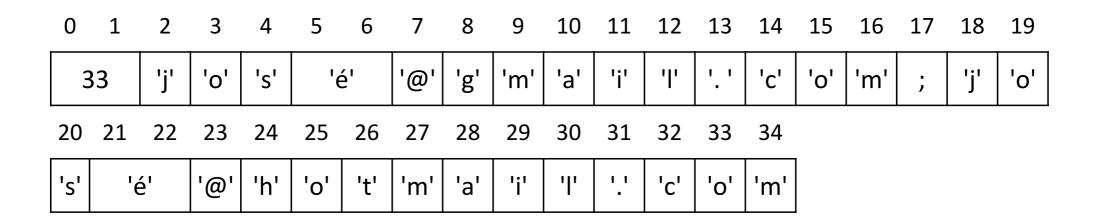
Campos com múltiplos valores

• Exemplo – emails de um cliente:



Campos com múltiplos valores

- Alternativa:
 - Usar uma string de tamanho variável e usar um separador entre os valores.



Campos com identificadores

- Normalmente, os campos sempre os mesmos e aparecem na mesma ordem
- Em algumas situações, porém, apenas alguns campos (de um conjunto muito grande) são usados
- Em outras situações, um campo pode receber informações diferentes:
 - Contato pode ser email, telefone, pager, etc.

Campos com identificadores

AND IN	INISTÉRIO DA PREVIDÊNCIA E ASSISTÊNCIA SOCIAL-MPAS	3. CÓDIGO DE PAGAMENTO	1600
(40)	STITUTO NACIONAL DO SEGURO SOCIAL - INSS	4. COMPETÊNCIA	11/2002
G	IUIA DA PREVIDÊNCIA SOCIAL – GPS	5, IDENTIFICADOR	4561098732
. NOME OU RAZ	ÃO SOCIAL/FONE/ENDEREÇO:	6. VALOR DO INSS	4520
Vera_	lauria da Jahra mital, 114 - Cambu - Curitiba - PR	7.	•
R. You	mulal, 114 - Cambii - Cambii	8.	
2.VENCIMENTO Uso exclusivo INS	S)	9. VALOR DE OUTRAS ENTIDADES	
ao estipulado em F	iada a utilização de GPS para recolhimento de receita de valor inferior Resolução publicada pelo INSS. A receita que resultar valor inferior	10. ATM/MULTA E JUROS	
deverá ser adicio subsequentes, até qu	mada à contribuição ou importância correspondente nos meses ue o total seja igual ou superior ao valor mínimo fixado	11. TOTAL	45,20
		12. AUTENTICAÇÃO BANG	CÁRIA

Byte	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
	1	53	'	-e	'r'	'a'	-	'L'	'ú'	'c'	ï	'a'	-	'd'	'a'	1 1	'S'	'i'	'I'	'v'	'a'	'\n'	'R'	:	'P'	'a'	' '	'm'	ï	't'	'a'	'ו'
Byte	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
	','	'1'	'1'	'4'	-1	'C'	'a'	'm'	'b'	'u'	ï	<u>-</u> !	'C'	'u'	'r'	'j'	't'	'i'	'b'	'a'	1_1	'P'	'R'	3	4	'1'	'6'	'0'	'0'	4	7	'1'
Byte	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91				
	'1'	'/'	'2'	'0'	'0'	'2'	5	10	'4'	'5'	'6'	'1'	'0'	'9'	'8'	'7'	'3'	'2'	6		45	,20	l 	11		45	,20	l 				

Arquivos CSV

Comma Separated Values (CSV)

 Formato de arquivo (RFC 4180) comumente usado no intercâmbio de dados entre sistemas

```
ID;Produto;Preço
1;Geladeira Branca;1.300,00
2;"TV LED 60"" Full HD";3.300,00
```

```
ID,Product,Price
1,Refrigerator,1300.00
2,"60"" Full HD LED TV",3300.00
```