
Introdução à Ciência da Computação: armazenamento de dados

Parte 2 – Texto

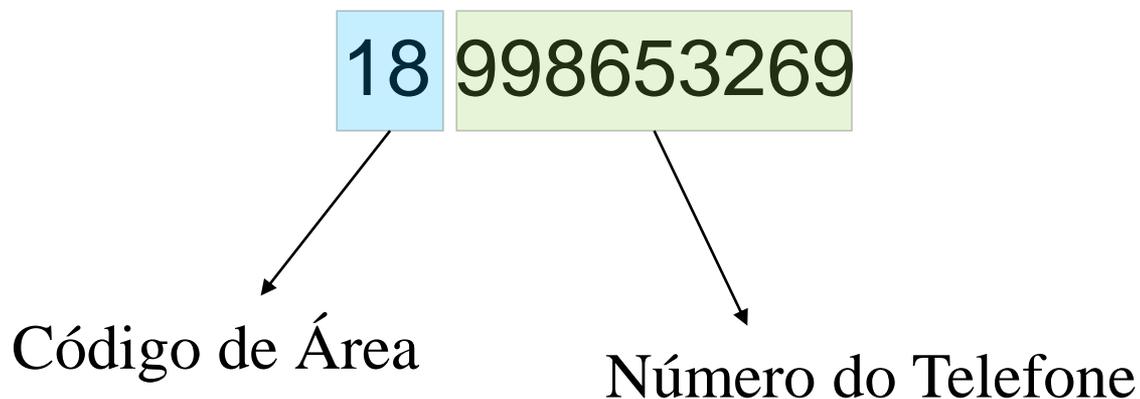
Prof. Danilo Medeiros Eler
danilo.eler@unesp.br

Conteúdo

- Representação e Armazenamento de Dados
 - Texto
 - Imagem
 - Número

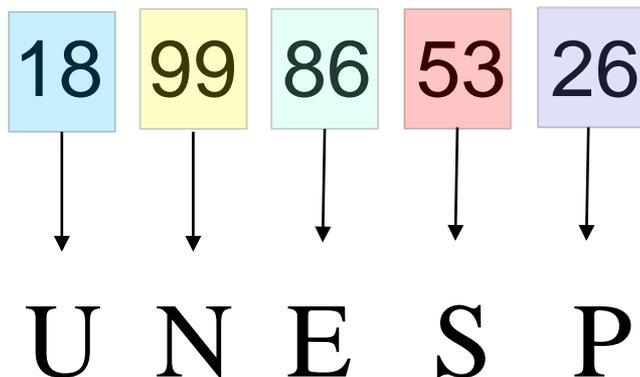
Ilustração

- Uma sequência de símbolos ainda poderia ser organizada de maneira que mais de um dado ou informação estivesse contido nele
 - Ex. telefone



Ilustração

- Uma sequência de símbolos ainda poderia ser organizada de maneira que mais de um dado ou informação estivesse contido nele
 - Ex. mensagem codificada



Padrão Binário – Ilustração

- Os padrões binários armazenados em um computador poderão ter diferentes significados dependendo do contexto da aplicação e do tipo de dado que representam
- O que o padrão binário abaixo representa?

000100111



Número decimal 39

Armazenamento de Texto

- Um texto, em qualquer idioma, é uma sequência de símbolos
 - 26 símbolos para letras maiúsculas (A, B, ..., Z)
 - 26 símbolos para letras minúsculas (a, b, ..., z)
 - 9 símbolos para números (0, 1, ..., 9)
 - Símbolos para pontuação (., ?, :, ;,, !)
 - Outros também podem ser utilizados para alinhamento e dar legibilidade ao texto

Armazenamento de Texto

- Cada símbolo utilizado pode ser representado por um padrão binário
- Por exemplo:
 - A palavra TEXTO
 - T → 1000011
 - E → 0001110
 - X → 1001100
 - T → 1000011
 - O → 1000000

Tamanho da Palavra

- O tamanho da palavra permite definir quantos símbolos poderão ser representados
 - Exemplo: palavra de 7 bits

0111010

Conseguiríamos representar apenas 128 padrões diferentes

Já seria possível representar o alfabeto e outros símbolos

Armazenamento de Texto

- Quantos bits são necessário para representar os símbolos utilizados em um idioma?

Armazenamento de Texto

- Quantos bits são necessário para representar os símbolos utilizados em um idioma?

$$\lceil \log_b x \rceil$$

x é a quantidade de padrões

b é a base

Armazenamento de Texto

- Exemplos:
 - Se quisermos representar as 26 letras do alfabeto com padrões binários

$$\lceil \log_b x \rceil$$

x é a quantidade de padrões

b é a base

Armazenamento de Texto

- Exemplos:
 - Se quisermos representar as 26 letras do alfabeto com padrões binários

$$\lceil \log_2 26 \rceil$$

x é a quantidade de padrões

b é a base

Armazenamento de Texto

- Exemplos:
 - Se quisermos representar as 26 letras do alfabeto com padrões binários

$$\lceil \log_2 26 \rceil = \lceil 4.70 \rceil$$

Armazenamento de Texto

- Exemplos:
 - Se quisermos representar as 26 letras do alfabeto com padrões binários

$$\lceil \log_2 26 \rceil = \lceil 4.70 \rceil = 5$$

Armazenamento de Texto

- Exemplos:
 - Se quisermos representar as 26 letras do alfabeto com padrões binários

$$\lceil \log_2 26 \rceil = \lceil 4.70 \rceil = 5$$

4 bits $2^4 = 16$ símbolos

Padrões de 0000 a 1111

Armazenamento de Texto

- Exemplos:
 - Se quisermos representar as 26 letras do alfabeto com padrões binários

$$\lceil \log_2 26 \rceil = \lceil 4.70 \rceil = 5$$

5 bits $2^5 = 32$ símbolos

Padrões de 00000 a 11111

Armazenamento de Texto

- Exemplos:
 - Se quisermos representar as 26 letras do alfabeto com padrões binários
 - Considerando também minúsculas e maiúsculas

$$\lceil \log_2 52 \rceil =$$

Armazenamento de Texto

- Exemplos:
 - Se quisermos representar as 26 letras do alfabeto com padrões binários
 - Considerando também minúsculas e maiúsculas

$$\lceil \log_2 52 \rceil = \lceil 5.70 \rceil = 6$$

6 bits $2^6 = 64$ símbolos

Padrões de 000000 a 111111

Armazenamento de Texto

- Exemplos:
 - Se quisermos representar as letras do alfabeto e outros símbolos
 - Suponha que tenhamos 128 símbolos

$$\lceil \log_2 128 \rceil =$$

Armazenamento de Texto

- Exemplos:
 - Se quisermos representar as letras do alfabeto e outros símbolos
 - Suponha que tenhamos 128 símbolos

$$\lceil \log_2 128 \rceil = \lceil 7 \rceil = 7$$

7 bits $2^7 = 128$ símbolos

Padrões de 0000000 a 1111111

Armazenamento de Texto

- O comprimento do **padrão binário** é dependente o número de símbolos que se deseja utilizar
- Quanto mais símbolos, maior será o comprimento do **padrão binário**

Comprimento do Padrão Binário

Quantidade de Símbolos	Comprimento do Padrão
2	1
4	2
8	3
16	4
32	5
64	6
128	7
256	8
65.536	16
4.294.967.296	32

Armazenamento de Texto

- Para representar texto é necessário
 - Decidir quais símbolos serão utilizados no texto
 - Associar cada símbolo com um padrão binário
- Diferentes conjuntos de padrões foram projetados para representar símbolos de texto
 - Cada conjunto é chamado de **código**
 - O processo de representação de símbolos é chamado de **codificação**

Armazenamento de Texto

- Códigos mais conhecidos
 - ASCII (American Standard Code for Information Interchange) – Código Padrão Americano para Troca de Informação (1960)
 - Desenvolvido pelo ANSI (American National Standards Institute) – Instituto Nacional Americano de Padronização
 - Utiliza 7 bits para cada símbolo
 - 128 diferentes símbolos podem ser representados
 - Códigos de 0 a 127

ASCII

NUL	DLE	SP	0	@	P	`	p
SOH	DC1 X0N	!	1	A	Q	a	q
STX	DC2	"	2	B	R	b	r
ETX	DC3 XOFF	#	3	C	S	c	s
EOT	DC4	\$	4	D	T	d	t
ENQ	NAK	%	5	E	U	e	u
ACK	SYN	&	6	F	V	f	v
BEL	ETB	'	7	G	W	g	w
BS	CAN	(8	H	X	h	x
HT	EM)	9	I	Y	i	y
LF	SUB	*	:	J	Z	j	z
VT	ESC	+	;	K	[k	{
FF	FS IS4	,	<	L	\	l	
CR	GS IS3	-	=	M]	m	}
SO	RS IS2	.	>	N	^	n	~
SI	US IS1	/	?	O	_	o	DEL

Armazenamento de Texto

- *Extended ASCII*
 - Aumentou 1 bit na codificação
 - Padrões binários com 8 bits permitiram utilizar 256 símbolos distintos (0 a 255)

Extended ASCII

	-0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-A	-B	-C	-D	-E	-F	
0-		☺	☹	♥	♦	♣	♠	•	◼	◊	♠	♀	♂	♪	♫	☀	0-
1-	▶	◀	↕	!!	¶	§	—	↕	↑	↓	→	←	L	↔	▲	▼	1-
2-		!	"	#	\$	%	&	'	()	*	+	,	-	.	/	2-
3-	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?	3-
4-	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	4-
5-	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_	5-
6-	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	6-
7-	p	q	r	s	t	u	v	w	x	y	z	{		}	~	△	7-
8-	Ç	ü	é	â	ã	à	á	ç	ê	ë	è	ï	î	í	Ä	Å	8-
9-	É	æ	Æ	ô	ö	ò	û	ù	ÿ	Ö	Ü	ø	£	¥	Pls	f	9-
A-	á	í	ó	ú	ñ	Ñ	ª	º	¿	¬	½	¼	¡	«	»		A-
B-	☄	☄	☄		†	‡	‡	π	τ	‡		π	‡	‡	‡	τ	B-
C-	L	⊥	⊥	†	—	†	†	‡	ℓ	ℓ	‡	π	‡	=	‡	±	C-
D-	⊥	π	π	ℓ	ℓ	F	π	‡	‡	J	γ	■	■	■	■	■	D-
E-	α	β	Γ	π	Σ	σ	μ	τ	Φ	Θ	Ω	δ	∞	φ	ε	∩	E-
F-	≡	±	≥	≤			÷	≈	°	.	.	√	n	²	■		F-
	-0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-A	-B	-C	-D	-E	-F	

Armazenamento de Texto

- Unicode
 - Utiliza 32 bits
 - $2^{32} = 4.294.967.296$ símbolos
 - Diferentes seção de código são alocados a diferentes símbolos de diferentes idiomas do mundo
 - Assim como símbolos gráficos e especiais

Unicode

Símbolo

Fonte: (texto normal) Subconjunto: Grego básico

ú	Û	ü	Ŵ	ŵ	Ŷ	ŷ	ÿ	Ž	ž	Ẑ	ẑ	Ž	ž	ƒ	f
À	á	Æ	æ	Ó	ó	^	v	-	˘	·	°	˘	˜	”	;
'	ˆ	Α	·	Ε	Η	Ι	Ο	Υ	Ω	ι	Α	Β	Γ	Δ	Ε
Ζ	Η	Θ	Ι	Κ	Λ	Μ	Ν	Ξ	Ο	Π	Ρ	Σ	Τ	Υ	Φ
Χ	Ψ	Ω	İ	ÿ	ά	έ	ή	ι	ϑ	α	β	γ	δ	ε	ζ
η	θ	ι	κ	λ	μ	ν	ξ	ο	π	ρ	ς	σ	τ	υ	φ

Símbolos usados recentemente:

√	—	Δ	μ	€	Σ	∅	ø	ε	β	α	■	€	£	¥	©
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

GREEK SMALL LETTER ALPHA ... Código do caractere: 03AC de: Unicode (hex)

Inserir Cancelar

Bibliografia

BIBLIOGRAFIA BÁSICA

1. BROOKSHEAR, J. G. **Ciência da computação**: uma visão abrangente. 5ª ed., Bookman Editora, 2000. 499p.
2. FOROUZAN, B. A., MOSHARRAF, F. **Fundamentos da Ciência da Computação**. 2ª ed., São Paulo: Cengage Learning, 2011. 560p.

BIBLIOGRAFIA COMPLEMENTAR

1. BROOKSHEAR, J. G. **Ciência da computação**: uma visão abrangente. 5ª ed., Bookman Editora, 2000. 499p.
2. CORMEN, T.H., Leiserson, C.E., Rivest R.L., Stein, C. **Algoritmos**: teoria e Prática. Rio de Janeiro: Editora Campus, 2002. 916p.
3. PLAUGER, P. L. **A Biblioteca Standard C**. Rio de Janeiro: Editora Campus, 1994. 614p.
4. PRATA, S. **C primer plus**, 4ª ed. SAMS Publishing, 2002. 931p.