



# Arquitetura de Computadores em SI

- Painel
- Meus cursos
- Arquitetura de Computadores em SI 2021-2
- Codificação em binário
- Caracteres e Codificação ASCII

## Caracteres e Codificação ASCII

Um [byte](#) pode ser interpretado como um caractere (letra, dígito, sinal de pontuação, etc.). Essa interpretação tem por base a [tabela ASCII](#), uma tabela arbitrária mas universalmente aceita que associa um caractere a cada byte. ([ASCII](#) é a sigla do American Standard Code for Information Interchange.) Por razões históricas, a tabela ASCII usa apenas bytes cujo primeiro bit é 0 e portanto, tem apenas 128 linhas. Eis uma pequena amostra da tabela:

byte	caractere
00111111	?
01000000	@
01000001	A
01000010	B
01000011	C
01100001	a
01100010	b
01100011	c
01111110	~

O conjunto de caracteres coberto pela tabela é conhecido como *alfabeto ASCII*. A parte principal desse alfabeto consiste nos seguintes caracteres:

```
! " # $ % & ' ( ) * + , - . /
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
{ | } ~
```

(O primeiro caractere da lista é um espaço.)

É cômodo usar atalhos verbais ao falar de caracteres ASCII. Por exemplo, em vez de dizer “o caractere A” podemos dizer “o caractere 65”, pois o byte que corresponde a A na [tabela ASCII](#) vale 65 em [notação binária](#).

Além dos noventa e cinco caracteres “normais”, o alfabeto ASCII contém trinta e três caracteres “especiais”, conhecidos como *caracteres de controle*. Esses caracteres não são símbolos tipográficos como os outros e por isso precisam ser indicados por uma notação especial: uma barra invertida seguida de um dígito ou letra. Eis os caracteres de controle mais comuns:

byte	caractere
00000000	\0 <i>caractere nulo (null)</i>
00001001	\t <i>tabulação horizontal (tab)</i>
00001010	\n <i>fim de linha (newline)</i>
00001011	\v <i>tabulação vertical</i>
00001100	\f <i>fim de página (new page)</i>
00001101	\r <i>carriage return</i>

O caractere `\0` é usado para marcar o fim de uma [string](#) e não ocupa espaço algum ao ser exibido; o caractere `\n` marca o fim de uma linha de texto e produz uma mudança de linha ao ser exibido; o caractere `\f` marca o fim de uma página; e assim por diante. Embora o espaço não seja um caractere de controle, podemos usar a notação `\`  (barra invertida seguida de um espaço) para torná-lo mais visível.

Os caracteres `\` , `\t`, `\n`, `\v`, `\f` e `\r` são conhecidos como *brancos* (= *white-spaces*).

O alfabeto ASCII não contém letras com [sinais diacríticos](#), como é, Á, ã, ç, etc. Cada uma dessas e muitas outras letras é representada por 2 ou mais bytes num esquema de codificação conhecido como UTF-8. Trataremos disso na página *Unicode e UTF-8*.

(Este conteúdo foi obtido e adaptado de <https://www.ime.usp.br/~pf/algoritmos/index.html>)

Última atualização: sexta, 3 Set 2021, 17:02

[<](#) [Aula 7 - vídeo](#)

Seguir para... ▾

[Codificação Unicode e UTF-8](#) [▶](#)

Manter contato

<https://www.rj.senac.br>



Obter o aplicativo para dispositivos móveis