

# Codificação de Fonte

Teoria da Informação - AULA 08  
Prof<sup>a</sup>. Verusca Severo

Universidade de Pernambuco  
Escola Politécnica de Pernambuco

21 de julho de 2021

## Objetivos:

- Estudar códigos de fonte.
- Como codificar uma fonte digital de informação que emite símbolos de um dado alfabeto.

- Consideremos o problema da codificação de símbolos de um alfabeto estendido, em que esses símbolos são estatisticamente independentes. Consideremos um alfabeto de 37 símbolos equiprováveis:

26 letras  $\Rightarrow$  a, b, c, ..., z  
1 espaço em branco  $\Rightarrow$  " "  
10 algarismos  $\Rightarrow$  0, 1, 2, ..., 9

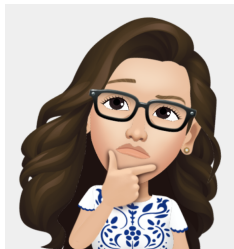
- Suponhamos que para codificar estes símbolos dispomos apenas de símbolos binários (*bits*).

# Introdução - Contextualização

- Precisamos usar **pelo menos**

$$32 = 2^5 < 37 \text{ símbolos} < 2^6 = 64$$

- 6 *bits* por símbolo para codificar os 37 símbolos.
- Qual o problema dessa estratégia de codificação?



# Introdução - Contextualização

- Qual o problema dessa estratégia de codificação?
  - Ela resulta em **desperdício**, uma vez que sobram:

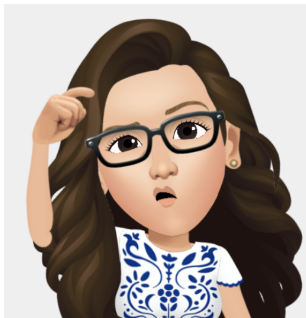
$$2^6 - 37 = 64 - 37 = 27 \text{ palavras código}^*$$

- \*que não correspondem a qualquer símbolo do alfabeto. Por exemplo:

Palavra código 1	⇒ 000000	⇒ símbolo "a"
⋮	⋮	⋮
Palavra código 26	⇒ 011010	⇒ símbolo "z"
Palavra código 27	⇒ 011011	⇒ símbolo " "
Palavra código 28	⇒ 011100	⇒ símbolo "0"
⋮	⋮	⋮
Palavra código 37	⇒ 100101	⇒ símbolo "9"
Palavra código 38	⇒ 100110	⇒ nenhum símbolo
⋮	⋮	⋮
Palavra código 64	⇒ 111111	⇒ nenhum símbolo

# Introdução - Contextualização

- **Pergunta:** Essa estratégia de codificação apresentada para os 37 símbolos é?
  - a) Eficiente ( )
  - b) Eficaz ( )



- **Pergunta:** Essa estratégia de codificação apresentada para os 37 símbolos é?
  - a Eficiente ( )
    - A eficiência consiste no ato de “fazer certo as coisas”
    - fazer da melhor forma possível o que foi demandado.
  - b Eficaz ( )
    - A eficácia consiste em “fazer as coisas certas”
    - saber o que fazer.

- **Pergunta:** Essa estratégia de codificação apresentada para os 37 símbolos é?

a Eficiente ( )

- A eficiência consiste no ato de “fazer certo as coisas”
- fazer da melhor forma possível o que foi demandado.

b Eficaz ( X )

- A eficácia consiste em “fazer as coisas certas”
- saber o que fazer.

Logo, esse código é somente eficaz!



- Como vimos, no exemplo anterior, a eficiência da codificação está associada à economia que se usa na escolha dos comprimentos das palavras código (no exemplo anterior, o comprimento é igual a 6).
- Além disso, para uma dada taxa de geração de símbolos, quanto maior for o comprimento das palavras código, maiores serão as necessidades em termos de largura de banda para transmissão e em termos de memória para armazenamento.
- Como, em geral, os símbolos de uma fonte não são equiprováveis, é razoável pensar em códigos de comprimento variável, permitindo com que símbolos que ocorrem mais vezes (com maior probabilidade de ocorrência) recebam palavras código de menor tamanho.

- O que significa **codificar**?



- O que significa **codificar**?
- Vejamos e julguemos alguns exemplos:

Símbolo	Código A	Código B	Código C	Código D
$s_1$	00	0	0	00
$s_2$	11	1	10	01
$s_3$	11	01	11	10

- O que significa **codificar**?
  - Vejamos e julgemos alguns exemplos:

Símbolo	Código A	Código B	Código C	Código D
$s_1$	00	0	0	00
$s_2$	11	1	10	01
$s_3$	11	01	11	10

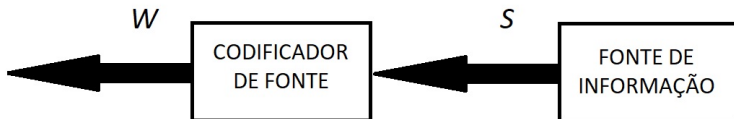
- **Código A:** incorreto! (atribuição de uma mesma palavra código para mais de um símbolo - não há "exclusividade")
- **Código B:** incorreto! (a estrutura do código não permite a decodificação correta - haverá incerteza na decodificação entre os símbolos)
- **Código C:** correto!
- **Código D:** correto!

- O que significa **codificar**?
  - consiste em atribuir uma palavra código ÚNICA a cada um dos símbolos gerados por uma fonte de informação.
- O problema inverso da codificação tem de ter solução única, isto é, qualquer código de fonte tem de ser **UNIVOCAMENTE DECODIFICÁVEL**.
  - Como identificar se um código é UNIVOCAMENTE DECODIFICÁVEL?

- Como identificar se o código é UNIVOCAMENTE DECODIFICÁVEL?
  - 1 Para **códigos de comprimento fixo** (palavras código de mesmo tamanho):
    - basta que cada palavra código corresponda a um e somente um símbolo (ou mensagem).
  - 2 Para **códigos de comprimento variável** (palavras código de tamanho variado):
    - além de que cada palavra código corresponda a um e somente um símbolo (ou mensagem),
    - é necessário também que a estrutura do código permita identificar sem ambiguidade o início e o fim de cada palavra código.

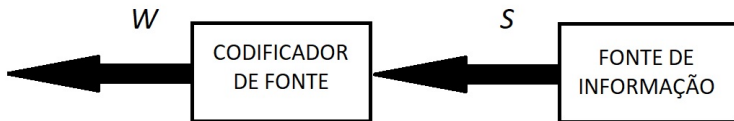
# Codificação de Fonte - comprimento médio de um código

- Vamos usar uma abordagem mais modesta para a codificação de fonte, considerando como é possível fazer a codificação eficiente de uma única variável aleatória.
- Imaginemos uma situação conceitual, na qual uma variável aleatória  $S$  é codificada em dígitos  $D$ -ários.



# Codificação de Fonte - comprimento médio de um código

- Imaginemos uma situação conceitual, na qual uma variável aleatória  $S$  é codificada em dígitos  $D$ -ários.



- $S$  denota a variável aleatória que assume os símbolos do alfabeto  $S = \{s_1, s_2, \dots, s_K\}$
- A designação “**um código para  $S$** ” é usada para significar uma lista  $(w_1, w_2, \dots, w_K)$  de sequências  $D$ -árias nas quais  $w_i$  denota a palavra código para  $s_i$ , isto é:

$$S = s_i \Leftrightarrow W = w_i$$



# Codificação de Fonte - comprimento médio de um código

- O comprimento médio, denotado por  $L$ , das palavras código será usado como um indicador de qualidade do código.
  - Quanto menor for  $L$ , melhor será o código.
- Se  $w_i = [x_{i1}, x_{i2}, \dots, x_{il_i}]$  for a palavra código para  $s_i$  e  $l_i$  for o comprimento desta palavra, então é possível escrever o comprimento médio destas palavras (dos  $w_i$ 's que compõem o código) como

$$L = \sum_{i=1}^K l_i P(s_i),$$

em que  $P(s_i)$  corresponde a probabilidade da fonte emitir o símbolo  $s_i$  (probabilidade com que o símbolo  $s_i$  ocorre).

- **Exemplo 1:** Seja uma fonte discreta sem memória que emite símbolos do alfabeto  $S = \{s_1, s_2, s_3, s_4\}$  com distribuição de probabilidades  $P(S = s_1) = \frac{1}{2}$ ,  $P(S = s_2) = \frac{1}{4}$ ,  $P(S = s_3) = P(S = s_4) = \frac{1}{8}$ . Calcule o comprimento médio de cada um dos códigos apresentados na tabela abaixo e determine aquele que apresenta o menor valor.

Símbolo	Código A	Código B	Código C	Código D
$s_1$	00	0	0	0
$s_2$	01	1	01	10
$s_3$	10	10	011	110
$s_4$	11	11	0111	111

- **Solução (Exemplo 1):**

$$L = \sum_{i=1}^4 l_i P(s_i) = l_1 P(s_1) + l_2 P(s_2) + l_3 P(s_3) + l_4 P(s_4)$$

- **Código A:**  $L_A = 2 \times \frac{1}{2} + 2 \times \frac{1}{4} + 2 \times \frac{1}{8} + 2 \times \frac{1}{8} = 2$
- **Código B:**  $L_B = 1 \times \frac{1}{2} + 1 \times \frac{1}{4} + 2 \times \frac{1}{8} + 2 \times \frac{1}{8} = 1,25$
- **Código C:**  $L_C = 1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} + 4 \times \frac{1}{8} = 1,875$
- **Código D:**  $L_D = 1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} + 3 \times \frac{1}{8} = 1,75$

O **Código B** é o código que apresenta o menor comprimento médio!

# Codificação de Fonte - comprimento médio de um código

- **Exemplo 2:** Qual dos códigos do **Exemplo 1** é mais eficiente, em termos de comprimento médio?

Símbolo	Código A	Código B	Código C	Código D
$s_1$	00	0	0	0
$s_2$	01	1	01	10
$s_3$	10	10	011	110
$s_4$	11	11	0111	111

- **Solução (Exemplo 2):**

O **Código D** é o código mais eficiente!

- **Código A:** Não é o mais eficiente, pois apresenta o maior comprimento médio entre os códigos apresentados.
- **Código B:** Não é o mais eficiente, pois mesmo apresentando o menor comprimento médio, ele não é univocamente decodificável.
- **Código C:** Não é o mais eficiente, pois apresenta o segundo maior valor de comprimento médio entre os códigos apresentados.
- **Código D:** É o mais eficiente, pois é univocamente decodificável e apresenta o menor comprimento médio entre os códigos univocamente decodificáveis apresentados.

- Códigos univocamente decodificáveis, basicamente, são aqueles em que para cada palavra código, a sequência de dígitos que a compõe é diferente da sequência de dígitos que compõe qualquer outra palavra código.
- Existe uma classe especial de códigos univocamente decodificáveis, os **Códigos Prefixo**.
- **Código Prefixo** é um código no qual nenhuma palavra código é prefixa de qualquer outra.

- **Exemplo 2:** Identifique quais os códigos apresentados na tabela abaixo são Códigos Prefixo.

Símbolo	Código A	Código B	Código C	Código D
$s_1$	00	0	0	0
$s_2$	01	1	01	10
$s_3$	10	10	011	110
$s_4$	11	11	0111	111

- **Exemplo 2:** Identifique quais os códigos apresentados na tabela abaixo são Códigos Prefixo.

Símbolo	Código A	Código B	Código C	Código D
$s_1$	00	0	0	0
$s_2$	01	1	01	10
$s_3$	10	10	011	110
$s_4$	11	11	0111	111

- **Solução:**

Somente os **Código A** e **Código D** são Código Prefixo!



- Os Códigos Prefixo são sempre univocamente decodificáveis. Além disso, eles apresentam a característica de serem códigos **decodificáveis instantaneamente**.
- Em um código **decodificável instantaneamente**, a palavra código é reconhecida (decodificada) assim que seu último dígito for recebido.
- Ou seja, você detecta (o decodificador detecta) a palavra código sem precisar do dígito da próxima palavra código.

- Exemplos de códigos **decodificáveis instantaneamente**:

Símbolo	Código A	Código B	Código C
$s_1$	0	1	1
$s_2$	01	01	01
$s_3$	011	001	000
$s_4$	0111	0001	001

- Código A** não é decodificável instantaneamente.
- Os **Código B** e **Código C** são decodificáveis instantaneamente.