

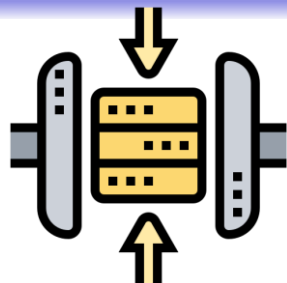
Algoritmos e Estruturas de Dados III

Aula 9 – Compressão de Dados

Prof. Felipe Lara

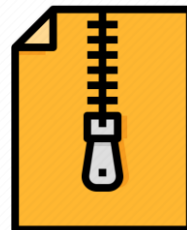


Roteiro do Conteúdo



Compressão de Dados

- Introdução
- Classificação
- Simetria, perda e adaptabilidade
- Métricas



Tipos de Compressão de Dados

- Compressão com e sem perdas
- Redução de quantidade e tamanho de símbolos
- Codificação RLE
- Métodos Estatísticos
- Métodos de Dicionário

Compressão de Dados

Introdução



Compressão de Dados - Introdução

Objetivo

Codificar um conjunto de informações de modo que o código resultante seja **menor** que o original

Justificativa

Redução do espaço ocupado

Aumento da velocidade de transmissão dos dados

Compressão de Dados - Introdução

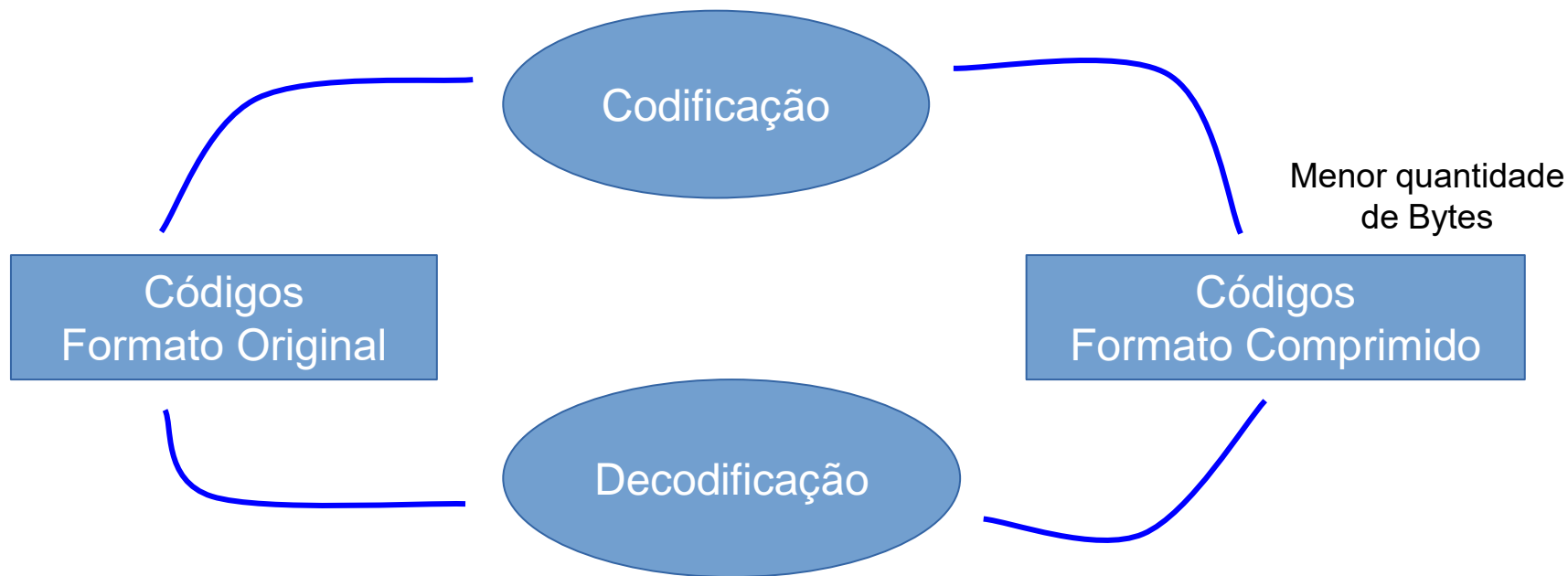
Conceito

Compressão é um processo de codificação de dados que busca **reduzir o número de bits** necessários para se representar uma informação.

A compressão de dados envolve dois processos:

- Em um deles os dados são comprimidos (codificados) para terem seu tamanho reduzido
- No segundo processo eles são descomprimidos (decodificados) para voltar ao formato original

Compressão de Dados - Introdução



Exemplo

"AAAAAAAAAABBBBBCCD"

(são 10 letras A, 5 letras B, 2 letras C e 1 letra D, totalizando 18 letras)

Se $A = 00$, $B = 01$, $C = 10$ e $D = 11$,

Qual a compactação resultante?

Quantos bytes serão usados?

Qual a economia?

Exemplo

"AAAAAAAAAABBBBBCCD"

(são 10 letras A, 5 letras B, 2 letras C e 1 letra D, totalizando 18 letras)

Se $A = 0$, $B = 10$, $C = 110$ e $D = 111$

Qual a compactação resultante?

Quantos bytes serão usados?

Qual a economia?

Exemplo

"AAAAAAAAAABBBBBCCD"

(são 10 letras A, 5 letras B, 2 letras C e 1 letra D, totalizando 18 letras)

Se $A = 00$, $B = 01$, $C = 10$ e $D = 11$, então a mensagem completa gastará $18 \cdot 2 = 36$ bits.

Se $A = 0$, $B = 10$, $C = 110$ e $D = 111$, então a mensagem completa gastará $10 \cdot 1 + 5 \cdot 2 + 2 \cdot 3 + 1 \cdot 3 = 29$ bits.

Antes da Compactação



Compressão de Dados - Introdução

Antes da Compressão – Racionalização da representação

- Eliminação de itens redundantes
- Uso de Notação Codificada
- Codificação de textos
- Supressão de espaços inúteis

Compressão de Dados - Introdução

Antes da Compressão – Racionalização da representação

- **Eliminação de itens redundantes**

- Uso de Notação Codificada
- Codificação de textos
- Supressão de espaços inúteis

Aluno {Nome, Matrícula, nota1, nota2, nota3, média}

=

Aluno {Nome, Matrícula, nota1, nota2, nota3}

Compressão de Dados - Introdução

Antes da Compressão – Racionalização da representação

- Eliminação de itens redundantes
- **Uso de Notação Codificada**
- Codificação de textos
- Supressão de espaços inúteis

17 de Abril de 2001 \Rightarrow 19 B

17/Abril/2001 \Rightarrow 13 B

17/04/01 \Rightarrow 3 B (Notação Binária)

00010001 00000100 00000001

Bits realmente utilizados : 10001 0100 0000001 \Rightarrow 2 B

Compressão de Dados - Introdução

Antes da Compressão – Racionalização da representação

- Eliminação de itens redundantes
- Uso de Notação Codificada
- **Codificação de textos**
- Supressão de espaços inúteis

Diminuição da ocorrência de textos em tabelas

Funcionário {CódF, Nome, **NomeDept**, DataNasc}

Equipamento {CódE, Nome, Desc, **NomeDept**}

Funcionário {CódF, Nome, **CodDept**, DataNasc}

Equipamento {CódE, Nome, Desc, **CodDept**}

Departamento {**CodDept**, **NomeDept**}

Compressão de Dados - Introdução

Antes da Compressão – Racionalização da representação

- Eliminação de itens redundantes
- Uso de Notação Codificada
- Codificação de textos
- **Supressão de espaços inúteis**

Campos com tam. fixo devem ser avaliados com muito cuidado.

Ex. Nome: 70 caracteres (70 bytes)
(José da Silva⇒ 13 bytes)

Classificação da Compactação de Dados



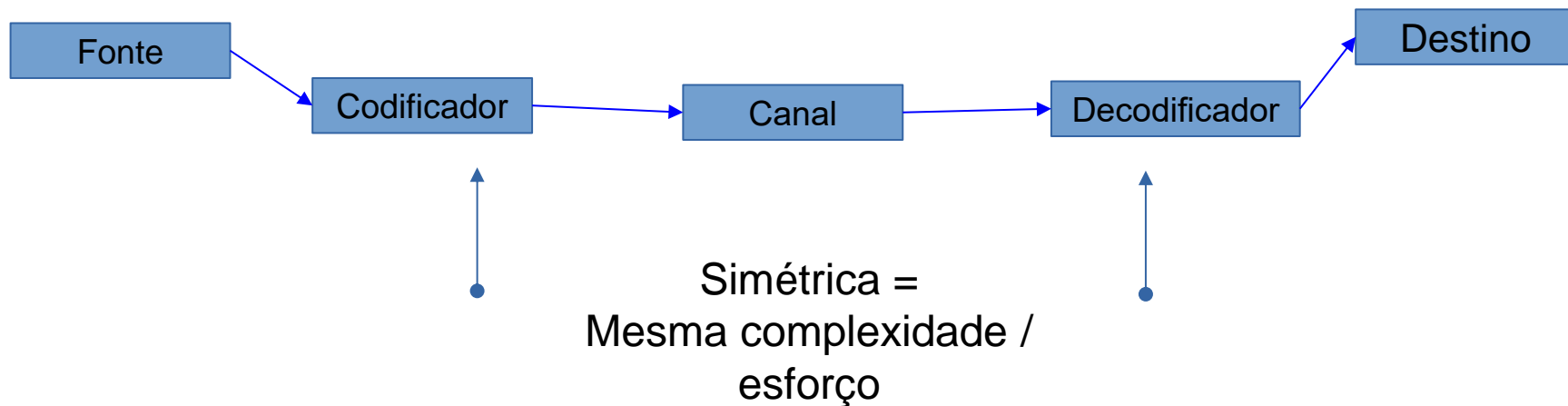
Compressão de Dados - Classificação

- Quanto a Simetria
- Quanto a Perda
- Quanto a Adaptabilidade

Compressão de Dados - Classificação

- **Quanto a Simetria**
- Quanto a Perda
- Quanto a Adaptabilidade

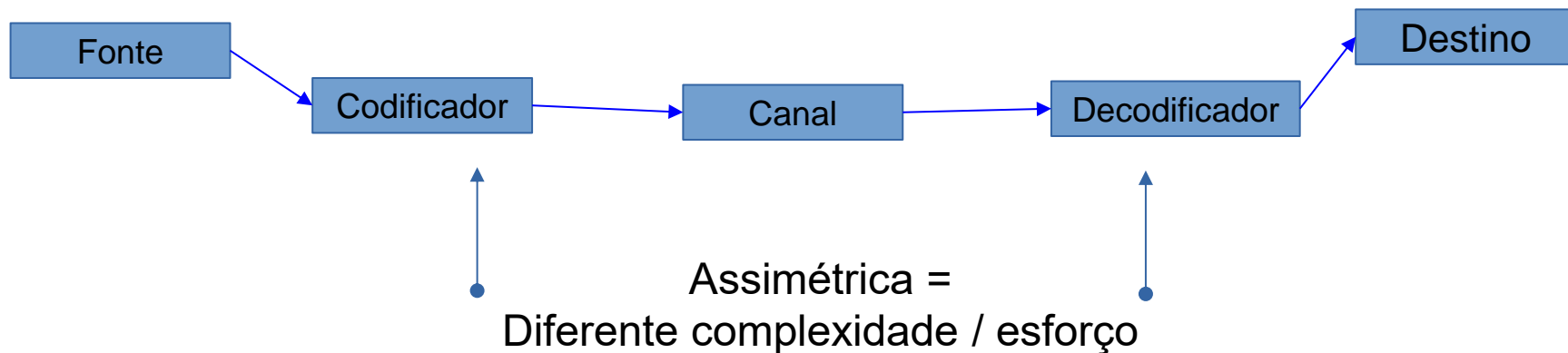
Procedimentos simétricos são indicados em aplicações que envolvem a transmissão e apresentação das imagens simultaneamente



Compressão de Dados - Classificação

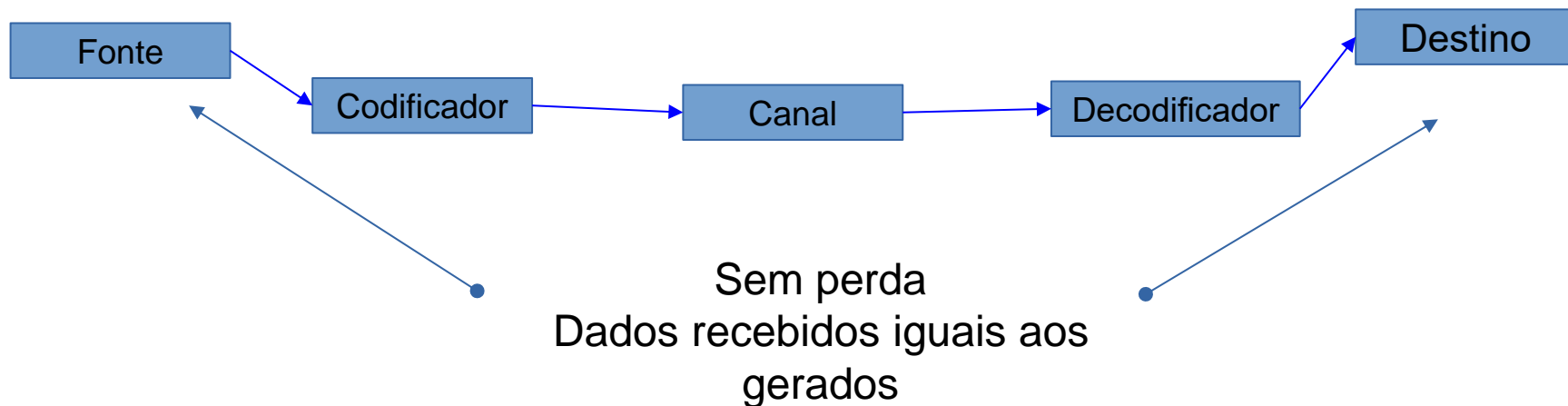
- Quanto a Simetria
- Quanto a Perda
- Quanto a Adaptabilidade

Procedimentos de backup tendem a privilegiar uma compressão mais rápida em detrimento do tempo necessário à descompressão



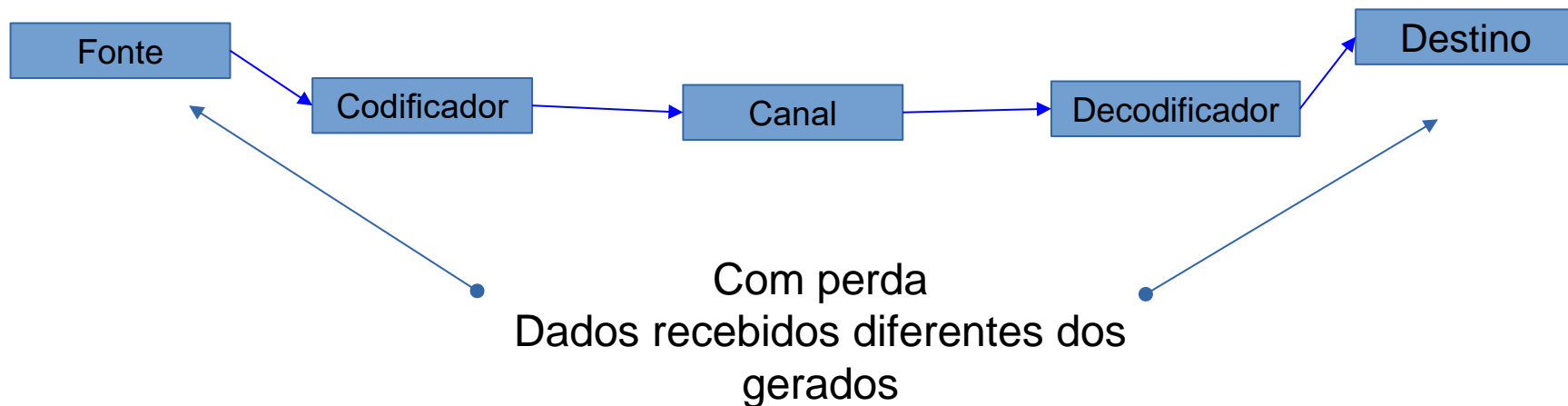
Compressão de Dados - Classificação

- Quanto a Simetria
- **Quanto a Perda**
- Quanto a Adaptabilidade



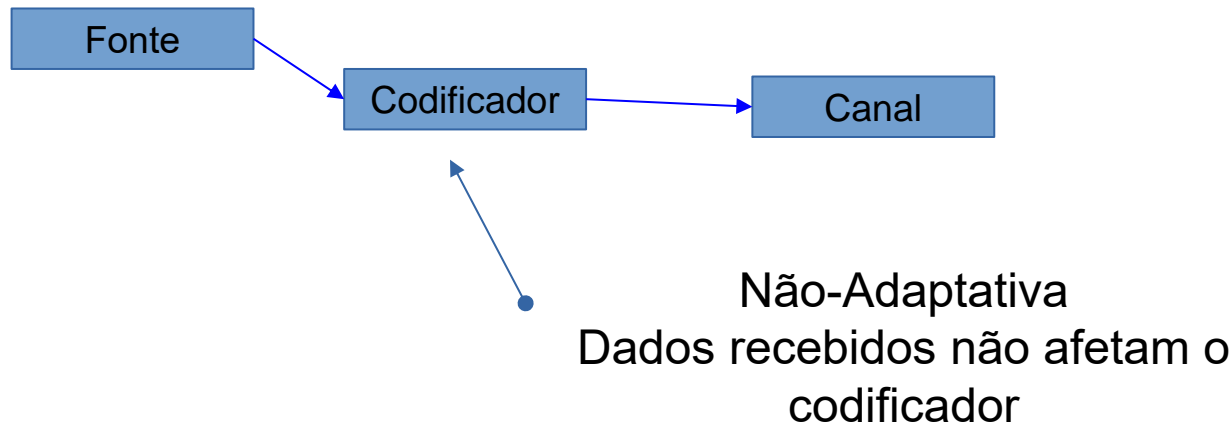
Compressão de Dados - Classificação

- Quanto a Simetria
- **Quanto a Perda**
- Quanto a Adaptabilidade



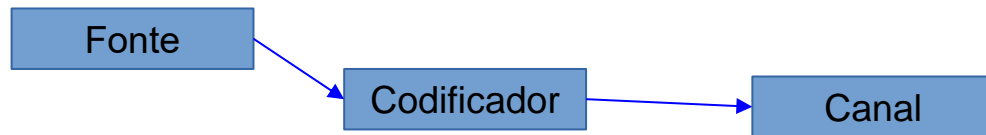
Compressão de Dados - Classificação

- Quanto a Simetria
- Quanto a Perda
- **Quanto a Adaptabilidade**



Compressão de Dados - Classificação

- Quanto a Simetria
- Quanto a Perda
- **Quanto a Adaptabilidade**



Adaptativa
Dados recebidos afetam o
codificador dinamicamente

Compressão de Dados - Métricas

- Taxa de Compressão
- Fator de Compressão
- Percentual de Redução
- Taxa de Bits

Compressão de Dados - Métricas

- **Taxa de Compressão**
- Fator de Compressão
- Percentual de Redução
- Taxa de Bits

$$T_c = \text{Tamanho Final} \div \text{Tamanho Inicial}$$

Compressão de Dados - Métricas

- Taxa de Compressão
- **Fator de Compressão**
- Percentual de Redução
- Taxa de Bits

$$Fc = \text{Tamanho Inicial} \div \text{Tamanho Final}$$

Compressão de Dados - Métricas

- Taxa de Compressão
- Fator de Compressão
- **Percentual de Redução**
- Taxa de Bits

$$Pr = 100 \times (1 - T_c)$$

Compressão de Dados - Métricas

- Taxa de Compressão
- Fator de Compressão
- Percentual de Redução
- **Taxa de Bits**

bpc \equiv bits por character
bpp \equiv bits por pixel

Tipos de Compressão de Dados

Compressão sem Perdas x Com Perdas

Sem Perdas



- Permite a recuperação exata dos dados originais após o processo de descompressão
- Remoção (recuperável) das redundâncias
- Aplicada na compressão de dados, textos, programas, imagens médicas ...
- Exemplos: ZIP

Com Perdas

- Informação obtida após a descompressão é diferente da original (antes da compressão)
- Informação suficientemente "parecida" para que seja de alguma forma útil.
- Eliminação de detalhes
- Aplicada na compressão de imagens, áudio, vídeo, ...
- Exemplos: JPEG, MP3, MP4

Compressão - Quantidade e Tamanho dos Símbolos

Redução da Quantidade de Símbolos



- Um símbolo passa a representar um conjunto de outros símbolos
- Ex.:
 - Ao invés de indexarmos cada letra, indexamos palavras
 - Um pixel pode representar um conjunto de pixels

Redução do Tamanho dos Símbolos

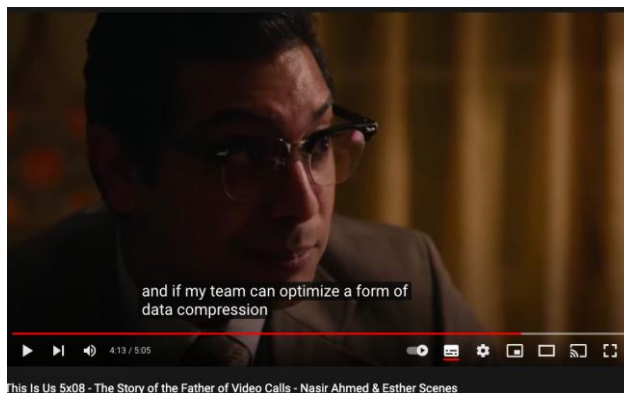
- Um símbolo pode ser representado com menos bits do que o usual
- Ex.:
 - Podemos usar menos de 1 byte para representar uma letra
 - Um pixel pode usar menos de 3 bytes

Compressão com Perdas

- Transformada Discreta de Cosseno
- Desenvolvida em 1974 por N. Ahmed, T. Natarajan and K. R. Rao
- Uma das ferramentas mais usadas em processamento de imagens
- <https://ieeexplore.ieee.org/document/1672377/>

Compressão com Perdas

- Em plena Pandemia
- Com vídeo-chamadas em grande uso
- Uma homenagem aos que permitiram este feito



<https://www.youtube.com/watch?v=W29r-zJtqcY>

Exemplos e Métodos de Compressão de Dados

Codificação RLE

- RLE = Run-length Encoding
- Compressão sem perda de dados
- Sequências longas de valores repetidos são armazenadas como um único valor e sua contagem no lugar de sua sequência original.
- Útil em dados com muitas repetições de valores
 - `aaaaabbbbbbbbbbccccdddeeeee`
 - `5a10b4c3d6e`

Métodos Estatísticos

- Utilizam códigos de comprimentos variáveis.
- Dados na informação original que aparecem com maior frequência são representados por palavras-código menores
- Dados de menor incidência são representados por palavras-código maiores
- Ex: Shannon-Fano / Huffman

Métodos de Dicionário

- Os símbolos (ou conjunto de símbolos) são substituídos por códigos a partir de um “dicionário”
- Os códigos possuem tamanho fixo
- Os dicionários podem ser estáticos ou dinâmicos
- Ex: LZ77 / LZ78 / LZW