

Algoritmos e Estruturas de Dados III

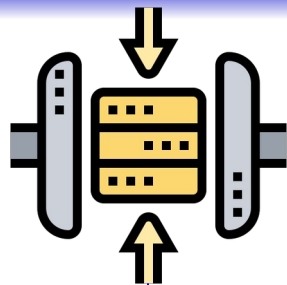
Aula 8.1 – Compressão de Dados

Prof. Hayala Curto
2022



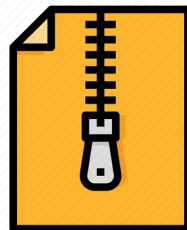
PUC Minas

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

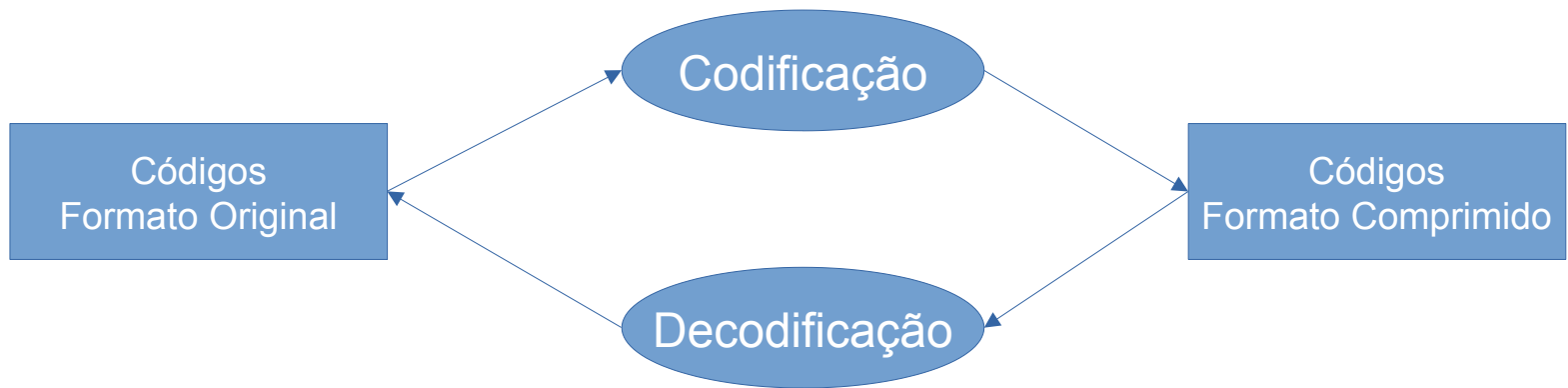
Justificativa

Redução do espaço ocupado

Aumento da velocidade de transmissão dos dados



Compressão da Dados - Introdução



Compressão de Dados - Introdução

Antes da Compressão – Racionalização da representação

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

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

17 de Abril de 2001 \Rightarrow 19 bytes

17/Abril/2001 \Rightarrow 13 bytes

17/04/2001 \Rightarrow 10 bytes

17/04/01 \Rightarrow 8 bytes

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

00010001 00000100 00000001

Bits realmente utilizados : 10001 0100 0000001 \Rightarrow 2 bytes

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 {Nome, CódF, NomeDept, DataNasc}

Equipamento {Nome, CódE, Desc, NomeDept}

Funcionário {Nome, CódF, CodDept, DataNasc}

Equipamento {Nome, CódE, 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
(José da Silva \Rightarrow 13 bytes)

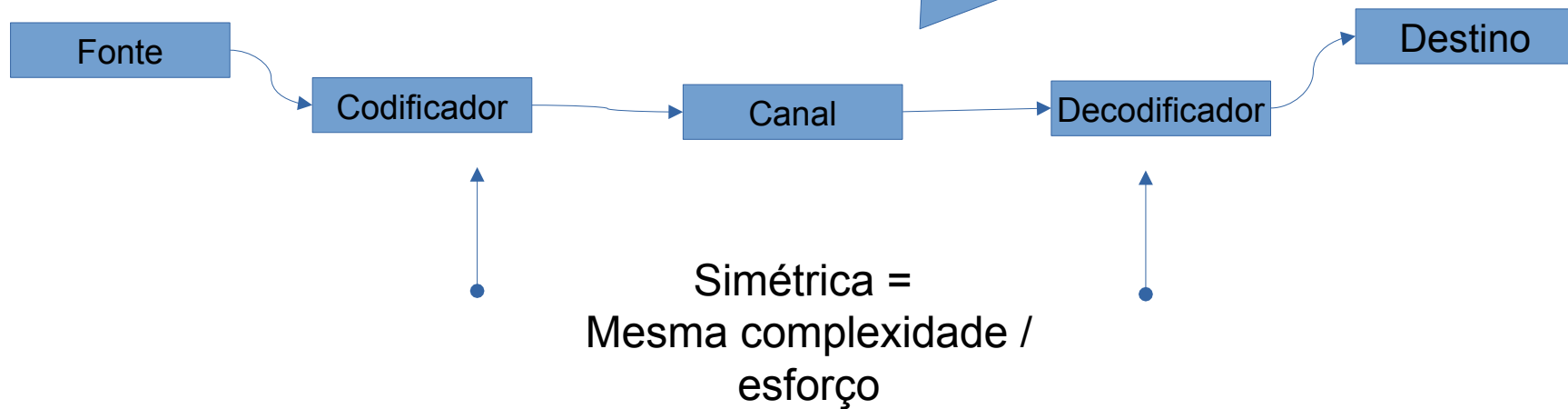
Compressão de Dados - Classificação

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

Compressão de Dados - Classificação

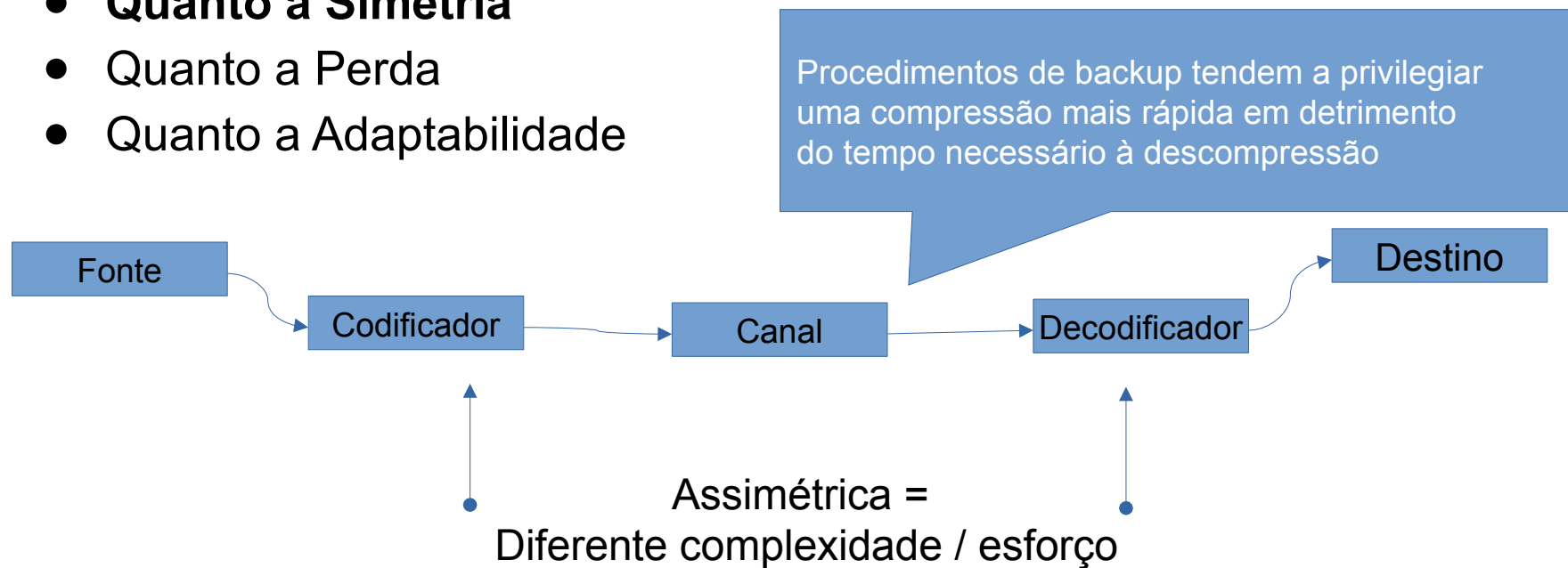
- Quanto a Simetria
- Quanto a Perda
- Quanto a Adaptabilidade

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



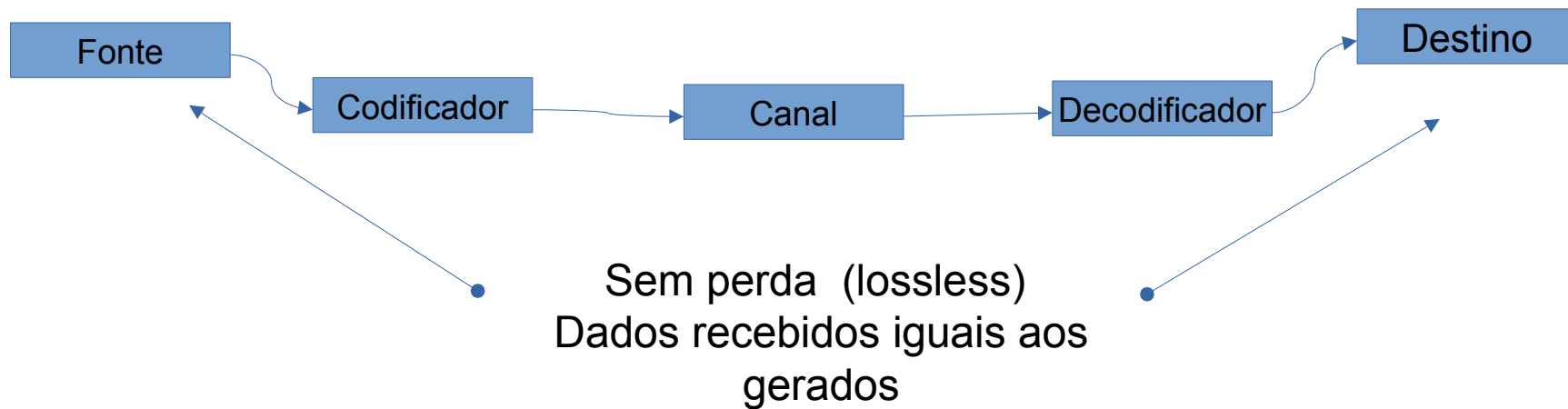
Compressão de Dados - Classificação

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



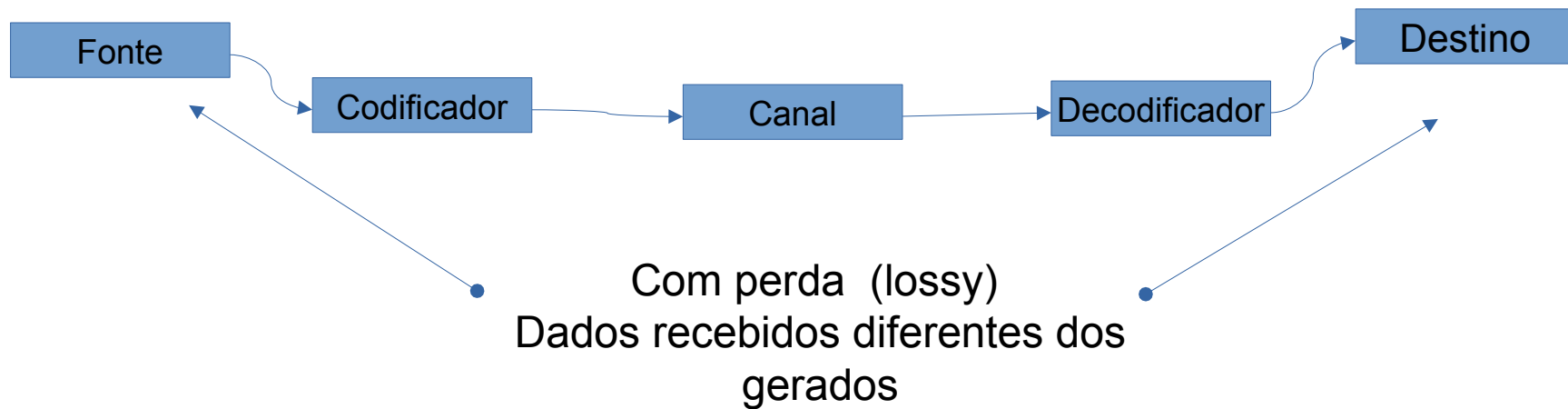
Compressão de Dados - Classificação

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



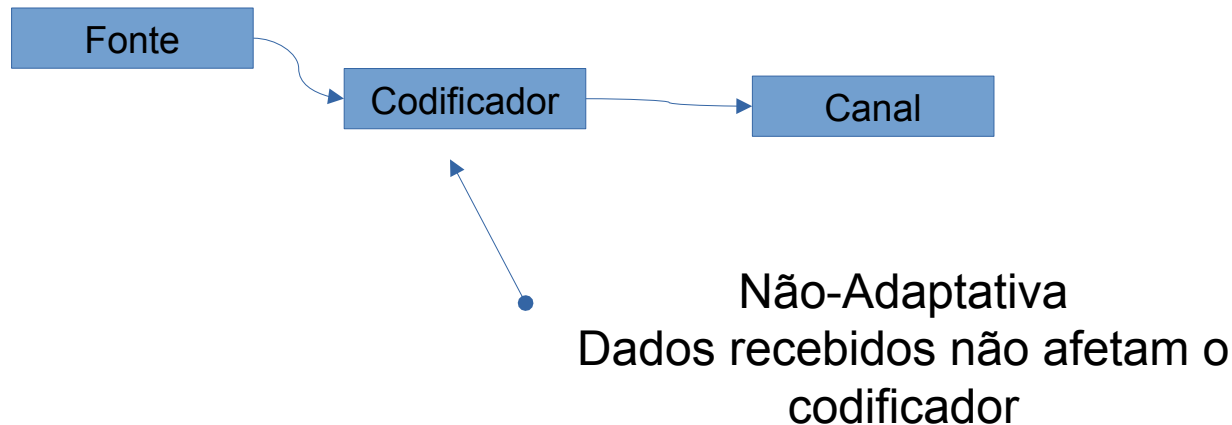
Compressão de Dados - Classificação

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



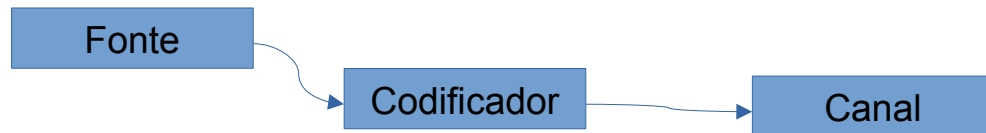
Compressão de Dados - Classificação

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



Compressão de Dados - Classificação

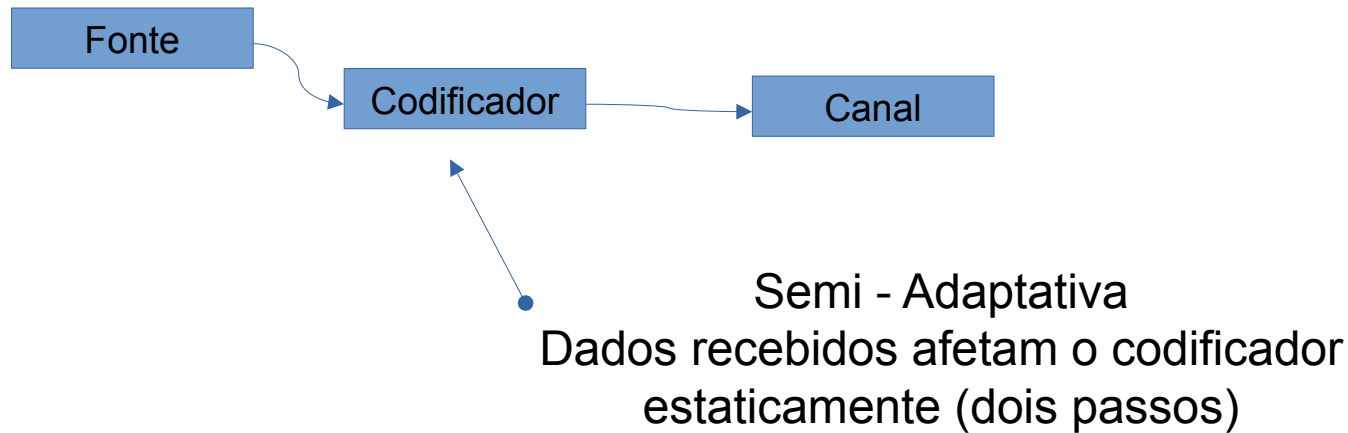
- Quanto a Simetria
- Quanto a Perda
- **Quanto a Adaptabilidade**



Adaptativa
Dados recebidos afetam o
codificador dinamicamente

Compressão de Dados - Classificação

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



Compressão de Dados - Métricas

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

Compressão de Dados - Métricas

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

$$Gc = 100 \times \log_e (\text{Tamanho de referência} \div \text{Tamanho Final})$$

Compressão de Dados - Métricas

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

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

Compressão de Dados - Métricas

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

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

Tipos de Compressão de Dados

Compressão sem Perdas

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

Compressão com Perdas

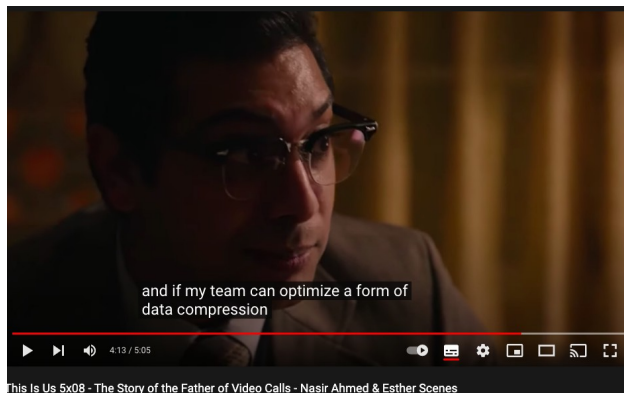
- Informação obtida após a descompressão é diferente da original (antes da compressão)
- Informação suficientemente "parecida" para que seja de alguma forma útil.
- Eliminação de detalhes
- Aplicada a compressão de imagens, áudio, vídeo, ...
- Exemplos: JPEG, MP3, MP4

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 video-chamadas em grande uso
- Uma homenagem aos que permitiram este feito



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

Compressão - 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

Compressão - Símbolos

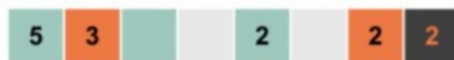
- Redução do Tamanho dos Símbolos
- Um símbolo pode ser representado com menos bits do que o usual
- Ex.:
- Podemos usar menos de 1 byte para representar uma letra
- Um pixel pode usar menos de 3 bytes

Codificação 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`

Codificação RLE

Lossless pixel compression



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

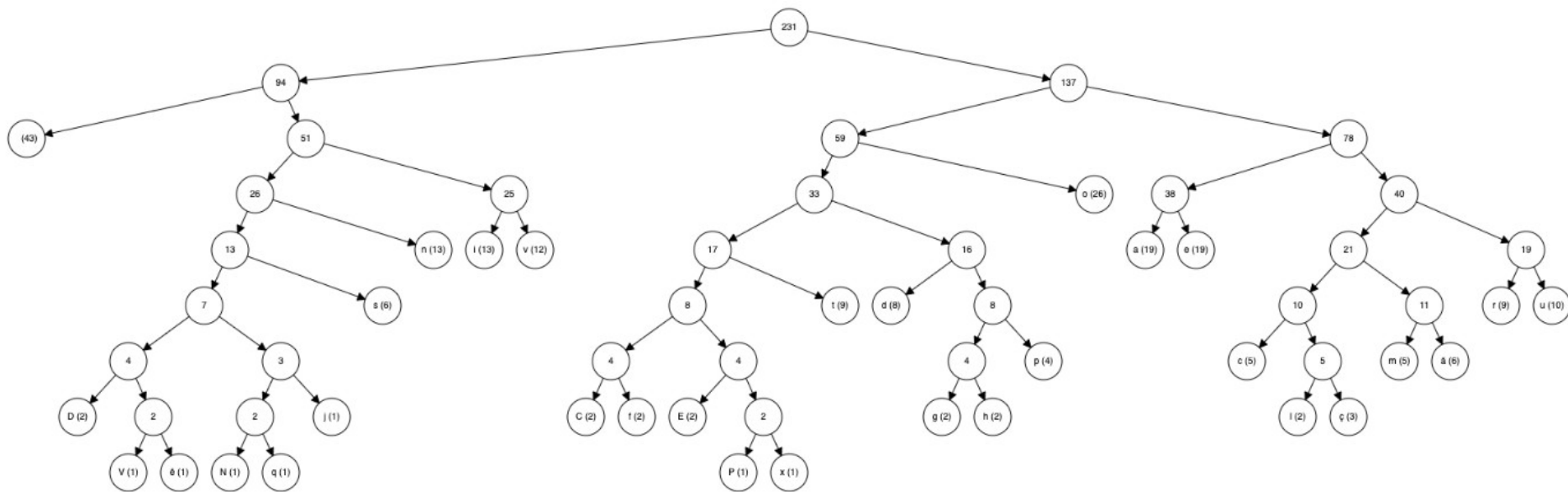
<https://cmps-people.ok.ubc.ca/ylucet/DS/Huffman.html>

Métodos Estatísticos - Huffman

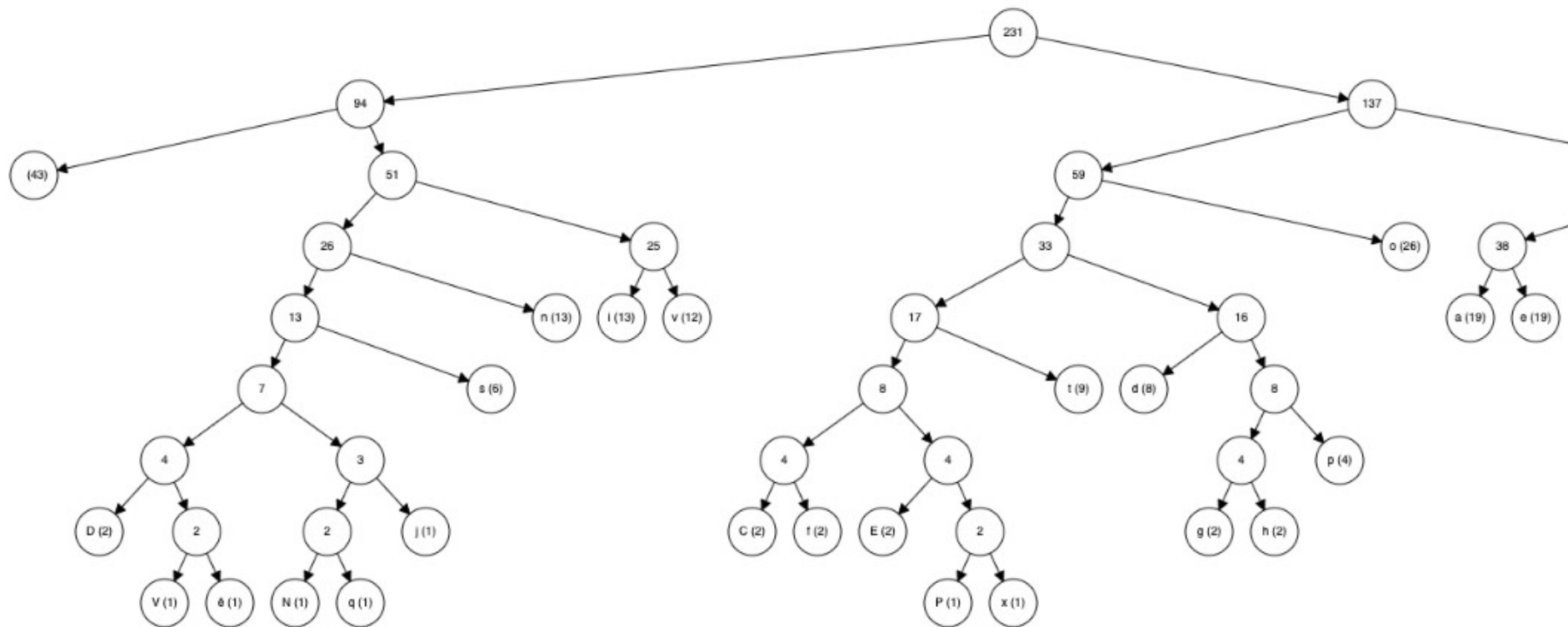
“Você me pergunta pela minha paixão
Digo que estou encantada
Como uma nova invenção
Eu vou ficar nesta cidade
Não vou voltar pro sertão
Pois vejo vir vindo no vento
Cheiro de nova estação
Eu sei de tudo na ferida viva do meu coração...”



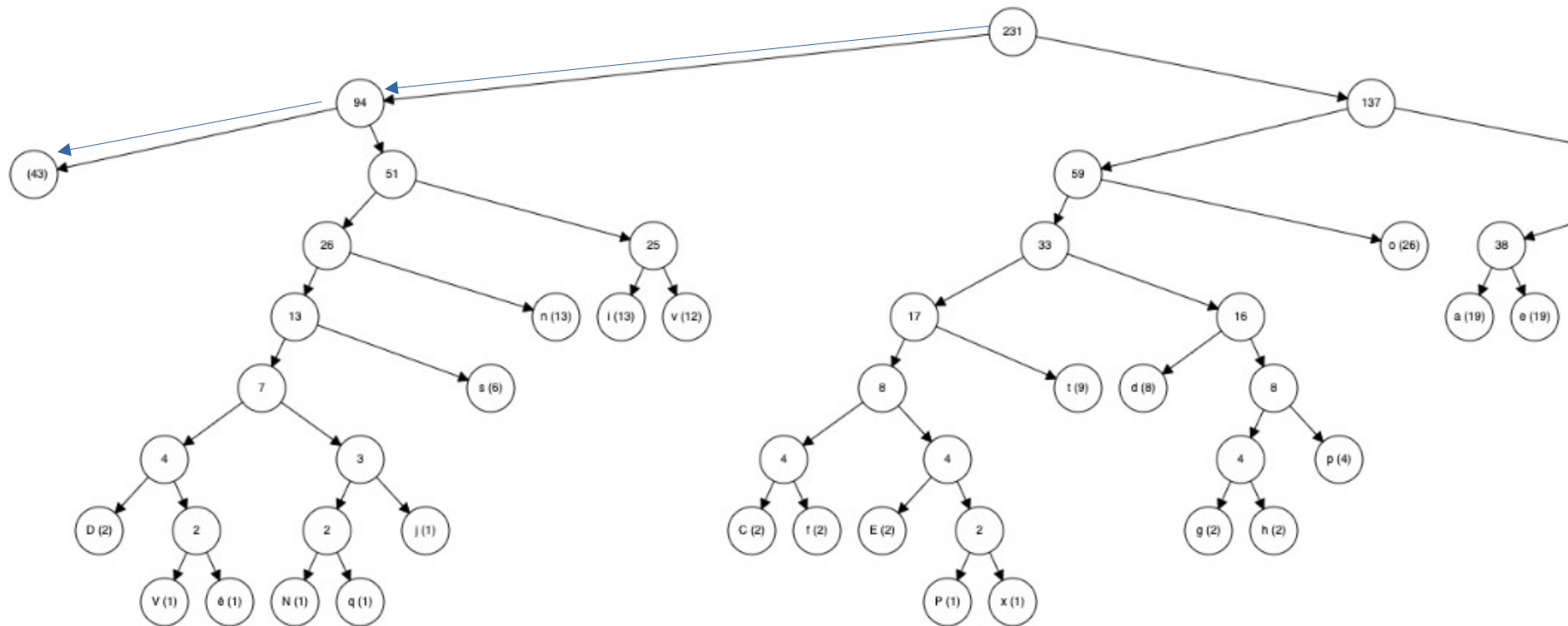
Métodos Estatísticos - Huffman



Métodos Estatísticos - Huffman

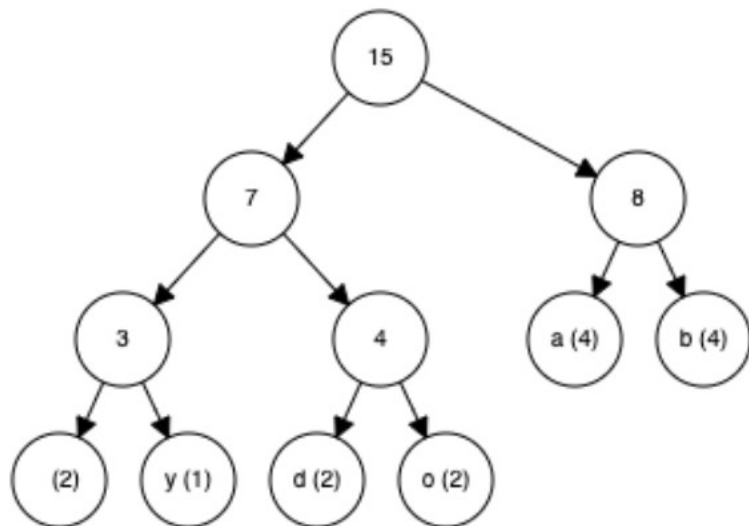


Métodos Estatísticos - Huffman



Métodos Estatísticos - Huffman

yabba dabba doo



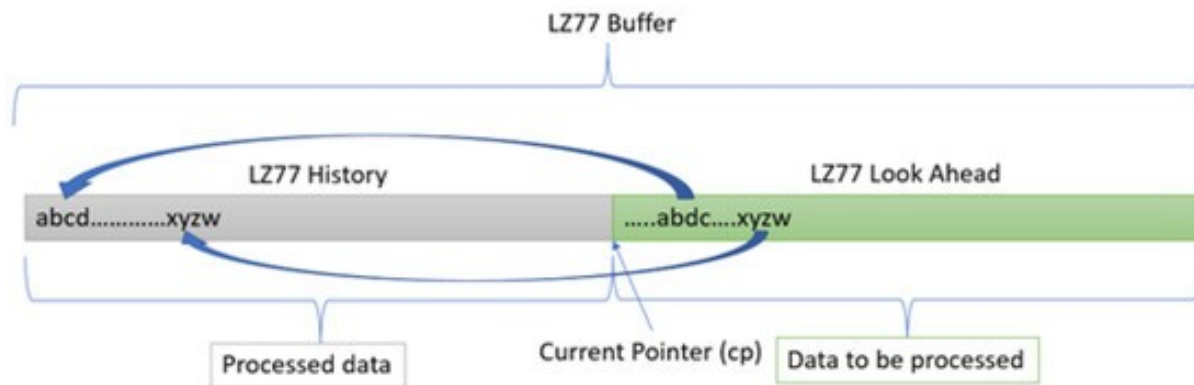
000	
001	y
010	d
011	o
10	a
11	b

001 10 11 11 10 000 010 10 11 11 10 000 010 011 011

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

Métodos de Dicionário - LZ77



Compressão de Dados – Sistemas de Arquivos

File system	Internal snapshotting / branching	encryption	deduplication	Data checksum/ ECC	Persistent Cache	Multiple Devices	compression
BeeGFS	No	No	Yes	No	No	No	Yes
APFS	Yes	Yes	Yes ^[51]	No	No	No	Yes
NTFS	No	Yes	Yes ^{[ca][53]}	No	No	No	Yes
F2FS	No	Yes, experimental ^[56]	No	No	No	No	Yes
Reiser4	?	Yes ^[cc]	?	No	No	No	Yes
NSS	Yes	Yes	?	No	No	No	Yes
NWFS	?	No	?	No	No	No	Yes
Fossil	Yes	No	Yes	No	No	No	Yes
ZFS	Yes	Yes ^[ce]	Yes	Yes	Yes	Yes	Yes
Btrfs	Yes	No	Yes	Yes ^[cf]	No	Yes	Yes
SquashFS	No	No	Yes	Yes	No	No	Yes
BlueStore/Cephfs	Yes	No	No	Yes	Yes	Yes	Yes