

Carlos Matos

2020245868

Mariana Magueijo

2020246886

João Pino

2020210945

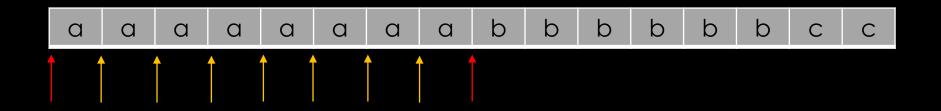
INTRODUÇÃO

- A compressão consiste na utilização de um conjunto de métodos e outros pormenores práticos com o intuito da redução do espaço armazenado em unidades de memória secundária ou mesmo primária de um sistema computacional.
- A necessidade de compressão de dados é algo comumente relacionado à vida das pessoas, seja de maneira direta ou indireta.

INTRODUÇÃO

- Na internet, a compressão ajuda a diminuir a quantidade de tráfego na grande rede, aumentando a velocidade de navegação, realização de downloads de arquivos e visualização de vídeos.
- Na vida offline, arquivos compactados são preferíveis quando há interesse de armazenamento de maior número de dados possível no menor espaço de memória secundária disponível, como em pen-drives, memmory cards, discos rígidos e demais unidades de memória.

RUN LENGTH ENCODING (RLE)



Codificação:

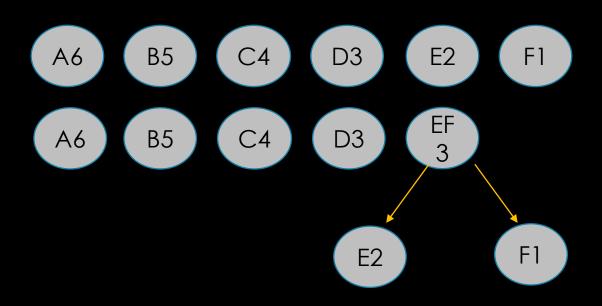
8 a # 6 b c c

SHANNON FANO

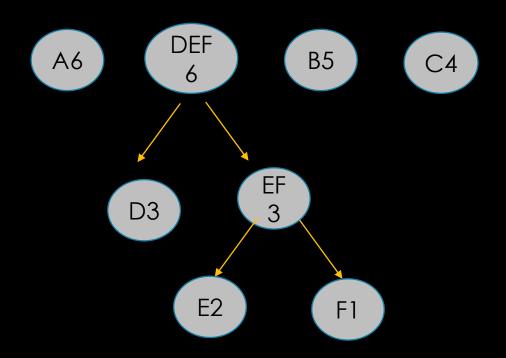
Caracteres	Α	В	С	D	E	F
Contagem	6	5	4	3	2	1

Caracteres	Α	В	С	D	E	F	
Frequência	6	5	4	3	2	1	
	0				1		
					1 1		
	0	1	0	0			
				U	0	1	
Códigos	00	01	10	110	1110	1111	

HUFFMAN ENCODING



HUFFMAN ENCODING



ADEFBC 21 BC ADEF 12 DEF B5 C4 **A6** EF D3 E2

HUFFMAN ENCODING

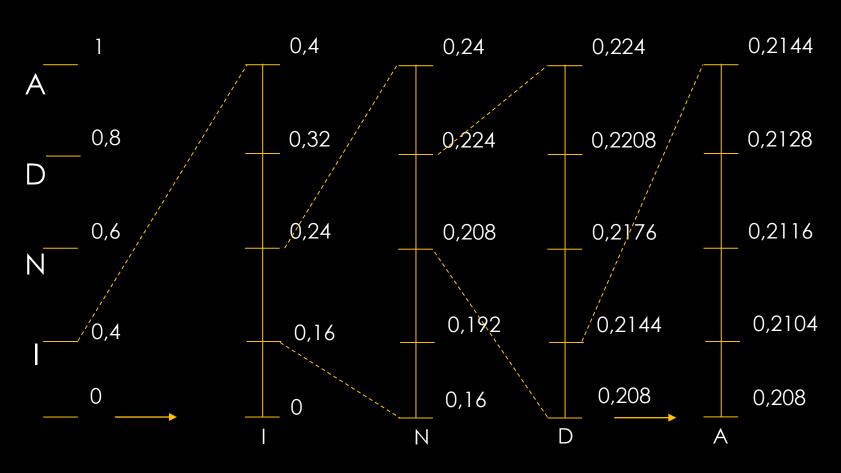
Caracteres	Α	В	С	D	Е	F
Código	10	00	01	110	1110	1111

ARITHMETIC ENCODING

Mensagem: INDIA

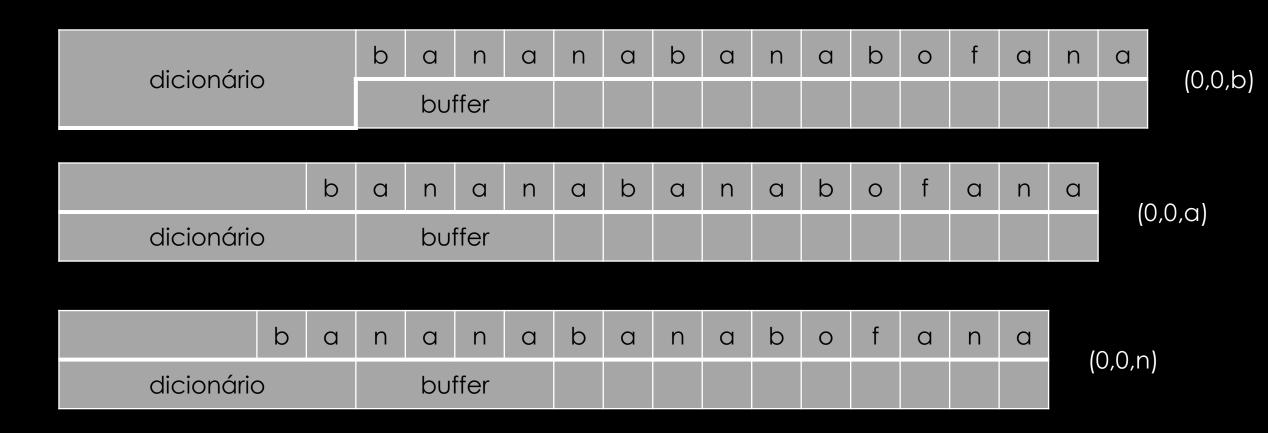
Caracter	Prob	CDF
I	2/5 = 0.4	0,4
Ν	1/5 = 0.2	0,6
D	1/5 = 0.2	0,8
Α	1/5 = 0.2	1

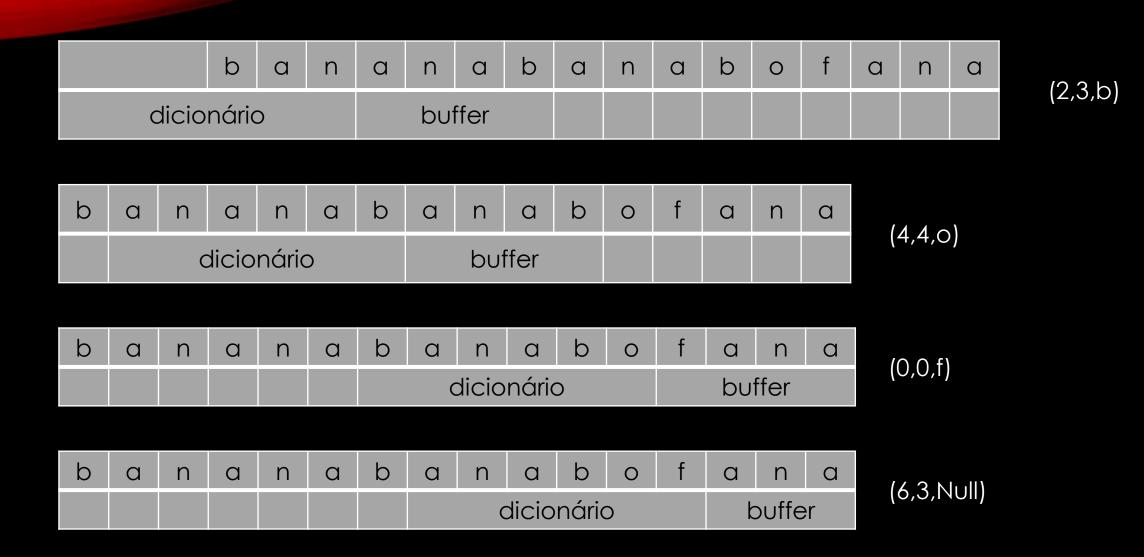
ARITHMETIC ENCODING



$$X = LI + (LS - LI) * PC$$

LZ77





LZW

String = "LAILAALAALAALAA"

Dicionário de arranque = {1-A, 2-I, 3-L}

S	С	Saída	Código	Sequência
-	-	-	1	Α
-	-	-	2	I
-	-	-	3	L

S	С	Saída	Código	Sequência
-	-	-	1	Α
-	-	-	2	1
-	-	-	3	L
L	Α	3	4	LA
Α	1	1	5	Al
	L	2	6	IL
L	Α	-	-	-
LA	Α	4	7	LAA
Α	L	1	8	AL
L	Α	-	-	-
LA	Α	-	-	-
LAA	L	7	9	LAAL
L	Α	-	-	-

LZW

S = LAILAALAALAALAA

TRANSFORMADA MOVE-TO-FRON (MTF)

Input	Output	Tabela de símbolos
-	-	abcdefghijklmnopqrstuvwxyz
U	20,	uabcdefghijklmnopqrstvwxyz
un	20,14,	nuabcdefghijkImopqrstvwxyz
uni	20,14,10,	inuabcdefghjklmopqrstvwxyz
univ	20,14,10,21,	vinuabcdefghjklmopqrstwxyz
unive	20,14,10,21,8,	evinuabcdfghjklmopqrstwxyz
univer	20,14,10,21,8,19,	revinuabcdfghjklmopqstwxyz
univers	20,14,10,21,8,19,20,	srevinuabcdfghjklmopqtwxyz
universi	20,14,10,21,8,19,20,4,	isrevnuabcdfghjklmopqtwxyz

TRANSFORMADA DE BURROWS-WHEELER (BWT)

```
i. ← A C C E L E R A T E ←
ii. C C E L E R A T E A
```

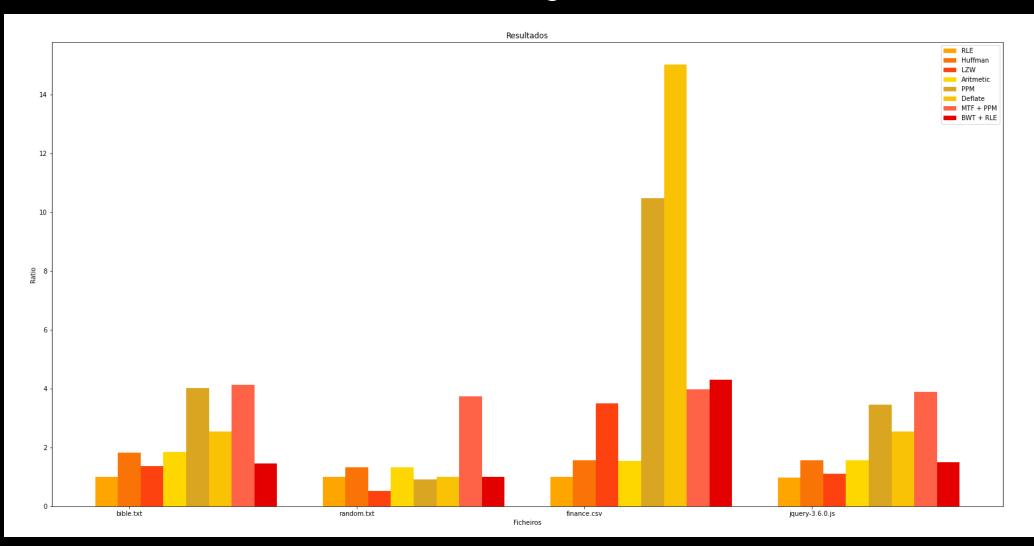
	0	1	2	3	4	5	6	7	8	9
0	Α	С	С	Е	L	Е	R	Α	Т	Е
1	С	С	Е	L	Е	R	Α	Т	Е	Α
2	С	Е	L	Е	R	Α	Т	Е	Α	С
3	Е	L	Е	R	Α	Т	Е	Α	С	С
4	L	Е	R	Α	Т	Е	Α	С	С	Е
5	Е	R	Α	Т	Е	Α	С	С	Е	L
6	R	Α	Т	Е	Α	С	С	Е	L	Е
7	Α	Т	Е	Α	С	С	Е	L	Е	R
8	T	Е	Α	С	С	Е	L	Е	R	Α
9	Е	Α	С	С	Е	L	Е	R	Α	Т

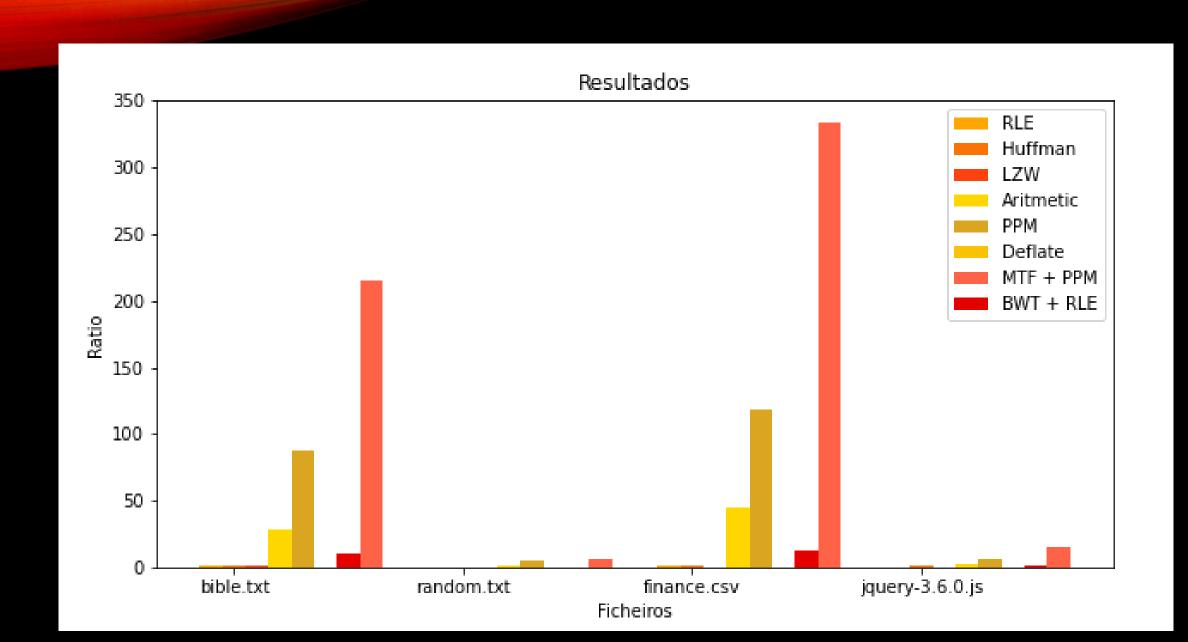
TRANSFORMADA DE BURROWS-WHEELER (BWT)

	0	1	2	3	4	5	6	7	8	9
0	Α	С	С	Е	L	Е	R	Α	Т	Е
1	Α	Т	Е	Α	С	С	Е	L	Е	R
2	С	С	Е	L	Е	R	Α	Т	Е	Α
3	С	Е	L	Е	R	Α	Т	Е	Α	С
4	Е	Α	С	С	Е	L	Е	R	Α	Т
5	Е	L	Е	R	Α	Т	Е	Α	С	С
6	Е	R	Α	Т	Е	Α	С	С	Е	L
7	L	Е	R	Α	Т	Е	Α	С	С	Е
8	R	Α	Т	Е	Α	С	С	Е	L	Е
9	T	Е	Α	С	С	Е	L	Е	R	Α

R

APRESENTAÇÃO DE RESULTADOS





```
import numpy as np
ratio RLE = [1, 1, 1, 1]
ratio huffman = [1.824424181, 1.333297779, 1.5512563, 1.568565636]
ratio lzw = [1.364579066, 0.520034322, 3.504529726, 1.115936257]
ratio aritmetic = [1.841289569, 1.315443304, 1.543219793, 1.570033459]
ratio_ppm = [4.016863835, 0.914151987, 10.46781649, 3.446019369]
ratio deflate = [2.546855847, 0.990373569, 15.01532149, 2.542555066]
ratio MTF PPM = [4.131364009, 3.727186922, 3.973788972, 3.87686516]
ratio_BWT_RLE = [1.452477769, 1.000009999, 4.293432308, 1.492352321]
barWidth = 0.1
plt.figure(figsize = (10, 5))
tempo RLE = [1.26353883, 0.03387904, 1.82790899, 0.09184503]
tempo huffman = [0.95987796, 0.0284574, 1.51949, 0.67985439]
tempo lzw = [0.8240661, 0.0221047, 0.0221047, 0.04380488]
tempo aritmetic = [28.454261, 0.8381381, 45.585161, 2.24748301]
tempo_ppm = [87.878869, 5.534121, 118.840321, 6.99088811]
tempo deflate = [0.14804697, 0.00349307, 0.03987812, 0.009869]
tempo MTF PPM = [214.617506, 6.11745357, 333.403995, 15.812921]
tempo BWT RLE = [9.74310994, 0.24683928, 13.2666859, 0.70176959]
rl = np.arange(len(ratio_RLE))
r2 = [x + barWidth for x in r1]
r3 = [x + barWidth for x in r2]
r4 = [x + barWidth for x in r3]
r5 = [x + barWidth for x in r4]
r6 = [x + barWidth for x in r5]
r7 = [x + barWidth for x in r6]
r8 = [x + barWidth for x in r7]
```

```
plt.bar(rl,ratio RLE, color = '#FFA500', width = barWidth, label = 'RLE')
plt.bar(r2,ratio huffman, color = '#F97306', width = barWidth, label = 'Huffman')
plt.bar(r3,ratio lzw, color = '#FE420F', width = barWidth, label = 'LZW')
plt.bar(r4,ratio aritmetic, color = '#FFD700', width = barWidth, label = 'Aritmetic')
plt.bar(r5,ratio ppm, color = '#DAA520', width = barWidth, label = 'PPM')
plt.bar(r6,ratio_deflate, color = '#FAC205', width = barWidth, label = 'Deflate')
plt.bar(r7,ratio MTF PPM, color = '#FF6347', width = barWidth, label = 'MTF + PPM')
plt.bar(r8,ratio_BWT_RLE, color = '#E50000', width = barWidth, label = 'BWT + RLE')
plt.bar(r1,tempo RLE, color = '#FFA500', width = barWidth, label = 'RLE')
plt.bar(r2,tempo huffman, color = '#F97306', width = barWidth, label = 'Huffman')
plt.bar(r3,tempo lzw, color = '#FE420F', width = barWidth, label = 'LZW')
plt.bar(r4,tempo aritmetic, color = '#FFD700', width = barWidth, label = 'Aritmetic')
plt.bar(r5,tempo ppm, color = '#DAA520', width = barWidth, label = 'PPM')
plt.bar(r6,tempo deflate, color = '#FAC205', width = barWidth, label = 'Deflate')
plt.bar(r7,tempo MTF PPM, color = '#FF6347', width = barWidth, label = 'MTF + PPM')
plt.bar(r8,tempo_BWT_RLE, color = '#E50000', width = barWidth, label = 'BWT + RLE')
Ficheiros = ['bible.txt', 'random.txt', 'finance.csv', 'jquery-3.6.0.js']
plt.xlabel('Ficheiros')
plt.xticks([r+barWidth for r in range(len(ratio RLE))], Ficheiros)
plt.ylabel('Ratio')
plt.title('Resultados')
plt.legend()
plt.show()
```

CONCLUSÃO

- Concluindo, tomamos consciência da importância dos algoritmos de compressão para a sociedade da informação, incluindo as suas implementações para compensar diversas facetas que a computação tem.
- O texto comprimido ocupa menos espaço de armazenamento ou seja, menos tempo para ser lido do disco ou ser transmitido por um canal de comunicação e para ser pesquisado.