

# Algoritmos e Estruturas de Dados III

Aula 9.4 – Métodos de Dicionário  
LZW

Prof. Felipe Lara



**PUC Minas**

# Métodos de Dicionário



# 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 - Dicionário Estático

## Exemplo Simples (Dicionário Estático)

- um dicionário que possui um código para as palavras da língua portuguesa. Para 500.000 palavras, seriam necessários 19 bits para cada código ( $2^{19} = 524.288$ )

# Métodos de Dicionário - Dicionário Estático

## Exemplo Simples (Dicionário Estático)

- um dicionário que possui um código para as palavras da língua portuguesa. Para 500.000 palavras, seriam necessários 19 bits para cada código ( $2^{19} = 524.288$ )
- o algoritmo irá substituir cada palavra do arquivo pelo seu código. Caso a palavra não existir no dicionário, ela é colocada sem substituição no arquivo de saída
- o arquivo compactado irá conter códigos e palavras

# Métodos de Dicionário - Dicionário Estático

## Exemplo Simples (Dicionário Estático)

- um dicionário que possui um código para as palavras da língua portuguesa. Para 500.000 palavras, seriam necessários 19 bits para cada código ( $2^{19} = 524.288$ )
- o algoritmo irá substituir cada palavra do arquivo pelo seu código. Caso a palavra não existir no dicionário, ela é colocada sem substituição no arquivo de saída
- o arquivo compactado irá conter códigos e palavras

**Possível problema:** como saber diferenciar entre os códigos e as palavras (todos são conjuntos de bits)

**Solução:** usar um bit extra para informar se é um código ou não

# Métodos de Dicionário - Dicionário Estático

## Exemplo Simples (Dicionário Estático)

No caso de muitas falhas (palavras não encontradas no dicionário), o tamanho do arquivo irá aumentar !!!

# Métodos de Dicionário - Dicionário Estático

## Exemplo Simples (Dicionário Estático)

No caso de muitas falhas (palavras não encontradas no dicionário), o tamanho do arquivo irá aumentar !!!

Perguntas:

- o que aconteceria se o texto estivesse em inglês?

Esse método de dicionário estático não é bom para compactadores de uso geral mas pode adequado para compactadores específicos



# Métodos de Dicionário

- Na verdade, os algoritmos de uso geral utilizam **dicionários dinâmicos** (métodos adaptativos)
- Os dicionários **são construídos à medida que o texto vai sendo lido**. Se uma palavra não existe no dicionário ela pode ser inserida, e palavras “fora de uso” podem ser removidas do dicionário

# Métodos de Dicionário

- Na verdade, os algoritmos de uso geral utilizam **dicionários dinâmicos** (métodos adaptativos)
- Os dicionários **são construídos à medida que o texto vai sendo lido**. Se uma palavra não existe no dicionário ela pode ser inserida, e palavras “fora de uso” podem ser removidas do dicionário
- A grande maioria dos métodos de dicionários existentes hoje são extensões ou modificações dos algoritmos de Lempel-Ziv (LZ) (década de 70)
- Boa parte dos programas comerciais (zip, compress, formato gif, etc.) se baseiam em variações/combinações destes algoritmos

LZW



**PUC Minas**

# LZW

- LZW – Lempel-Ziv-Welch (1984)
- Versão melhorada do algoritmo LZ77/LZ78 de fácil implementação e que possui grande potencial para outras versões
- Equivale a utilizar um dicionário modificado que foi inicializado com todos os símbolos do alfabeto
- <https://planetcalc.com/9045/>
- <https://www.csfieldguide.org.nz/en/interactives/lzw-compression/>

# LZW Codificação



# LZW - Algoritmo

1) **Inicializar o dicionário** (com símbolos básicos).

2) Repetir até o fim do texto:

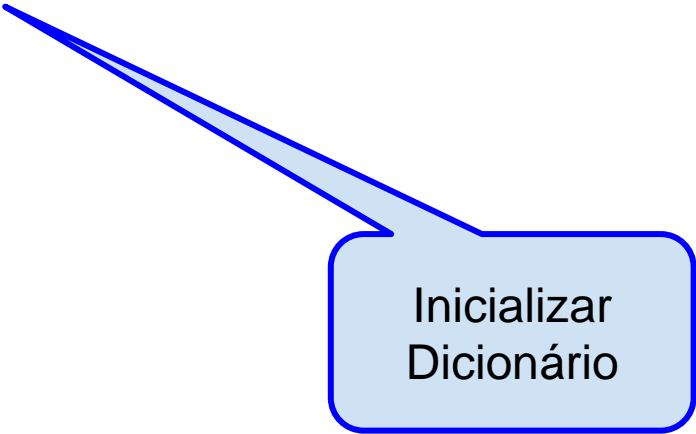
- A partir da posição atual, achar a maior string **w** existente no dicionário.
- Escrever o índice de **w** na saída.
- Olhar o próximo caráter **a** que não fez parte de **w**.
- Escrever **wa** no dicionário.
- Avançar para a posição de **a**.

Dicionário

0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

Arquivo:

w	a	b	b	a	w	a	b	b	a
---	---	---	---	---	---	---	---	---	---

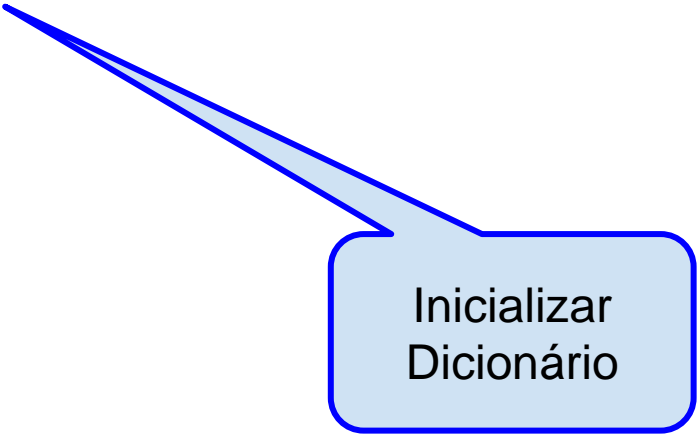


Dicionário

0	a
1	b
2	w
3	
4	
5	
6	
7	
8	
9	
10	

Arquivo:

w	a	b	b	a	w	a	b	b	a
---	---	---	---	---	---	---	---	---	---



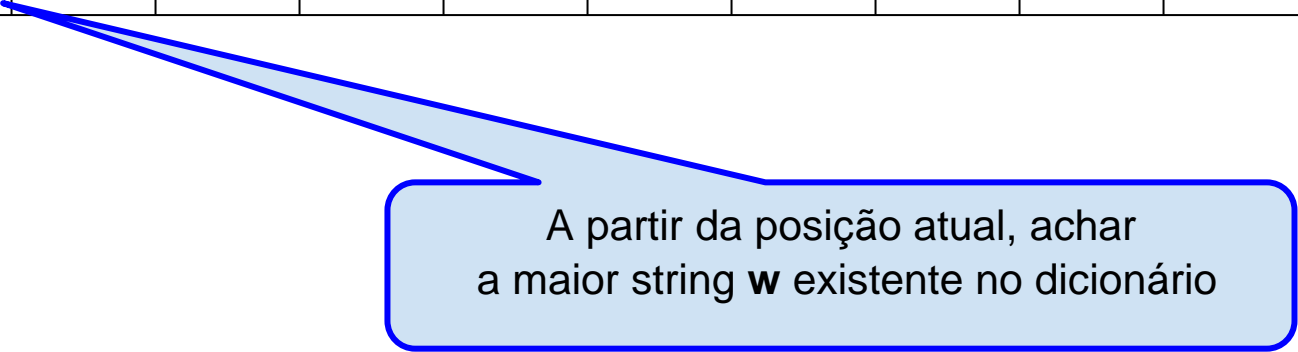


## Dicionário

0	a
1	b
2	w
3	
4	
5	
6	
7	
8	
9	
10	

Arquivo:

w	a	b	b	a	w	a	b	b	a
---	---	---	---	---	---	---	---	---	---



A partir da posição atual, achar a maior string **w** existente no dicionário

## Dicionário

0	a
1	b
<b>2</b>	<b>w</b>
3	
4	
5	
6	
7	
8	
9	
10	

Arquivo:

<b>w</b>	a	b	b	a	w	a	b	b	a
----------	---	---	---	---	---	---	---	---	---

2



Escrever o índice de **w** na saída.

## Dicionário

0	a
1	b
2	w
3	wa
4	
5	
6	
7	
8	
9	
10	

Arquivo:

w	a	b	b	a	w	a	b	b	a
---	---	---	---	---	---	---	---	---	---

2

Olhar o próximo carácter **a** que não fez parte de **w**.

Escrever **wa** no dicionário

Dicionário

0	a
1	b
2	w
3	wa
4	
5	
6	
7	
8	
9	
10	

Arquivo:

w	a	b	b	a	w	a	b	b	a
2	0								

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	
6	
7	
8	
9	
10	

Arquivo:

w	a	b	b	a	w	a	b	b	a
2	0								

Dicionário

0	a
1	<b>b</b>
2	w
3	wa
4	ab
5	
6	
7	
8	
9	
10	

Arquivo:

w	a	<b>b</b>	b	a	w	a	b	b	a
2	0	1							

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	
7	
8	
9	
10	

Arquivo:

w	a	b	b	a	w	a	b	b	a
2	0	1							

Dicionário

0	a
1	<b>b</b>
2	w
3	wa
4	ab
5	bb
6	
7	
8	
9	
10	

Arquivo:

w	a	b	<b>b</b>	a	w	a	b	b	a
2	0	1	1						



Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	
8	
9	
10	

Arquivo:

w	a	b	b	a	w	a	b	b	a
2	0	1	1						

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	
8	
9	
10	

Arquivo:

w	a	b	b	a	w	a	b	b	a
2	0	1	1	0					

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	aw
8	
9	
10	

Arquivo:

w	a	b	b	a	w	a	b	b	a
2	0	1	1	0					

Dicionário

0	a
1	b
2	w
3	<b>wa</b>
4	ab
5	bb
6	ba
7	aw
8	
9	
10	

Arquivo:

w	a	b	b	a	w	a	b	b	a
2	0	1	1	0	3				

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	aw
8	wab
9	
10	

Arquivo:

w	a	b	b	a	w	a	b	b	a
2	0	1	1	0	3				

Dicionário

0	a
1	b
2	w
3	wa
4	ab
<b>5</b>	<b>bb</b>
6	ba
7	aw
8	wab
9	
10	

Arquivo:

w	a	b	b	a	w	a	b	b	a
2	0	1	1	0		3		5	

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	aw
8	wab
9	bba
10	

Arquivo:

w	a	b	b	a	w	a	b	b	a
2	0	1	1	0		3		5	

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	aw
8	wab
9	bba
10	

Arquivo:

w	a	b	b	a	w	a	b	b	a
2	0	1	1	0	3		5		0



Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	aw
8	wab
9	bba
10	

Arquivo:

w	a	b	b	a	w	a	b	b	a
2	0	1	1	0	3		5		0

## Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	aw
8	wab
9	bba
10	

## Arquivo:

w	a	b	b	a	w	a	b	b	a
2	0	1	1	0	3		5		0
0010	0000	0001	0001	0000	0011		0101		0000

Tamanho fixo

## Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	aw
8	wab
9	bba
10	

Arquivo:

w	a	b	b	a	w	a	b	b	a
2	0	1	1	0	3		5		0
0010	0000	0001	0001	0000	0011		0101		0000

Total de bits do arquivo original:

$$10 \text{ letras} * 1\text{B} = 10 \text{ letras} * 8 \text{ b} = 80 \text{ b}$$

$$8 \text{ números de 4 bits} = 8 * 4 = 32 \text{ bits}$$

$$\text{Compressão} = 1 - 32/80 = 60\%$$

# LZW Decodificação



# LZW - Algoritmo

- Inicializar o dicionário (com símbolos básicos).
- Decodificar o 1º índice, escrevê-lo na saída e armazená-lo em **w**.
- Colocar **w?** no dicionário.
- Repetir até o fim dos índices:
  - Decodificar o primeiro símbolo **s** do próximo índice.
  - Trocar o **?** da última entrada no dicionário por **s**.
  - Decodificar o resto do índice, escrevê-lo na saída e armazená-lo em **w**.
  - Colocar **w?** no dicionário.

Dicionário

0	a
1	b
2	w
3	
4	
5	
6	
7	
8	
9	
10	

Arquivo:

0010	0000	0001	0001	0000	0011	0101	0000

Dicionário

0	a
1	b
2	w
3	
4	
5	
6	
7	
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
0010	0000	0001	0001	0000	0011	0101	0000

Dicionário

0	a
1	b
2	w
3	
4	
5	
6	
7	
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
---	---	---	---	---	---	---	---



Dicionário

0	a
1	b
2	w
3	
4	
5	
6	
7	
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
---	---	---	---	---	---	---	---

w

Dicionário

0	a
1	b
2	w
3	w?
4	
5	
6	
7	
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
---	---	---	---	---	---	---	---

w

Dicionário

0	a
1	b
2	w
3	w?
4	
5	
6	
7	
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
w	a						

Dicionário

0	a
1	b
2	w
3	wa
4	
5	
6	
7	
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
w	a						

Dicionário

0	a
1	b
2	w
3	wa
4	a?
5	
6	
7	
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
w	a						

Dicionário

0	a
1	b
2	w
3	wa
4	a?
5	
6	
7	
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
w	a	b					

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	
6	
7	
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
w	a	b					

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	b?
6	
7	
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
w	a	b					



Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	b?
6	
7	
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
w	a	b	b				

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	
7	
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
w	a	b	b				

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	b?
7	
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
w	a	b	b				

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	b?
7	
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
w	a	b	b	a			

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
w	a	b	b	a			

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	a?
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
w	a	b	b	a			

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	a?
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
w	a	b	b	a	w	a	

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	aw
8	
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
w	a	b	b	a	w	a	



Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	aw
8	wa?
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
w	a	b	b	a	w	a	

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	aw
8	wa?
9	
10	

Arquivo:

2	0	1	1	0	3	5	0
w	a	b	b	a	w	a	b

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	aw
8	wab
9	
10	

Arquivo:

2	0	1	1	0	3	5	0	
w	a	b	b	a	w	a	b	b

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	aw
8	wab
9	bb?
10	

Arquivo:

2	0	1	1	0	3	5	0	
w	a	b	b	a	w	a	b	b

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	aw
8	wab
9	bba
10	

Arquivo:

2	0	1	1	0	3		5		0
w	a	b	b	a	w	a	b	b	a

Dicionário

0	a
1	b
2	w
3	wa
4	ab
5	bb
6	ba
7	aw
8	wab
9	bba
10	a?

Arquivo:

2	0	1	1	0	3	5	0		
w	a	b	b	a	w	a	b	b	a

<https://www.geeksforgeeks.org/lzw-lempel-ziv-welch-compression-technique/>

<https://www.youtube.com/watch?v=7TqQS1sMyaY>

<https://www.youtube.com/watch?v=as3fuSWa6xs>

<https://www.youtube.com/watch?v=59Oo9b10r9A>