Algoritmos e Estruturas de Dados III

Aula 9.4 – Métodos de Dicionário LZW

Prof. Felipe Lara



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

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 (2¹⁹ = 524.288)

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 (2¹⁹ = 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

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 (2¹⁹ = 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

Exemplo Simples (Dicionário Estático)

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

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



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.

וט	Cionanio	,										
	0		Arquivo): 								
	1		w	a	b	b	а	W	а	b	b	а
	2											
	3											
	4											
	5											
	6					icializa cionário						
	7				<u> </u>	olollall.						
	8											
	9											
	10											

1010110110		- A	_								
0	а	Arquivo									
1	b	$\left[\left[\begin{array}{cc} \mathbf{w} \end{array} \right] \right]$	а	b	b	а	W	а	b	b	а
2	W										
3											
4											
5											
6					icializa cionári						
7				5,							
8		1									
9											
10											

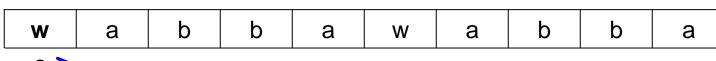
0	а
1	b
2	W
3	
4	
5	
6	
7	
8	
9	
10	

Arquivo:

			_				_	_	
l w	a	l h	l h	a	۱۸/	а	b	l h	า
**	u			u	V V	u			l a
		•	•	•	•		•	•	•

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

U	icionano	1	
	0	а	Arquivo:
	1	b	w
	2	W	2
	3		
	4		
	5		
	6		
	7		
	8		
	9		
	10		



Escrever o índice de ${\bf w}$ na saída.

ט	lcionano									
	0	а	Arquivo	:						
	1	b	W	а	b	b	а	W	а	b
	2	W	2							
	3	wa								
	4									
	5									
	6					har o n	róvimo	caráte	er a que	não
	7					παι ο ρ	/IOXIIIIC		W .	Tiao
	8						Escre	ever w a	a no dic	cionári
	9									
	10									

b

fez parte de

a

0	а
1	b
2	W
3	wa
4	
5	
6	
7	
8	
9	
10	

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

_

0	а
1	b
2	W
3	wa
4	ab
5	
6	
7	
8	
9	
10	

Arquivo:

W	а	b	b	а	W	а	b	b	а
---	---	---	---	---	---	---	---	---	---

_

0	а
1	b
2	W
3	wa
4	ab
5	
6	
7	
8	
9	
10	

w a b b a w a b b a	a w a b b	w a	а	h	h	а	\\/	
---------------------	-----------	-----	---	---	---	---	-----	--

0	а
1	b
2	W
3	wa
4	ab
5	bb
6	
7	
8	
9	
10	

Α	rq	ui	V	0
			-	_

w a b b a w a b b a

2 0

0	а
1	b
2	W
3	wa
4	ab
5	bb
6	
7	
8	
9	
10	

W	а	b	b	а	W	а	b	b	а

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

Arquivo:

W	а	b	b	а	W	а	b	b	а	
	_	_	_			•		•	•	

2 0 1

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

W	а	b	b	а	W	а	b	b	а	
_	_	_		_						

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

W	а	b	b	а	W	а	b	b	а
2	0	1	1	0					

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

W	а	b	b	а	W	а	b	b	а
2	0	1	1	0	3	3			

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

W	а	b	b	а	W	а	b	b	а
2	0	1	1	0	3	3			

0	а
1	b
2	W
3	wa
4	ab
5	bb
6	ba
7	aw
8	wab
9	
10	
I	

W	а	b	b	а	W	а	b	b	а
2	0	1	1	0	3	3	5	5	

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

W	а	b	b	а	W	а	b	b	а
2	0	1	1	0		3	5	5	

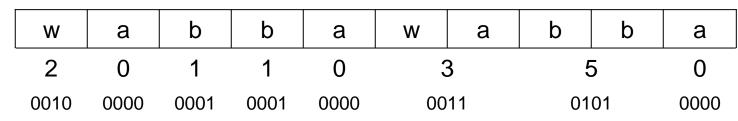
0	а
1	b
2	W
3	wa
4	ab
5	bb
6	ba
7	aw
8	wab
9	bba
10	
i	I

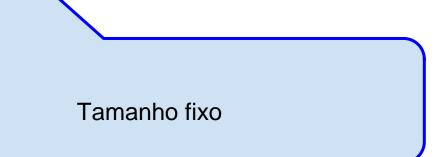
W	а	b	b	а	W	а	b	b	а
2	0	1	1	0	3	3	5	5	0

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

W	а	b	b	а	W	а	b	b	а
2	0	1	1	0	3	3	Ţ.	5	0

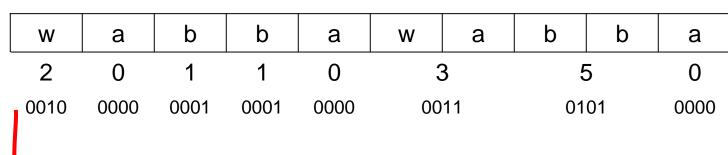
0	а
1	b
2	W
3	wa
4	ab
5	bb
6	ba
7	aw
8	wab
9	bba
10	





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

Arquivo:



Total de bits do arquivo original: 10 letras * 1B = 10 letras * 8 b = 80 b

8 números de 4 bits = 8 * 4 = 32 bits

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.

Dicioná	ário								
0	а]	Arquivo):					
1	b								
2	W		0010	0000	0001	0001	0000	0011	0101
3									
4		-							
5		-							
6		•							
7									
8									
9									

Dicioná	ırio							
0	а		Arquivo):				
1	b	•	2	0	1	1	0	3
2	W		0010	0000	0001	0001	0000	0011
3								
4								
5								
6								
7								
8								
9								

0	а
1	b
2	W
3	
4	
5	
6	
7	
8	
9	

10

Arquivo:

2	0	1	1	0	3	5	0

0	а
1	b
2	W
3	
4	
5	
6	
7	
8	
9	

10

Λ	•
$\Delta r \cap$	\cup
\neg IU	luivo

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

W

0	а
1	b
2	W
3	w?
4	
5	
6	
7	
8	
9	

10

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

W

_		
	0	а
	1	р
	2	W
	3	w?
	4	
	5	
	6	
	7	
	8	
	9	
	10	

Λ	
Arquivo	•
Marvo	•
•	

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

W

0	а
1	b
2	W
3	wa
4	
5	
6	
7	
8	
9	
10	

Arquivo:	
----------	--

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

W

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

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

W

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

Arquivo:

0 b W

0

3

5

0

0	а
1	b
2	W
3	wa
4	ab
5	
6	
7	
8	
9	
10	

Arquivo:

2 0 1 1 0 w a b

3

5

0	а
1	b
2	W
3	wa
4	ab
5	b?
6	
7	
8	
9	
10	

Arquivo:

b W a

0

0

3

5

Dicioná	rio								
0	а	Arquivo):						
1	b	2	0	1	1	0	3	5	0
2	w								
3	wa	W	a	b	b				
4	ab								
5	b?								
6									
7									
8									

Dicioná	ırio									
0	а	Arqu	iivo:							
1	b	2		0	1	1	0	3	5	0
2	w									
3	wa	W		a	b	b				
4	ab									
5	bb									
6										
7										
8										

Dicioná	ário							
0	а	Arquiv	0:					
1	b	2	0	1	1	0	3	5
2	w	L						
3	wa	W	а	b	b			
4	ab							
5	bb							
6	b?							
7								
8								
9								

Dicioná	ırio							
0	а	Arquivo):					
1	b	2	0	1	1	0	3	5
2	w							
3	wa	W	а	b	b	а		
4	ab							
5	bb							
6	b?							
7								
8								
9								

Dicioná	ário							
0	а	Arquivo):					
1	b	2	0	1	1	0	3	5
2	W	L			l		9	
3	wa	W	а	b	b	а		
4	ab							
5	bb							
6	ba							
7								
8								
9								

Dicioná	rio								
0	а	Arquivo	:						
1	b	2	0	1	1	0	3	5	0
2	w								
3	wa	W	а	b	b	а			
4	ab								
5	bb								
6	ba								
7	a?								
8									

Dicioná	rio									
0	а	Arquivo):							
1	b	2	0	1	1	0	3	3	5	
2	w									
3	wa	W	а	b	b	а	W	а		
4	ab									
5	bb									
6	ba									
7	a?									
8										
9										

Diciona	ario									
0	а	Arquivo	:							
1	b	2	0	1	1	0	9	3	ļ	 5
2	w									
3	wa	W	а	b	b	а	W	а		
4	ab									
5	bb									
6	ba									
7	aw									
8										
9										
10										

	Diciona	ario									
	0	а		Arquivo	:						
	1	b		2	0	1	1	0	3	3	5
	2	w	L								
•	3	wa		W	а	b	b	а	W	а	
	4	ab									
	5	bb									
	6	ba									
	7	aw									
	8	wa?									
	9										
- 1											

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

Arquivo:

W

0

a

b

b

0 a

3 W

a

b

5

b

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

Arquivo:

W

0

a

b

b

0 a

W

3

a

b

5

b

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

Arquivo:	

2

W

0

a

1 b

b

0 a W

3

5 b

b

		=								
0	а	 Arquiv	0:							
1	b			1 .		_		_		
2	w	2	0	1	1	0		3		5
3	wa	w	а	b	b	a	W	a	b	
4	ab									
5	bb									
6	ba									
7	aw									
8	wab									
9	bba									
10										

b

_			_									
	0	а		Arquivo):							
	1	b										
ļ	2	W		2	0	1	1	0	3		5	
	3	wa		W	a	b	b	а	W	а	b	
	4	ab										
	5	bb										
	6	ba										
Ī	7	aw										
	8	wab										
	9	bba										
	10	a?										

b

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