

COMPRESSÃO DE DADOS MÉTODOS DE DICIONÁRIO

Sérgio Mergen

Modelos baseado em dicionário

- O dicionário armazena uma tabela de códigos
 - Cada código equivale a uma sequência de caracteres
- São mapeadas as sequências de caracteres mais comuns
 - Que se repetem bastante ao longo do arquivo

Modelos baseado em dicionário

- Os métodos mais simples de compressão usam pequenos dicionários estáticos
 - Codificação diatômica (ou codificação de digramas)
 - Substituição de padrões

Codificação diatômica

- Substitui pares de caracteres frequentes por um código
- Os identificadores são normalmente caracteres não imprimíveis

AVIÃO COMERCIAL



AVI# \$ERCI&

Amostra do dicionário

id	par
#	ÃO
\$	CO
&	AL
...	...

- Um dicionário para o código de caracteres ASCII pode conter os 128 caracteres bem como 128 pares de caracteres mais comuns

Codificação diatômica

- As palavras de código de saída têm 8 bits cada
 - Início "zero" indica um dos caracteres ASCII
 - Início "um" indica um código
- No melhor dos casos
 - cada par de caracteres é substituído por uma palavra de código, reduzindo de 8 bits/caracter para 4 bits/caracter
- No pior dos casos
 - cada caracter permanece representado com 8 bits
- Problemas:
 - O tamanho da tabela é limitado pelo número de códigos disponíveis ($2^7 = 128$)
 - Não é uma abordagem universal

Substituição de padrões

- Substitui trechos que se repetem por um código
- Codificação: CN
 - C = símbolo especial
 - N = identificador do trecho no dicionário

*BAB**CA**BAB**D**DCC**BAB**B**DCC*



#1CA#1D#5#1B#5

Amostra do dicionário

id	Binário	trecho
00000001	1	BAB
00000002	2	DCC
...

- Alguns problemas:
 - A tabela comporta o máximo de $2^8 = 256$ identificadores
 - Continua não sendo uma abordagem universal

Modelos baseado em dicionário

- Problemas dos modelos estáticos:
 - Os códigos criados são dependentes do domínio
 - Necessário codificar sequências curtas para a abordagem ser mais genérica
 - Sob pena de taxas menores de compressão
- Solução?
 - Uso de dicionários semi-estáticos

Modelos baseado em dicionário

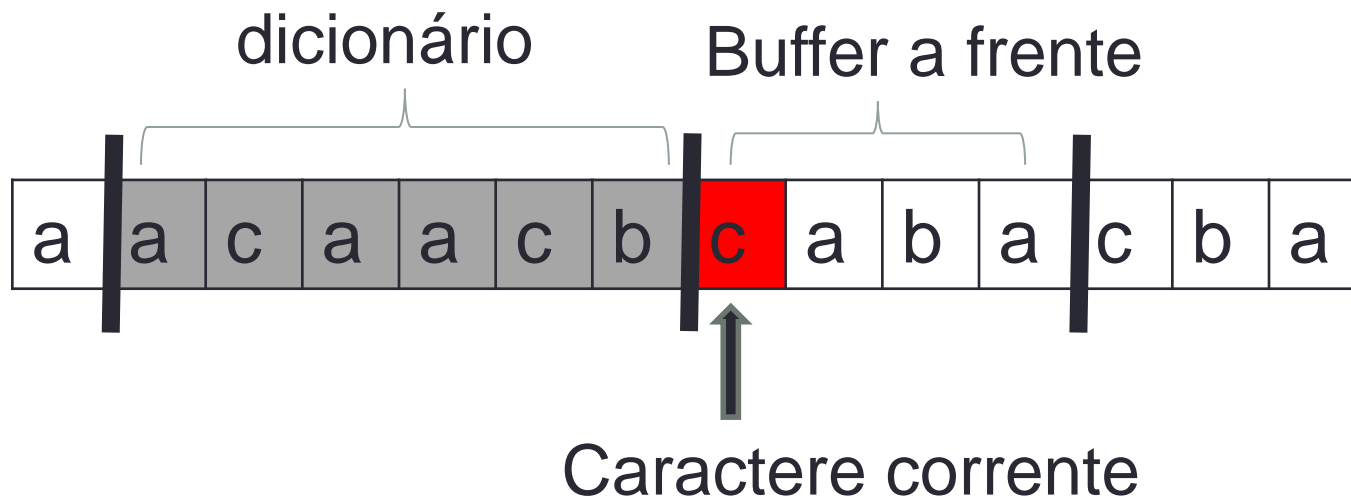
- Dicionário semi-estático
 - um novo dicionário é construído para todo arquivo que se deseja comprimir
- Problemas
 - Custo de criação do dicionário
 - Custo de armazenamento do dicionário
- Solução?
 - Uso de dicionários adaptativos

Dicionários adaptativos

- O dicionário é composto por sequências de caracteres que aparecem em alguma parte anterior no texto
- Alguns esquemas adaptativos
 - Codificadores Lempel-Ziv
 - LZ77
 - LZSS
 - LZ78
 - LZW

LZ77 - Compressão

- Dicionário
 - Cadeia de onde se consegue extrair padrões
- Buffer a frente (lookahead buffer)
 - Cadeia de caracteres onde se busca algum padrão
- O tamanho das estruturas é prefixado
 - O buffer é menor do que o dicionário



LZ77 - Compressão

- Um cursor avança pelos caracteres de entrada
 - Fazendo o dicionário e o buffer deslizarem para frente
- A partir do caractere corrente
 - pode-se gerar um código
- Formato de cada código gerado: $\langle a, b, c \rangle$
 - a: um offset relativo a distância a se voltar a partir do caractere corrente, até encontrar o início de um padrão
 - b: um número relativo ao tamanho do padrão encontrado
 - c: o próximo caractere de entrada

LZ77 - Compressão

- Quando gerar um código?
 - A partir do caractere corrente
 - Encontrar a maior sequência dentro do buffer que possua correspondência no dicionário
 - Essa sequência será codificada
- E se não existir padrão?
 - O offset e a distância são marcados com zero (0)

Exemplo de codificação – LZ77

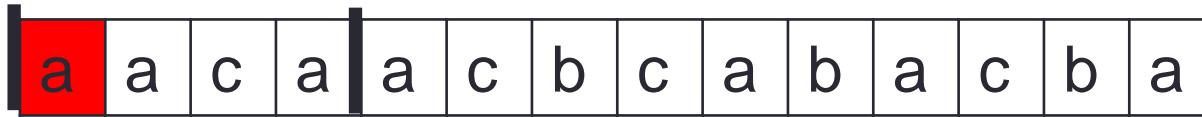
Próximo caractere não existe no dicionário.

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(, ,)

Exemplo de codificação – LZ77

Adicionar caractere na saída, sem referência ao dicionário



(0, 0, a)

Exemplo de codificação – LZ77

Próximo caractere existe no dicionário(**a**).

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

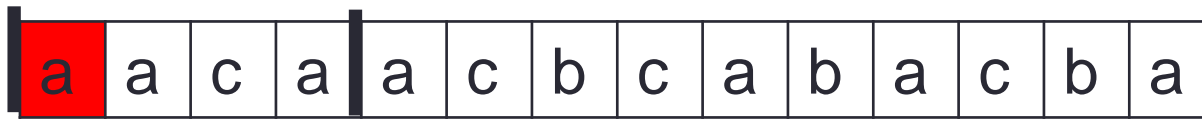
(0, 0, a)

1	a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

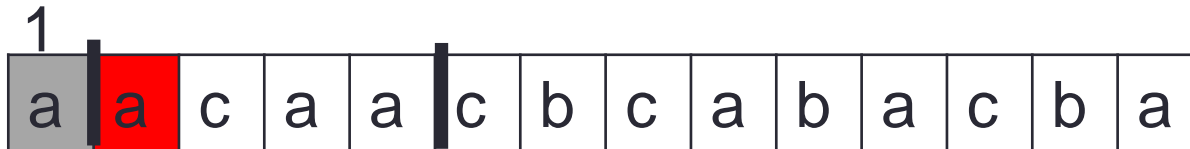
(, ,)

Exemplo de codificação – LZ77

A partir do início do buffer, descobrir a posição desse caractere no dicionário (1)



(0, 0, a)



(1, ,)

Exemplo de codificação – LZ77

Acrescentar o número de caracteres mapeados ($\{a\} = 1$)



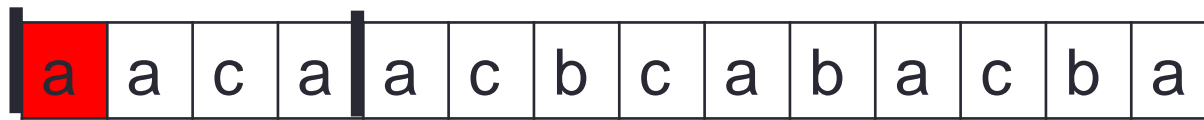
(0, 0, a)



(1, 1,)

Exemplo de codificação – LZ77

Acrescentar o próximo caractere



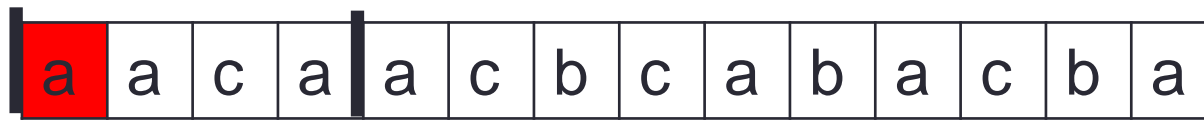
(0, 0, a)



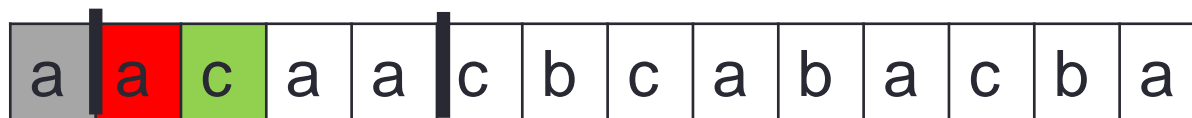
(1, 1, c)

Exemplo de codificação – LZ77

Próximo caractere existe no dicionário (**a**).



(0, 0, a)



(1, 1, c)



(, ,)

Exemplo de codificação – LZ77

No entanto, existe sequência maior no dicionário (**aac**).

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(0, 0, a)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

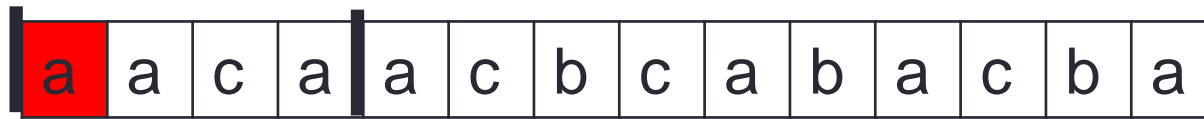
(1, 1, c)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

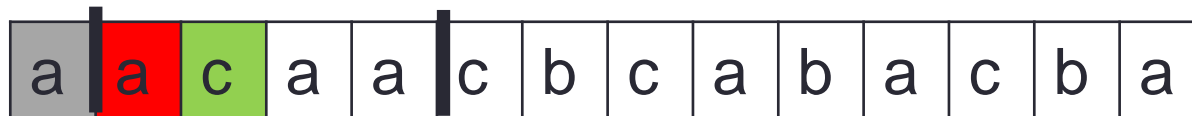
(, ,)

Exemplo de codificação – LZ77

A partir do início do buffer, descobrir a posição dessa sequência no dicionário (3)



(0, 0, a)



(1, 1, c)



(3, ,)

Exemplo de codificação – LZ77

Acrescentar o número de caracteres mapeados ($\{aac\} = 3$)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(0, 0, a)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

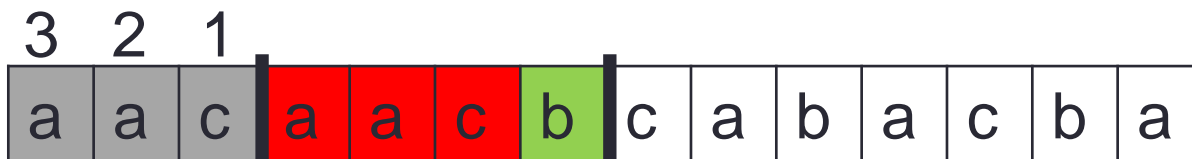
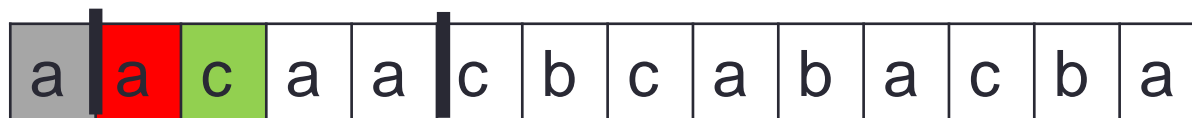
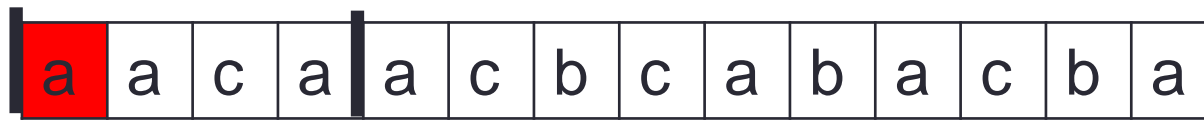
(1, 1, c)

3	2	1	a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

(3, 3,)

Exemplo de codificação – LZ77

Acrescentar o próximo caractere (b)



Exemplo de codificação – LZ77

Próximo caractere existe no dicionário (**c**).

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(0, 0, a)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(1, 1, c)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(3, 3, b)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

Exemplo de codificação – LZ77

No entanto, existe sequência maior no dicionário (**ca**).

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(0, 0, a)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(1, 1, c)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(3, 3, b)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(, ,)

Exemplo de codificação – LZ77

A partir do início do buffer, descobrir a posição dessa sequência no dicionário (5)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(0, 0, a)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(1, 1, c)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

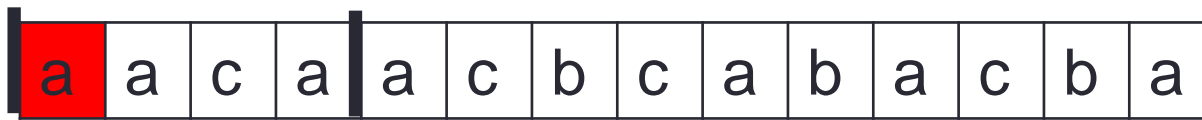
(3, 3, b)

		5	4	3	2	1							
a	a	c	a	a	c	b	c	a	b	a	c	b	a

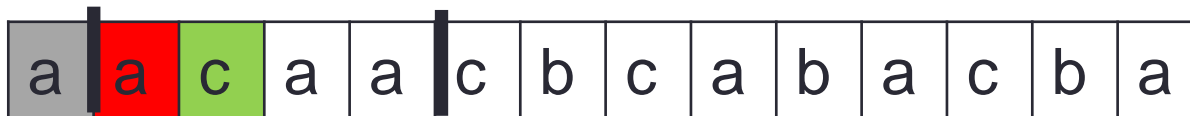
(5, ,)

Exemplo de codificação – LZ77

Acrescentar o número de caracteres mapeados ($\{ca\} = 2$)



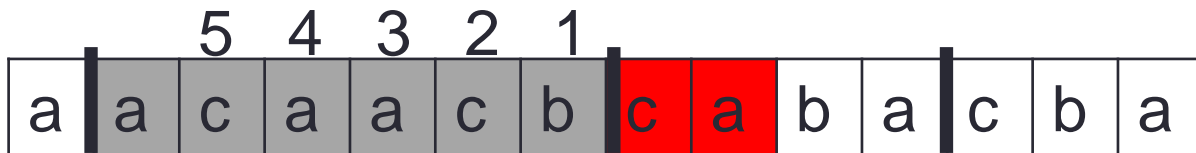
(0, 0, a)



(1, 1, c)



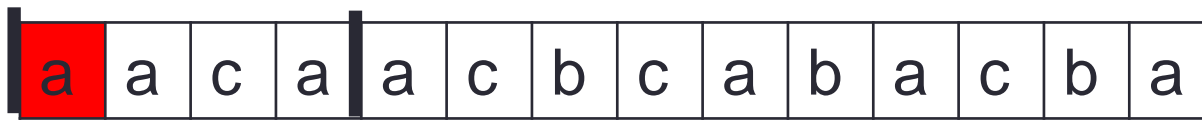
(3, 3, b)



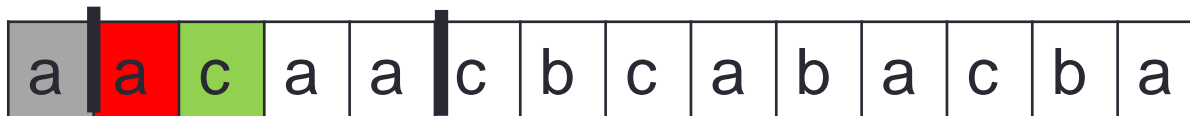
(5, 2,)

Exemplo de codificação – LZ77

Acrescentar o próximo caractere (b)



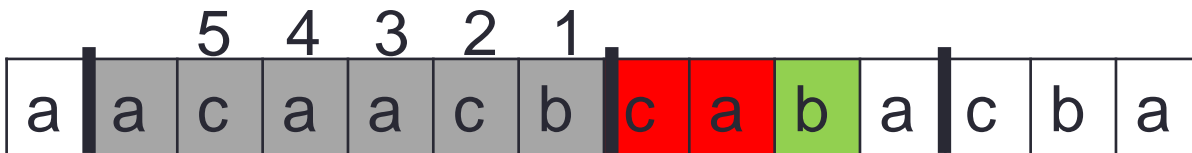
(0, 0, a)



(1, 1, c)



(3, 3, b)



(5, 2, b)

Exemplo de codificação – LZ77

Próximo caractere existe no dicionário (**a**).

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(0, 0, a)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(1, 1, c)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(3, 3, b)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(5, 2, b)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(, ,)

Exemplo de codificação – LZ77

No entanto, existe sequência maior no dicionário (**acb**).

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(0, 0, a)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(1, 1, c)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(3, 3, b)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

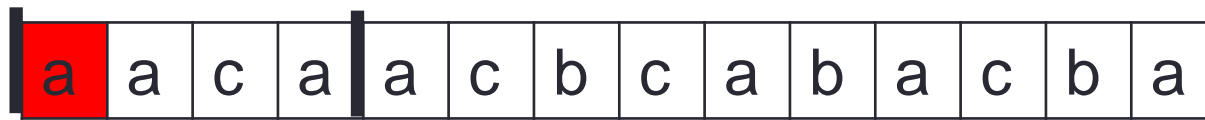
(5, 2, b)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

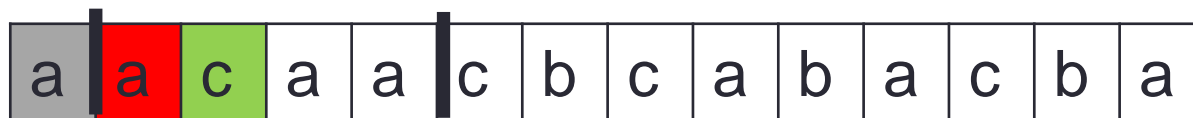
(, ,)

Exemplo de codificação – LZ77

A partir do início do buffer, descobrir a posição dessa sequência no dicionário (6)



(0, 0, a)



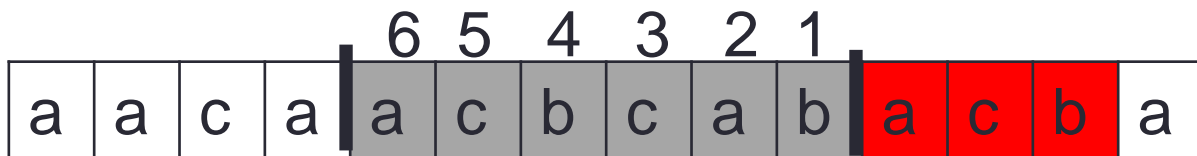
(1, 1, c)



(3, 3, b)



(5, 2, b)



(6, ,)

Exemplo de codificação – LZ77

Acrescentar o número de caracteres mapeados ($\{acb\} = 3$)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(0, 0, a)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(1, 1, c)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(3, 3, b)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

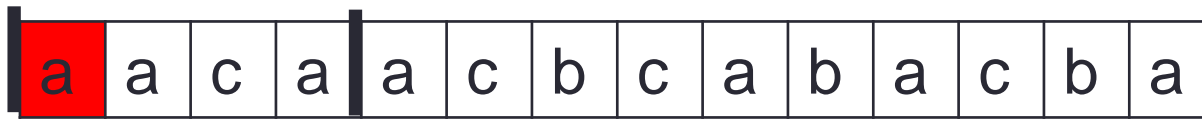
(5, 2, b)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

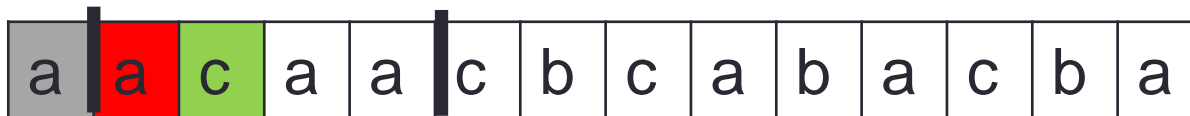
(6, 3,)

Exemplo de codificação – LZ77

Acrescentar o próximo caractere (a)



(0, 0, a)



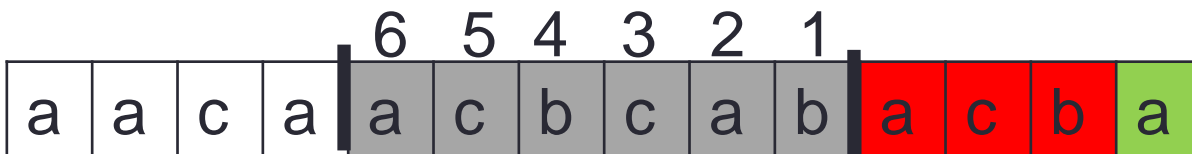
(1, 1, c)



(3, 3, b)



(5, 2, b)



(6, 3, a)

Exemplo de decodificação – LZ77

Offset zero significa que não existe nada no buffer.
O caractere é copiado para a saída

(0, 0, a)



(1, 1, c)

(3, 3, b)

(5, 2, b)

(6, 3, a)

Exemplo de decodificação – LZ77

A sequência de 1 caractere encontrado 1 caractere atrás é copiada para a saída

(0, 0, a)



1

(1, 1, c)



(3, 3, b)

(5, 2, b)

(6, 3, a)

Exemplo de decodificação – LZ77

Em seguida, é copiado o caractere codificado (c)

(0, 0, a)



(1, 1, c)



(3, 3, b)

(5, 2, b)

(6, 3, a)

Exemplo de decodificação – LZ77

A sequência de 3 caractere encontrados 3 caractere atrás é copiada para a saída

(0, 0, a)



(1, 1, c)



3 2 1

(3, 3, b)



(5, 2, b)

(6, 3, a)

Exemplo de decodificação – LZ77

Em seguida, é copiado o caractere codificado (b)

(0, 0, a)



(1, 1, c)



(3, 3, b)



(5, 2, b)

(6, 3, a)

Exemplo de decodificação – LZ77

A sequência de 2 caractere encontrados 5 caractere atrás é copiada para a saída

(0, 0, a)



(1, 1, c)



(3, 3, b)



5 4 3 2 1

(5, 2, b)



(6, 3, a)

Exemplo de decodificação – LZ77

Em seguida, é copiado o caractere codificado (b)

(0, 0, a)



(1, 1, c)



(3, 3, b)



(5, 2, b)



(6, 3, a)

Exemplo de decodificação – LZ77

A sequência de 3 caractere encontrados 6 caractere atrás é copiada para a saída

(0, 0, a)



(1, 1, c)



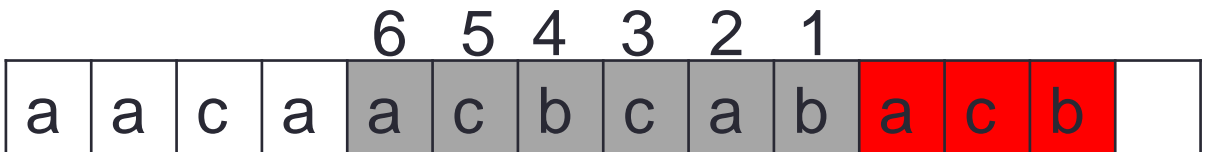
(3, 3, b)



(5, 2, b)



(6, 3, a)



Exemplo de decodificação – LZ77

Em seguida, é copiado o caractere codificado (a)

(0, 0, a)



(1, 1, c)



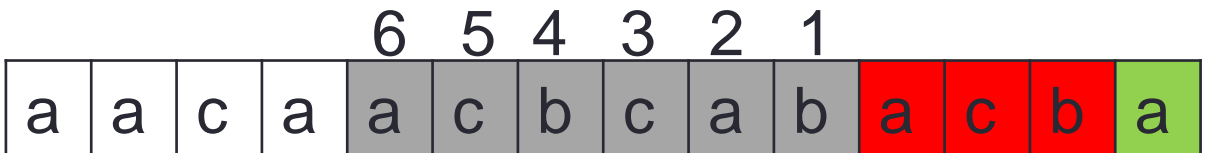
(3, 3, b)



(5, 2, b)



(6, 3, a)



LZ77

- O dicionário é transmitido sem custo extra
 - Seus dados estão no próprio texto decodificado.
- Deve-se escolher com cuidado o tamanho do buffer e do dicionário
 - Senão gasta-se muito tempo na busca sem benefício
- Para texto em inglês
 - Um dicionário de poucos milhares de caracteres
 - Buffer de no máximo 16 caracteres

LZSS (Lempel-Ziv-Storer-Szymanski)

- Extensão do LZ77
 - Cria códigos menores para padrões curtos
 - Usado pelo gzip
- Para codificar padrões:
 - Usa $\langle 1, a, b \rangle$
 - 1: o bit 1
 - A: início do padrão
 - B: tamanho do padrão
- Para codificar caracteres simples
 - Usa $\langle 0, c \rangle$
 - 0: o bit 0
 - C: o caractere a codificar

Exemplo de codificação – LZSS

- O formato mais curto é usado quando
 - o tamanho do padrão for < 3

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(1, a)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(1, a)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(1, c)

a	a	c	a	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---	---

(0, 3, 3)

Análise do LZ77

- A taxa de compressão depende do tamanho do dicionário
- No entanto, quanto maior o dicionário
 - Menor será a eficiência
- Como trabalhar com dicionários maiores?
 - LZ78

LZ78

Algoritmo básico:

- Manter dicionário de padrões com um id para cada entrada
 - Usando uma **Trie**
- Codificação
 - A partir da entrada, encontrar o maior padrão S existente no dicionário
 - Codificar com o id correspondente + o próximo caractere na entrada c
 - Adicionar Sc ao dicionário
- Se não existir padrão no dicionário
 - Codificar com id = 0

Exemplo de codificação – LZ78

Próximo caractere (a) não existe no dicionário .

Dicionário

pos	padrão
-----	--------

a	a	c	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---

Exemplo de codificação – LZ78

Adicionar código com pos = 0 e caractere = próprio caractere (a)

Adicionar no dicionário o caractere

a	a	c	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---

(0, a)

Dicionário

pos	padrão
1	a

Exemplo de codificação – LZ78

A maior sequência no dicionário é (a), na posição 1.

a	a	c	a	c	b	c	a	b	a	c	b	a
a	a	c	a	c	b	c	a	b	a	c	b	a

(0, a)

Dicionário

pos	padrão
1	a

Exemplo de codificação – LZ78

Adicionar código com pos = 1 e caractere = próprio caractere (c)

Adicionar no dicionário o caractere (a) + o próximo caractere (c)

Dicionário

a	a	c	a	c	b	c	a	b	a	c	b	a	(0, a)
a	a	c	a	c	b	c	a	b	a	c	b	a	(1, c)

pos	padrão
1	a
2	ac

Exemplo de codificação – LZ78

A maior sequência no dicionário é (ac), na posição 2.

a	a	c	a	c	b	c	a	b	a	c	b	a	(0, a)
a	a	c	a	c	b	c	a	b	a	c	b	a	(1, c)
a	a	c	a	c	b	c	a	b	a	c	b	a	

Dicionário

pos	padrão
1	a
2	ac

Exemplo de codificação – LZ78

Adicionar código com pos = 2 e caractere = próprio caractere (b)

Adicionar no dicionário a sequência (ac) + o próximo caractere (b) **Dicionário**

a	a	c	a	c	b	c	a	b	a	c	b	a	(0, a)
a	a	c	a	c	b	c	a	b	a	c	b	a	(1, c)
a	a	c	a	c	b	c	a	b	a	c	b	a	(2, b)

pos	padrão
1	a
2	ac
3	acb

Exemplo de codificação – LZ78

Próximo caractere (c) não existe no dicionário.

a	a	c	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---

(0, a)

a	a	c	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---

(1, c)

a	a	c	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---

(2, b)

a	a	c	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---

Dicionário

pos	padrão
1	a
2	ac
3	acb

Exemplo de codificação – LZ78

Adicionar código com pos = 0 e caractere = próprio caractere (c)

Adicionar no dicionário o caractere

a	a	c	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---

(0, a)

a	a	c	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---

(1, c)

a	a	c	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---

(2, b)

a	a	c	a	c	b	c	a	b	a	c	b	a
---	---	---	---	---	---	---	---	---	---	---	---	---

(0, c)

Dicionário

pos	padrão
1	a
2	ac
3	acb
4	c

Exemplo de codificação – LZ78

A maior sequência no dicionário é (a), na posição 1.

a	a	c	a	c	b	c	a	b	a	c	b	a	(0, a)
a	a	c	a	c	b	c	a	b	a	c	b	a	(1, c)
a	a	c	a	c	b	c	a	b	a	c	b	a	(2, b)
a	a	c	a	c	b	c	a	b	a	c	b	a	(0, c)
a	a	c	a	c	b	c	a	b	a	c	b	a	

Dicionário

pos	padrão
1	a
2	ac
3	acb
4	c

Exemplo de codificação – LZ78

Adicionar código com pos = 1 e caractere = próprio caractere (b)

Adicionar no dicionário o caractere (a) + o próximo caractere (b)

Dicionário

a	a	c	a	c	b	c	a	b	a	c	b	a	(0, a)
a	a	c	a	c	b	c	a	b	a	c	b	a	(1, c)
a	a	c	a	c	b	c	a	b	a	c	b	a	(2, b)
a	a	c	a	c	b	c	a	b	a	c	b	a	(0, c)
a	a	c	a	c	b	c	a	b	a	c	b	a	(1, b)

pos	padrão
1	a
2	ac
3	acb
4	c
5	ab

Exemplo de codificação – LZ78

A maior sequência no dicionário é (acb), na posição 3.

a	a	c	a	c	b	c	a	b	a	c	b	a	(0, a)
a	a	c	a	c	b	c	a	b	a	c	b	a	(1, c)
a	a	c	a	c	b	c	a	b	a	c	b	a	(2, b)
a	a	c	a	c	b	c	a	b	a	c	b	a	(0, c)
a	a	c	a	c	b	c	a	b	a	c	b	a	(1, b)
a	a	c	a	c	b	c	a	b	a	c	b	a	

Dicionário

pos	padrão
1	a
2	ac
3	acb
4	c
5	ab

Exemplo de codificação – LZ78

Adicionar código com pos = 3 e caractere = próximo caractere (a)

Adicionar no dicionário a sequência (acb) + o próx. caractere (a)

Dicionário

a	a	c	a	c	b	c	a	b	a	c	b	a	(0, a)
a	a	c	a	c	b	c	a	b	a	c	b	a	(1, c)
a	a	c	a	c	b	c	a	b	a	c	b	a	(2, b)
a	a	c	a	c	b	c	a	b	a	c	b	a	(0, c)
a	a	c	a	c	b	c	a	b	a	c	b	a	(1, b)
a	a	c	a	c	b	c	a	b	a	c	b	a	(3, a)

pos	padrão
1	a
2	ac
3	acb
4	c
5	ab
6	acba

Exemplo de decodificação – LZ78

Código zero significa que nada no dicionário pode ser aproveitado

Dicionário

pos	padrão
-----	--------

(0, a)

--	--	--	--	--	--	--	--	--	--	--	--	--	--

(1, c)

(2, b)

(0, c)

(1, b)

(3, a)

Exemplo de decodificação – LZ78

O caractere codificado é copiado para a saída

O caractere codificado é copiado para o dicionário

(0, a)

a													
---	--	--	--	--	--	--	--	--	--	--	--	--	--

(1, c)

(2, b)

(0, c)

(1, b)

(3, a)

Dicionário

pos	padrão
1	a

Exemplo de decodificação – LZ78

A entrada do dicionário na posição 1 + o caractere codificado (c)
são copiados para a saída

Esse conteúdo é copiado para o dicionário

(0, a)	<div><div>a</div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>
(1, c)	<div><div>a</div><div>a</div><div>c</div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>
(2, b)	
(0, c)	
(1, b)	
(3, a)	

Dicionário

pos	padrão
1	a
2	ac

Exemplo de decodificação – LZ78

A entrada do dicionário na posição 2 + o caractere codificado (b)
são copiados para a saída
Esse conteúdo é copiado para o dicionário

(0, a)	<div>a</div>
(1, c)	<div>aac</div>
(2, b)	<div>aacab</div>
(0, c)	
(1, b)	
(3, a)	

Dicionário

pos	padrão
1	a
2	ac
3	acb

Exemplo de decodificação – LZ78

Código zero significa que nada no dicionário pode ser aproveitado

(0, a)



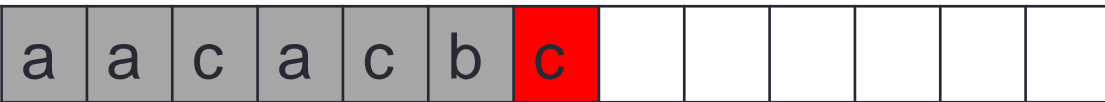
(1, c)



(2, b)



(0, c)



(1, b)

(3, a)

Dicionário

pos	padrão
1	a
2	ac
3	acb

Exemplo de decodificação – LZ78

O caractere codificado é copiado para a saída

O caractere codificado é copiado para o dicionário

(0, a)



(1, c)



(2, b)



(0, c)



(1, b)

(3, a)

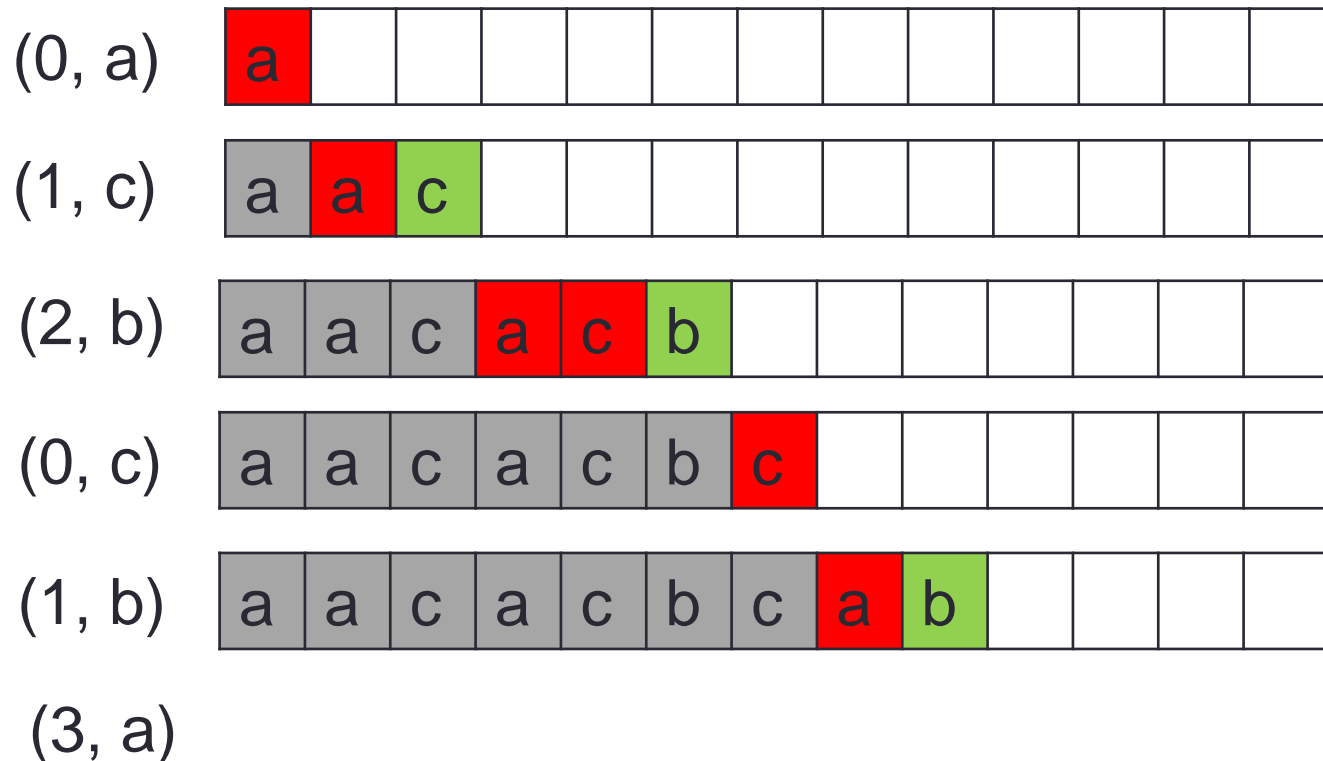
Dicionário

pos	padrão
1	a
2	ac
3	acb
4	c

Exemplo de decodificação – LZ78

A entrada do dicionário na posição 1 + o caractere codificado (b)
são copiados para a saída

Esse conteúdo é copiado para o dicionário



Dicionário

pos	padrão
1	a
2	ac
3	acb
4	c
5	ab

Exemplo de decodificação – LZ78

A entrada do dicionário na posição 3 + o caractere codificado (a)
são copiados para a saída

Esse conteúdo é copiado para o dicionário

(0, a)	a												
(1, c)	a	a	c										
(2, b)	a	a	c	a	c	b							
(0, c)	a	a	c	a	c	b	c						
(1, b)	a	a	c	a	c	b	c	a	b				
(3, a)	a	a	c	a	c	b	c	a	b	a	c	b	a

Dicionário

pos	padrão
1	a
2	ac
3	acb
4	c
5	ab
6	acba

Na próxima aula

- Estratégia LZW
 - Uma variação do LZ78
- O método de Burrows Wheeler
- Uma análise de todas as estratégias

Exercício

- Dada a seguinte cadeia de caracteres:
 - a_aba_da_aba_da_baba
- Mostre a codificação que seria obtida pelos seguintes métodos de compressão:
 - LZ77
 - LZ78
- Para o método LZ77, considere um dicionário de tamanho seis e um buffer de tamanho quatro.