

Universidade Estadual de Campinas - UNICAMP
Instituto de Computação - IC

Espalhamento (*hashing*): Parte 1

- 1 Introdução
- 2 Funções de espalhamento
- 3 Tratamento de colisões

- Tipo abstrato de dados com as operações básicas de Inserir, Remover e Buscar.

- Típa abstrato de dados com as operações básicas de Inserir, Remover e Buscar.

– **Tabela de espalhamento ou hashing** : matriz com $b \geq 1$ linhas e $s \geq 1$ colunas.

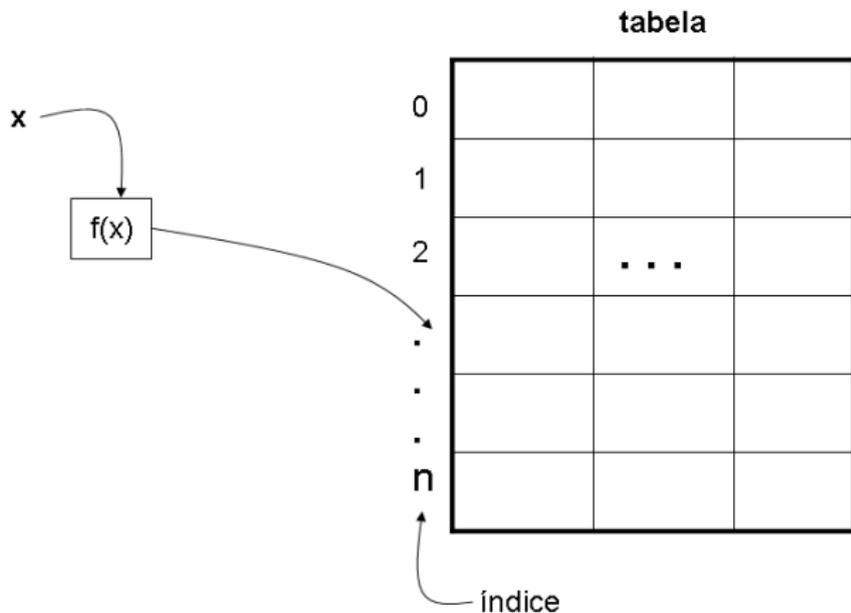
$f('joão') \rightarrow$

	1	2	3
0			
1			
2			
3	joão		
4			
5			
6			

– $b = 7$ e $s = 3$

Espalhamento

- Princípio: dado um nome (ou chave) x a ser inserido, removido ou consultado na tabela, se quisermos executar estas operações em tempo $O(1)$, a localização de x deve depender apenas desta chave.



- Exemplo de tabela de espalhamento ($b = 26$ e $s = 2$):

	1	2
0	antônio	áttila
1		
2	carlos	célio
3	douglas	
4	ernesto	estêvão
5		
...		
24		
25	zoroastro	

– Função de espalhamento muito simples: baseada no índice da primeira letra das chaves ($a \rightarrow 0$, $b \rightarrow 1$, ...).

- Vantagens e problemas:
 - Vantagens:
 - simplicidade
 - inserção e busca muito rápidas (se a função de espalhamento for eficiente)
 - Problemas:
 - escolha da função de espalhamento
 - tratamento de colisões
 - *overflow* da tabela

- Vantagens e problemas:
 - Vantagens:
 - simplicidade
 - inserção e busca muito rápidas (se a função de espalhamento for eficiente)
 - Problemas:
 - escolha da função de espalhamento
 - tratamento de colisões
 - *overflow* da tabela

- Técnicas para construção de funções de espalhamento
 - Propriedades desejáveis:
 - eficiência de cálculo
 - bom espalhamento (espalhamento mais uniforme)
 - Técnicas:
 - espalhamento mínimo perfeito
 - espalhamento pseudo-aleatório

- Exemplo de método de espalhamento:

Divisão:

A chave, tratada como um número na base 26, é dividida por um número p , preferencialmente *primo*, e $f(x) = x \bmod p$.

- Exemplo para $p = 53$:

$$f(\textit{carlos}) = (2 \times 26^5 + 0 \times 26^4 + 17 \times 26^3 + 11 \times 26^2 + 14 \times 26^1 + 18 \times 26^0) \bmod 53$$

$$f(\textit{carlos}) = 24.069.362 \bmod 53 = 48$$

- Exemplo de método de espalhamento:

Divisão:

A chave, tratada como um número na base 26, é dividida por um número p , preferencialmente *primo*, e $f(x) = x \bmod p$.

- Exemplo para $p = 53$:

$$f(\textit{carlos}) = (2 \times 26^5 + 0 \times 26^4 + 17 \times 26^3 + 11 \times 26^2 + 14 \times 26^1 + 18 \times 26^0) \bmod 53$$

$$f(\textit{carlos}) = 24.069.362 \bmod 53 = 48$$

- Seleção de algarismos:

– A chave é tratada como uma sequência de algarismos ou de *bytes* ou de *bits*, e uma subsequência é selecionada para representar o índice. Por exemplo, suponhamos que todos os nomes são representados pela sequência de dígitos $x = d_0d_1\dots d_{11}$ em alguma base conveniente; uma escolha da função de espalhamento seria $f(x) = d_3d_5d_9$.

– Exemplo: *carlos*. A sua representação poderia ser 020017111418. Neste caso, supusemos que cada letra é indicada por dois dígitos que indicam a sua posição no alfabeto, ou seja 00 para *a*, 01 para *b* etc. Teríamos então $f(\textit{carlos}) = 074$.

– O método *mid-square* consiste do cálculo do quadrado da representação numérica da chave antes das extração de k termos do meio do valor resultante.

- Seleção de algarismos:

– A chave é tratada como uma sequência de algarismos ou de *bytes* ou de *bits*, e uma subsequência é selecionada para representar o índice. Por exemplo, suponhamos que todos os nomes são representados pela sequência de dígitos $x = d_0d_1\dots d_{11}$ em alguma base conveniente; uma escolha da função de espalhamento seria $f(x) = d_3d_5d_9$.

– Exemplo: *carlos*. A sua representação poderia ser 020017111418. Neste caso, supusemos que cada letra é indicada por dois dígitos que indicam a sua posição no alfabeto, ou seja 00 para *a*, 01 para *b* etc. Teríamos então $f(\textit{carlos}) = 074$.

– O método *mid-square* consiste do cálculo do quadrado da representação numérica da chave antes das extração de k termos do meio do valor resultante.

- Seleção de algarismos:

– A chave é tratada como uma sequência de algarismos ou de *bytes* ou de *bits*, e uma subsequência é selecionada para representar o índice. Por exemplo, suponhamos que todos os nomes são representados pela sequência de dígitos $x = d_0d_1\dots d_{11}$ em alguma base conveniente; uma escolha da função de espalhamento seria $f(x) = d_3d_5d_9$.

– Exemplo: *carlos*. A sua representação poderia ser 020017111418. Neste caso, supusemos que cada letra é indicada por dois dígitos que indicam a sua posição no alfabeto, ou seja 00 para *a*, 01 para *b* etc. Teríamos então $f(\textit{carlos}) = 074$.

– O método *mid-square* consiste do cálculo do quadrado da representação numérica da chave antes das extração de k termos do meio do valor resultante.

- Dobramento (*folding*)

– A chave é tratada como uma sequência de algarismos ou de *bytes* ou de *bits*, e algumas subsequências são combinadas por operações convenientes para produzir um índice.

Por exemplo, suponhamos que todas as chaves são representadas pela sequência de dígitos $x = d_0d_1\dots d_{11}$ em alguma base conveniente; uma escolha da função seria $f(x) = d_3d_5d_9 \oplus d_7d_8d_{10}$, em que \oplus denota a operação de *ou exclusivo bit a bit*.

- Endereçamento aberto
 - Busca sistemática de outras entradas disponíveis na tabela:
 - reespalhamento linear
 - reespalhamento quadrático
 - reespalhamento duplo

- Reespalhamento linear

– Se houver colisão, procura-se a próxima posição consecutiva livre da tabela $\Rightarrow (f(x) + i) \bmod b, (i = 0, 1, \dots)$.

- Reespalhamento linear

- Se houver colisão, procura-se a próxima posição consecutiva livre da tabela $\Rightarrow (f(x) + i) \bmod b, (i = 0, 1, \dots)$.
- Exemplo: Inserir *antônio, carlos, douglas, célio, armando, zoroastro, átila* e *alfredo* (nesta ordem).

- Reespalhamento linear

– Se houver colisão, procura-se a próxima posição consecutiva livre da tabela $\Rightarrow (f(x) + i) \bmod b, (i = 0, 1, \dots)$.

– Exemplo: Inserir *antônio*, *carlos*, *douglas*, *célio*, *armando*, *zoroastro*, *áttila* e *alfredo* (nesta ordem).

0	antônio
1	armando
2	carlos
3	douglas
4	célio
5	áttila
6	alfredo
7	
...	
25	zoroastro

- Reespalhamento quadrático

– A busca de uma posição livre na tabela é dada por $(f(x) + i^2) \bmod b, (i = 0, 1, \dots)$.

- Reespalhamento quadrático

- A busca de uma posição livre na tabela é dada por $(f(x) + i^2) \bmod b$, ($i = 0, 1, \dots$).
- Exemplo: Inserir *antônio*, *carlos*, *douglas*, *célio*, *armando*, *zoroastro*, *átilla* e *alfredo* (nesta ordem).

- Reespalhamento quadrático

– A busca de uma posição livre na tabela é dada por $(f(x) + i^2) \bmod b, (i = 0, 1, \dots)$.

– Exemplo: Inserir *antônio*, *carlos*, *douglas*, *célio*, *armando*, *zoroastro*, *áttila* e *alfredo* (nesta ordem).

0	antônio
1	armando
2	carlos
3	douglas
4	áttila
5	
6	célio
7	
8	
9	alfredo
	...
25	zoroastro

- Reespalhamento duplo

– A busca de uma posição livre envolve duas funções de espalhamento: uma primária, $f(x)$, e uma secundária, $g(x)$.

- Reespalhamento duplo

- A busca de uma posição livre envolve duas funções de espalhamento: uma primária, $f(x)$, e uma secundária, $g(x)$.
- Esta busca é dada pela fórmula $(f(x) + i \times g(x)) \bmod b$, ($i = 0, 1, \dots$).

- Reespalhamento duplo

- A busca de uma posição livre envolve duas funções de espalhamento: uma primária, $f(x)$, e uma secundária, $g(x)$.
- Esta busca é dada pela fórmula $(f(x) + i \times g(x)) \bmod b$, ($i = 0, 1, \dots$).
- Exemplo: Inserir *antônio, carlos, douglas, célio, armando, zoroastro, átila e alfredo* (nesta ordem).

- Reespalhamento duplo

- A busca de uma posição livre envolve duas funções de espalhamento: uma primária, $f(x)$, e uma secundária, $g(x)$.
- Esta busca é dada pela fórmula $(f(x) + i \times g(x)) \bmod b$, ($i = 0, 1, \dots$).
- Exemplo: Inserir *antônio, carlos, douglas, célio, armando, zoroastro, átila e alfredo* (nesta ordem).

$$(f(x) + i \times g(x)) \bmod b,$$
$$(i = 0, 1, \dots)$$

Exemplo:

$$g(x) = (c \bmod 3) + 1$$

c = posição da 2ª letra da chave

• Reespalhamento duplo

- A busca de uma posição livre envolve duas funções de espalhamento: uma primária, $f(x)$, e uma secundária, $g(x)$.
- Esta busca é dada pela fórmula $(f(x) + i \times g(x)) \bmod b$, ($i = 0, 1, \dots$).
- Exemplo: Inserir *antônio*, *carlos*, *douglas*, *célio*, *armando*, *zoroastro*, *áttila* e *alfredo* (nesta ordem).

$$(f(x) + i \times g(x)) \bmod b,$$
$$(i = 0, 1, \dots)$$

Exemplo:

$$g(x) = (c \bmod 3) + 1$$

c = posição da 2ª letra da chave

0	antônio
1	
2	carlos
3	douglas
4	célio
5	
6	armando
7	
8	áttila
9	alfredo
	...
25	zoroastro

Referências:

- Aaron M. Tenenbaum et al. Estruturas de Dados Usando C, Makron Books, 1995.
- Apostila dos professores Thomasz e Lucchesi.