

Visão geral

Endereçamento
aberto

Visão geral
Sondagens
Desempenho
Remoções
Operações

Encadeamento

Visão geral
Remoções
Operações

Endereçamento
em baldes

Visão geral

Organização e Recuperação da Informação

Tabelas hash: tratamento de colisões

Jander Moreira
jander@dc.ufscar.br

Universidade Federal de São Carlos
Departamento de Computação

2014

Visão geral

Endereçamento
aberto

Visão geral
Sondagens
Desempenho
Remoções
Operações

Encadeamento

Visão geral
Remoções
Operações

Endereçamento
em baldes

Visão geral

1 Visão geral

2 Endereçamento aberto

Visão geral
Sondagens
Desempenho
Remoções
Operações

3 Encadeamento

Visão geral
Remoções
Operações

4 Endereçamento em baldes

Visão geral

Visão geral

Endereçamento
aberto

Visão geral
Sondagens
Desempenho
Remoções
Operações

Encadeamento

Visão geral
Remoções
Operações

Endereçamento
em baldes

Visão geral

Visão geral

Visão geral

Endereçamento
aberto

Visão geral
Sondagens
Desempenho
Remoções
Operações

Encadeamento

Visão geral
Remoções
Operações

Endereçamento
em baldes

Visão geral

Formas de tratamento de colisões mais usuais:

- Endereçamento aberto
- Encadeamento
- Endereçamento em balde (*bucket*)

Visão geral

Endereçamento
aberto

Visão geral
Sondagens
Desempenho
Remoções
Operações

Encadeamento

Visão geral
Remoções
Operações

Endereçamento
em baldes

Visão geral

Considerações sobre o tratamento de colisões:

- Uso das posições livres da própria tabela
- Reserva de parte da tabela exclusiva para colisões
- Uso de área externa à tabela

Visão geral

Endereçamento
aberto

Visão geral

Sondagens

Desempenho

Remoções

Operações

Encadeamento

Visão geral

Remoções

Operações

Endereçamento
em baldes

Visão geral

Endereçamento aberto

Endereçamento aberto

Visão geral

Endereçamento
aberto

Visão geral

Sondagens

Desempenho

Remoções

Operações

Encadeamento

Visão geral

Remoções

Operações

Endereçamento
em baldes

Visão geral

Endereçamento aberto:

- Escolha de outra posição da tabela
- Definição de uma função de sondagem (ou *rehash*)
- Tentativas para colisão em uma posição k :

$$(k + rh(1)) \% tam, (k + rh(2)) \% tam \dots$$

- Função $rh(n)$: função de sondagem
- Uso da divisão modular (ou equivalente)
- Término da sondagem: posição livre, repetição da mesma sondagem, tabela cheia

Endereçamento aberto

Visão geral

Endereçamento
aberto

Visão geral

Sondagens

Desempenho

Remoções

Operações

Encadeamento

Visão geral

Remoções

Operações

Endereçamento
em baldes

Visão geral

Alternativas de sondagem:

- Sondagem linear
- Sondagem quadrática
- Hash duplo

Visão geral

Endereçamento
aberto

Visão geral

Sondagens

Desempenho

Remoções

Operações

Encadeamento

Visão geral

Remoções

Operações

Endereçamento
em baldes

Visão geral

Sondagem linear

Características gerais:

- Função de sondagem: $rh(i) = i$
- Sondagens a partir da posição k : $k + 1, k + 2, k + 3 \dots$
(cíclica por causa da divisão modular)

Problema:

- Agrupamentos primários: aumento da chance de ocupação de algumas posições em função da ocupação de posições anteriores

Sondagem quadrática:

Características gerais:

- Função de sondagem: $rh(i) = (-1)^{i-1} \left\lfloor \frac{i+1}{2} \right\rfloor^2$
- Sondagens a partir da posição k :
 $k+1, k-1, k+4, k-4, k+9, k-9 \dots$
(cíclica pela divisão modular) (fechada na tabela pela divisão modular)
- Sobre o tamanho da tabela:
 - A tabela não pode ter tamanho par
 - Tamanho ideal: número primo na forma $4j+3, j \in \mathbb{N}$

Problema:

- Agrupamentos secundários: todas as colisões em k seguem a mesma sequência de sondagens

Endereçamento aberto

Visão geral

Endereçamento
aberto

Visão geral

Sondagens

Desempenho

Remoções

Operações

Encadeamento

Visão geral

Remoções

Operações

Endereçamento
em baldes

Visão geral

Hash duplo (ou *double hash*):

- Uso de duas funções:
 - Função de sondagem: $rh(e, c) = e + rh_2(c)$
 - Função de resolução de conflito: $rh_2(c)$
- Dependência do valor da chave que gerou a colisão
- Sondagens a partir da posição k :
 $k + rh_2(c), k + 2rh_2(c), k + 3rh_2(c) \dots$

Endereçamento aberto

Visão geral

Endereçamento
aberto

Visão geral

Sondagens

Desempenho

Remoções

Operações

Encadeamento

Visão geral

Remoções

Operações

Endereçamento
em baldes

Visão geral

Desempenho:

- Dependência da chance de colisões, i.e., da ocupação da tabela
- Fator de carga: $f = \frac{n}{tam}$
 n : número de registros na tabela
 tam : tamanho da tabela

Endereçamento aberto

Visão geral

Endereçamento
aberto

Visão geral

Sondagens

Desempenho

Remoções

Operações

Encadeamento

Visão geral

Remoções

Operações

Endereçamento
em baldes

Visão geral

Desempenho:

Valor esperado para o número de sondagens.

	sondagem linear	sondagem quadrática	hash duplo
sucesso	$\frac{1}{2} \left(1 + \frac{1}{1-f} \right)$	$1 - \log(1-f) - \frac{f}{2}$	$\frac{1}{f} \log \frac{1}{1-f}$
fracasso	$\frac{1}{2} \left(1 + \frac{1}{(1-f)^2} \right)$	$\frac{1}{1-f} - f - \log(1-f)$	$\frac{1}{1-f}$

Endereçamento aberto

Visão geral

Endereçamento
aberto

Visão geral

Sondagens

Desempenho

Remoções

Operações

Encadeamento

Visão geral

Remoções

Operações

Endereçamento
em baldes

Visão geral

Remoções:

- Evitação do término prematuro da busca
- Marcação “disponível ” \neq “vazia”

Endereçamento aberto

Visão geral

Endereçamento
aberto

Visão geral

Sondagens

Desempenho

Remoções

Operações

Encadeamento

Visão geral

Remoções

Operações

Endereçamento
em baldes

Visão geral

Operações:

- Criação: posições vazias e inserção 1 a 1
- Busca: função hash e sondagens
- Inserção: função hash e sondagens
- Remoção: marcação de disponível
- Alteração: remoção e reinserção se houver alteração de chave
- Manutenção: Recriação da tabela

Visão geral

Endereçamento
aberto

Visão geral
Sondagens
Desempenho
Remoções
Operações

Encadeamento

Visão geral
Remoções
Operações

Endereçamento
em baldes

Visão geral

Encadeamento

Visão geral

Endereçamento
abertoVisão geral
Sondagens
Desempenho
Remoções
Operações

Encadeamento

Visão geral
Remoções
OperaçõesEndereçamento
em baldes

Visão geral

Encadeamento:

- Uso de um campo de ligação entre as posições do vetor
- Encadeamento de colisões

Visão geral

Endereçamento
abertoVisão geral
Sondagens
Desempenho
Remoções
Operações

Encadeamento

Visão geral
Remoções
OperaçõesEndereçamento
em baldes

Visão geral

Escolhas para a manutenção da lista encadeada:

- Interna
(nas outras posições livres da tabela)
- Externa
(e.g., uso de alocação dinâmica ou tabela separada)

Visão geral

Endereçamento
aberto

Visão geral
Sondagens
Desempenho
Remoções
Operações

Encadeamento

Visão geral
Remoções
Operações

Endereçamento
em baldes

Visão geral

Considerações sobre encadeamento em lista interna:

- Determinação de uma posição vazia
- Evitação da mistura de listas

Visão geral

Endereçamento
aberto

Visão geral
Sondagens
Desempenho
Remoções
Operações

Encadeamento

Visão geral
Remoções
Operações

Endereçamento
em baldes

Visão geral

Considerações sobre encadeamento em lista externa:

- Tabela contendo apenas ponteiros
- Alocação dinâmica dos nós da lista

Visão geral

Endereçamento
aberto

Visão geral
Sondagens
Desempenho
Remoções
Operações

Encadeamento

Visão geral
Remoções
Operações

Endereçamento
em baldes

Visão geral

Remoções:

- Remoção simples de lista encadeada
- Tratamento específico do primeiro registro da lista

Visão geral

Endereçamento
abertoVisão geral
Sondagens
Desempenho
Remoções
Operações

Encadeamento

Visão geral
Remoções
OperaçõesEndereçamento
em baldes

Visão geral

Operações:

- Criação: posições vazias e inserção 1 a 1
- Busca: função hash e busca no encadeamento
- Inserção: função hash, sondagem/alocação e encadeamento
- Remoção: remoção da lista encadeada
- Alteração: remoção e reinserção se houver alteração de chave
- Manutenção: Recriação da tabela

Visão geral

Endereçamento
aberto

Visão geral
Sondagens
Desempenho
Remoções
Operações

Encadeamento

Visão geral
Remoções
Operações

Endereçamento
em baldes

Visão geral

Endereçamento em baldes

Endereçamento em baldes

Visão geral

Endereçamento
aberto

Visão geral
Sondagens
Desempenho
Remoções
Operações

Encadeamento

Visão geral
Remoções
Operações

Endereçamento
em baldes

Visão geral

Baldes:

- Ampliação de cada posição para conter mais que um registro (“balde de registros”)
- Acomodação de colisões até a capacidade do balde
- Tratamento de baldes cheios
- Exigência de espaço para o armazenamento