

Organização e Recuperação da Informação

Reuso de espaço em arquivos

Jander Moreira

UFSCar – DC

4 de setembro de 2017

Este material é complementar, sendo apenas uma apresentação de aula e não consistindo em material suficiente para estudo.

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Agenda

① Introdução

② Compressão de dados
Reversibilidade③ Reuso de espaço
Introdução

Remoção e compactação
Reuso dinâmico:
tamanho fixo
Reuso dinâmico:
tamanho variável

④ Atividade

Material baseado em
FOLK, M.J & ZOELLICK, B. *File structures*, 2nd ed.
Addison-Wesley Publishing Company, Inc. 1992
Capítulo 5

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Introdução

Introdução

Compressão de
dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e
compactaçãoReuso dinâmico:
tamanho fixoReuso dinâmico:
tamanho variável

Atividade

Créditos

Organização de arquivos para desempenho

- Redução da necessidade de armazenamento (compressão de dados)
- Reuso de espaço
- Uso de ordenação e estruturas auxiliares

Compressão de dados

Compressão de dados

Compressão de dados

A codificação dos dados para reduzir seu tamanho é chamada compressão de dados.

- Substituição da representação atual por outra
- Nova representação exigindo menor número de bytes
- Necessidade de conhecer a nova codificação para retornar ao dado original

Compressão de dados

Técnicas comuns

- Redução de redundância
Eliminação dos bits desnecessários
- Supressão de sequências repetidas
Troca de sequências do mesmo byte pelo par $\langle n, \text{byte} \rangle$
- Códigos de comprimento variável
Uso de número de bits diferentes no novo código¹
- ...

¹E.g., codificação de Huffman

Compressão de dados

Sem perda \times Com perda

Sem perda

- Operação reversível
- Aplicação: documentos, valores de transações

Com perda

- Operação irreversível
- Descompressão obtém uma aproximação do dado original
- Aplicação: imagem, áudio, vídeo

Reuso de espaço

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Operações sobre arquivos

- Criação e população inicial
- **Inserção***
- **Remoção***
- **Alteração/modificação***
- Busca
- **Manutenção****

*Operações com impacto no uso do espaço

**Operação necessária para corrigir deteriorações

Reuso de espaço

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Remoção lógica \times remoção física

Remoção física

Uma remoção é chamada de física quando o espaço liberado por ela é imediatamente eliminado do arquivo, reduzindo seu tamanho



Remoção lógica

A remoção lógica é aquela em que há apenas uma marcação de um registro como inválido, permanecendo ocioso o espaço ocupado pelo registro eliminado



Reuso de espaço

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Compactação

A compactação de um arquivo corresponde à eliminação dos espaços com dados inúteis existentes no arquivo



Espaço ocioso: devido a remoções e atualizações

Compatação: operação de manutenção

Reuso de espaço

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico:
tamanho fixo

Reuso dinâmico:
tamanho variável

Atividade

Créditos

Remoção e compactação

- Marcação de registros como inválidos
- Compactação do arquivo

Reuso de espaço

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico:
tamanho fixo

Reuso dinâmico:
tamanho variável

Atividade

Créditos

Arquivo de dados

(com delimitador de campo e de registro)

```
000 A l g o r i t m o s - _ t e o r i a _ e _  
021 p r a t i c a | C o r m e n , _ T . _ H .  
042 _ a n d _ L e i s e r s o n , _ C . _ E .  
063 _ a n d _ R i v e s t , _ R . _ L . _ a n  
084 d _ S t e i n , _ C . | 2 0 0 2 | # C - _  
105 C o m o _ p r o g r a m a r | D e i t e l  
126 , _ P . _ a n d _ D e i t e l , _ H . | 2  
147 0 1 1 | # E s t r u t u r a _ d e _ d a d  
168 o s _ e _ a l g o r i t m o s _ e m _ C +  
189 + | D r o z d e k , _ A . | 2 0 1 0 | # A  
210 l g o r i t m o s _ e s t r u t u r a d o  
231 s | F a r r e r , _ H . | 1 9 8 9 | #
```

Reuso de espaço

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Após a remoção do terceiro registro
(Estrutura de dados...)

```

000 A l g o r i t m o s - _ t e o r i a _ e _
021 p r a t i c a | C o r m e n , _ T . _ H .
042 _ a n d _ L e i s e r s o n , _ C . _ E .
063 _ a n d _ R i v e s t , _ R . _ L . _ a n
084 d _ S t e i n , _ C . | 2 0 0 2 | # C - _
105 C o m o _ p r o g r a m a r | D e i t e l
126 , _ P . _ a n d _ D e i t e l , _ H . | 2
147 0 1 1 | # * : 5 7 : t u r a _ d e _ d a d
168 o s _ e _ a l g o r i t m o s _ e m _ C +
189 + | D r o z d e k , _ A . | 2 0 1 0 | # A
210 l g o r i t m o s _ e s t r u t u r a d o
231 s | F a r r e r , _ H . | 1 9 8 9 | #
  
```


Reuso de espaço

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Compacatação

(dados do terceiro registro excluídos)

```

000 A l g o r i t m o s - _ t e o r i a _ e _
021 p r a t i c a | C o r m e n , _ T . _ H .
042 _ a n d _ L e i s e r s o n , _ C . _ E .
063 _ a n d _ R i v e s t , _ R . _ L . _ a n
084 d _ S t e i n , _ C . | 2 0 0 2 | # C - _
105 C o m o _ p r o g r a m a r | D e i t e l
126 , _ P . _ a n d _ D e i t e l , _ H . | 2
147 0 1 1 | # A l g o r i t m o s _ e s t r u
168 t u r a d o s | F a r r e r , _ H . | 1 9
189 8 9 | #
  
```

Reuso de espaço

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico:
tamanho fixo

Reuso dinâmico:
tamanho variável

Atividade

Créditos

Quando compactar?

- Depende da aplicação
- Depende das disponibilidade do arquivo para a operação
- Depende do número de registros removidos

Reuso de espaço

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Arquivo de dados

(registros de tamanho fixo – 40 bytes)

```

000 A l g o r i t m o s - _ t e o r i a _ e
020 _ p r a t i c a _ _ _ _ _ _ _ _ 2 0 0 2
040 C - _ C o m o _ p r o g r a m a r _ _ _
060 _ _ _ _ _ _ _ _ _ _ _ _ _ _ 2 0 1 1
080 E s t r u t u r a _ d e _ d a d o s _ e
100 _ a l g o r i t m o s _ _ _ _ _ _ 2 0 1 0
120 A l g o r i t m o s _ e s t r u t u r a
140 d o s _ _ _ _ _ _ _ _ _ _ 1 9 8 9
160 C _ c o m p l e t o _ e _ t o t a l _ _
180 _ _ _ _ _ _ _ _ _ _ _ _ _ _ 1 9 9 1
  
```

Reuso de espaço

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Remoção do segundo registro

(Marcação com caractere especial *)

```

000 A l g o r i t m o s - _ t e o r i a _ e
020 _ p r a t i c a _ _ _ _ _ _ _ 2 0 0 2
040 * - _ C o m o _ p r o g r a m a r _ _ _
060 _ _ _ _ _ _ _ _ _ _ _ _ _ 2 0 1 1
080 E s t r u t u r a _ d e _ d a d o s _ e
100 _ a l g o r i t m o s _ _ _ _ _ 2 0 1 0
120 A l g o r i t m o s _ e s t r u t u r a
140 d o s _ _ _ _ _ _ _ _ _ _ 1 9 8 9
160 C _ c o m p l e t o _ e _ t o t a l _ _
180 _ _ _ _ _ _ _ _ _ _ _ _ _ 1 9 9 1

```

Como reaproveitar o espaço para uma nova inserção antes de uma compactação?

Arquivos com registros de tamanho fixo

Rastreamento das remoções:

- Manutenção de lista encadeada de registros removidos
 - Ponteiro externo para a posição do primeiro registro da lista
 - Cada registro removido mantém um ponteiro para o próximo

Reuso de espaço I

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Arquivo original

```

000 A l g o r i t m o s - _ t e o r i a _ e
020 _ p r a t i c a _ _ _ _ _ 2 0 0 2
040 C - _ C o m o _ p r o g r a m a r _ _ _
060 _ _ _ _ _ _ _ _ _ _ _ 2 0 1 1
080 E s t r u t u r a _ d e _ d a d o s _ e
100 _ a l g o r i t m o s _ _ _ _ _ 2 0 1 0
120 A l g o r i t m o s _ e s t r u t u r a
140 d o s _ _ _ _ _ _ _ _ _ 1 9 8 9
160 C _ c o m p l e t o _ e _ t o t a l _ _
180 _ _ _ _ _ _ _ _ _ _ _ 1 9 9 1

```

Reuso de espaço II

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Remoção do segundo registro ("C- Como programar")

```

000 A l g o r i t m o s - _ t e o r i a _ e
020 _ p r a t i c a _ _ _ _ _ 2 0 0 2
040 * : - 1 : m o _ p r o g r a m a r _ _ _
060 _ _ _ _ _ _ _ _ _ _ _ 2 0 1 1
080 E s t r u t u r a _ d e _ d a d o s _ e
100 _ a l g o r i t m o s _ _ _ _ _ 2 0 1 0
120 A l g o r i t m o s _ e s t r u t u r a
140 d o s _ _ _ _ _ _ _ _ _ 1 9 8 9
160 C _ c o m p l e t o _ e _ t o t a l _ _
180 _ _ _ _ _ _ _ _ _ _ _ 1 9 9 1
  
```

Início da lista: 40

Reuso de espaço III

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Remoção do quinto registro ("C completo e total")

```

000 A l g o r i t m o s - _ t e o r i a _ e
020 _ p r a t i c a _ _ _ _ _ 2 0 0 2
040 * : - 1 : m o _ p r o g r a m a r _ _ _
060 _ _ _ _ _ _ _ _ _ _ _ 2 0 1 1
080 E s t r u t u r a _ d e _ d a d o s _ e
100 _ a l g o r i t m o s _ _ _ _ _ 2 0 1 0
120 A l g o r i t m o s _ e s t r u t u r a
140 d o s _ _ _ _ _ _ _ _ _ 1 9 8 9
160 * : 4 0 : p l e t o _ e _ t o t a l _ _
180 _ _ _ _ _ _ _ _ _ _ _ 1 9 9 1

```

Início da lista: 160

Reuso de espaço

Reuso em arquivos com registros de tamanho fixo

- Manutenção de lista encadeada de áreas livres (registros removidos)
- Organização da lista como uma **pilha**
- Novas inserções usam os espaços guardados na pilha
- Inserção no fim do arquivo quando a pilha está vazia

- Possível substituição do endereço pelo RNN²

²*Relative record number*, ou número do registro

Reuso de espaço

Arquivos com registros de tamanho variável

Rastreamento das remoções:

- Manutenção de lista encadeada de registros removidos
 - Ponteiro externo para a posição do primeiro registro da lista
 - Cada registro removido mantém um ponteiro para o próximo

Igual a registros de tamanho fixo

Reuso de espaço I

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Registros de tamanho variável (com delimitadores)

```

000 A l g o r i t m o s - _ t e o r i a _ e
020 _ p r a t i c a | C o r m e n , _ T . _
040 H . _ a n d _ L e i s e r s o n , _ C .
060 _ E . _ a n d _ R i v e s t , _ R . _ L
080 . _ a n d _ S t e i n , _ C . | 2 0 0 2
100 | # C - _ C o m o _ p r o g r a m a r |
120 D e i t e l , _ P . _ a n d _ D e i t e
140 l , _ H . | 2 0 1 1 | # E s t r u t u r
160 a _ d e _ d a d o s _ e _ a l g o r i t
180 m o s _ e m _ C + + | D r o z d e k , _
200 A . | 2 0 1 0 | # A l g o r i t m o s _
220 e s t r u t u r a d o s | F a r r e r ,
240 _ H . | 1 9 8 9 | #
  
```

Reuso de espaço II

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Remoção do quarto registro (“Algoritmos estruturados”)

```

000 A l g o r i t m o s - _ t e o r i a _ e
020 _ p r a t i c a | C o r m e n , _ T . _
040 H . _ a n d _ L e i s e r s o n , _ C .
060 _ E . _ a n d _ R i v e s t , _ R . _ L
080 . _ a n d _ S t e i n , _ C . | 2 0 0 2
100 | # C - _ C o m o _ p r o g r a m a r |
120 D e i t e l , _ P . _ a n d _ D e i t e
140 l , _ H . | 2 0 1 1 | # E s t r u t u r
160 a _ d e _ d a d o s _ e _ a l g o r i t
180 m o s _ e m _ C + + | D r o z d e k , _
200 A . | 2 0 1 0 | # * : - 1 : 4 1 : o s _
220 e s t r u t u r a d o s | F a r r e r ,
240 _ H . | 1 9 8 9 | #

```

Reuso de espaço III

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Remoção do primeiro registro (“Algoritmos- teoria e pratica”)

```

000 * : 2 0 9 : 1 0 2 : - _ t e o r i a _ e
020 _ p r a t i c a | C o r m e n , _ T . _
040 H . _ a n d _ L e i s e r s o n , _ C .
060 _ E . _ a n d _ R i v e s t , _ R . _ L
080 . _ a n d _ S t e i n , _ C . | 2 0 0 2
100 | # C - _ C o m o _ p r o g r a m a r |
120 D e i t e l , _ P . _ a n d _ D e i t e
140 l , _ H . | 2 0 1 1 | # E s t r u t u r
160 a _ d e _ d a d o s _ e _ a l g o r i t
180 m o s _ e m _ C + + | D r o z d e k , _
200 A . | 2 0 1 0 | # * : - 1 : 4 1 : o s _
220 e s t r u t u r a d o s | F a r r e r ,
240 _ H . | 1 9 8 9 | #
  
```

Início da lista: 0

Exercício

Atividade

- Converse com o colega ao lado (2 ou 3 no grupo)
- Responda
 - Quando um novo registro **não** pode ser inserido no espaço de um dado registro removido?
 - Pode haver desperdício de espaço na inserção. Em que situação?

Reuso de espaço

Reuso em arquivos com registros de tamanho variável

- Manutenção de lista encadeada de áreas livres (registros removidos)
- Organização da lista depende de uma estratégia
- Novas inserções usam os espaços apenas quando o novo registro couber nos espaços
- Inserção no fim do arquivo quando não der para usar nenhum espaço da lista

Diferente de registros de tamanho fixo

Exercício

Esta atividade é dinâmica e deve ser concluída em 10 minutos.

Terminada, o professor coordenará um fechamento para a discussão.

Em pequenos grupos (tamanho livre)

- Discutam **uma única alternativa** como usar a lista de espaços disponíveis para o caso de um novo registro a ser inserido
- Levantem os prós e os contras da para essa alternativa

Exercício

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Pontos importantes:

- Qualquer registro da lista pode ser usado para a nova inserção?
- Há sempre um item da lista que seja exatamente do tamanho do novo registro?
- O que acontece quando não há nenhum registro de tamanho suficiente na lista de espaço disponível?
- O que ocorre com a diferença entre a área usada para a nova inserção e o espaço real usado?
- A ordem dos registros na lista influencia o desempenho na escolha do espaço?

Reuso de espaço

Fragmentação

A fragmentação do espaço de um arquivo ocorre quando é usado espaço no qual não há dado útil

- **Fragmentação interna:** espaço não usado *dentro* de um registro
Em registros de tamanho físico, por exemplo
- **Fragmentação externa:** espaço não usado *fora* dos registros que se torna não usável
Na lista de disponíveis, por exemplo

Reuso de espaço

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Estratégias de escolha

- *First-fit*
Escolha do primeiro espaço com tamanho suficiente
- *Best-fit*
Escolha do espaço com tamanho suficiente mais próximo do novo tamanho
- *Worst-fit*
Escolha do espaço com tamanho suficiente para o qual sobre o maior espaço

Reuso de espaço

Introdução

Compressão de dados

Reversibilidade

Reuso de espaço

Introdução

Remoção e compactação

Reuso dinâmico: tamanho fixo

Reuso dinâmico: tamanho variável

Atividade

Créditos

Combate à fragmentação interna

- Registros de tamanho fixo: não há
- Registros de tamanho variável: transformação em fragmentação externa

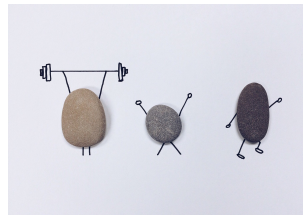
Combate à fragmentação externa

- Compactação do arquivo
- Coalescência de espaços fisicamente adjacentes

Atividade

Exercícios

A realização da lista de exercícios disponibilizada é essencial para a boa compreensão desta aula e correta absorção do conteúdo.



Créditos

Créditos

Jander Moreira

<http://www.dc.ufscar.br/~jander>

jander@dc.ufscar.br

Universidade Federal de São Carlos

<http://www.ufscar.br>

Departamento de Computação

<http://www.dc.ufscar.br>