

# Classificação e Pesquisa de Dados

## Aula 01

Apresentação da Disciplina; Introdução à Classificação; Definições; Formas de Representação de Resultados

UFRGS

INF01124

1

## Histórico do material

1. Preparado inicialmente pelo prof. Paulo Alberto Azeredo em forma textual
2. Adaptado para apresentação pela profa. Nina Edelweiss
3. Transformado para PowerPoint e aperfeiçoado pela Profa. Maria Aparecida M. Souto
4. Revisado e reformulado pelos prof. José Palazzo M. de Oliveira, profa. Daniela Remião de Macedo
5. Segunda revisão e reformulação pelo prof. Manuel Menezes de Oliveira Neto

2

## Objetivos da disciplina

- ◆ Capacitar o aluno para seleção e análise de:
  - Algoritmos para classificação de dados
  - Algoritmos para pesquisa de dados
  - Técnicas de organização e de compactação de arquivos
- ◆ Estruturas de dados consideradas
  - em memória vs. em disco

3

## Súmula

1. Métodos de Classificação de Dados
2. Introdução à Análise de Complexidade de Algoritmos
3. Métodos de Armazenamento e Pesquisa de Dados em Tabelas
4. Técnicas de Organização de Arquivos
5. Técnicas de Compactação de Arquivos

4

## Bibliografia

AZEREDO, P.A. Métodos de Classificação de Dados e Análise de suas Complexidades. Editora Campus, RJ, 1995.  
KNUTH, D.: The Art of Computer Programming: Sorting and Searching. Vol. 2. Addison-Wesley, Reading, Mass, 1973.  
CORMEN, T., LEISERSON, C. E RIVEST, R.: Introduction to Algorithms. The MIT Press. Cambridge, Massachusetts, 1990.  
SANTOS, C.S. e AZEREDO, P. A. Tabelas: Organização e Pesquisa. Série Livros Didáticos, Editora Sagra Luzzato, Porto Alegre, 2001.  
VELOSO, P.A.S; SANTOS, C.S; AZEREDO, P.A; FURTADO, A.L.: Estruturas de Dados. Editora Campus, Rio de Janeiro, 1985.  
FURTADO, A.L. e SANTOS, C.S. dos. Organização de Banco de Dados. Editora Campus, Rio de Janeiro, 1988.  
SZWARCFITER, Jayme L. e MARKENZON, Lilian. Estrutura de Dados e seus algoritmos. Rio de Janeiro: LTC, 1994.

5

## Classificação de Dados

UFRGS

INF01124

6

## Classificação (Sorting)

- ◆ Processo de organizar itens em ordem (de)crescente, segundo algum critério
- ◆ Também chamado de ordenação
- ◆ Aplicações de Sorting
  - Preparação de dados para facilitar pesquisas futuras
    - Exemplo: dicionários e listas telefônicas
  - Agrupar itens que apresentam mesmos valores
    - Para eliminação de elementos repetidos
  - Batimento entre itens presentes em mais de um arquivo
    - Para combinação de dados presentes nos vários arquivos
    - Para consolidação dos vários arquivos em um único

7

## Definições

- ◆ Sejam  $R_1, R_2, \dots, R_N$ ,  $N$  itens (chamados registros)
- ◆ Cada registro  $R_i$  é formado por uma chave  $C_i$  e por outros dados ditos satélites
- ◆ A ordenação dos registros é feita definindo-se uma relação de ordem " $<$ " sobre os valores das chaves
- ◆ O objetivo da ordenação é determinar uma permutação dos índices  $1 \leq i_1, i_2, \dots, i_N \leq N$  das chaves, tal que  $C_{i1} \leq C_{i2} \leq \dots \leq C_{iN}$ .
- ◆ Um conjunto de registros é chamado de arquivo

8

## Relação de Ordem

- ◆ Uma relação de ordem " $<$ " (leia-se precede) deve satisfazer as seguintes condições para quaisquer valores  $a, b$  e  $c$ :
  - (i) Uma e somente uma das seguintes possibilidades é verdadeira:  $a < b$ ,  $a = b$  ou  $b < a$  (lei da tricotomia)
  - (ii) Se  $a < b$  e  $b < c$ , então  $a < c$  (transitividade)
- ◆ As propriedades (i) e (ii) definem o conceito de ordem linear ou ordem total

9

## Mais Definições

- ◆ Um algoritmo de classificação é dito estável, se ele preserva a ordem relativa original dos registros com mesmo valor de chave.
- ◆ Classificação local: feita sobre a mesma área física onde se encontram as chaves (não há necessidade de memória extra).
- ◆ Algoritmos de ordenação podem ser classificados como internos (todos os registros mantidos em RAM) ou externos.

10

## Pergunta

- ◆ Qual a importância prática de algoritmos de classificação estável? Exemplifique.
  - Controle de filas de prioridades em sistemas operacionais.

11

## Exercício

- ◆ Escreva um algoritmo para classificar um conjunto de 5 números em ordem crescente.

12

## Formas de Representação do Resultado

- Reorganização Física
- Encadeamento
- Vetor Indireto de Ordenação (VIO)

13

## Classificação de Dados com Reorganização Física

	chave	dados satélites		chave	dados satélites
1	10		1	7	
2	19		2	10	
3	13		3	12	
4	12		4	13	
5	7		5	19	

Antes da classificação

Após a classificação

14

## Classificação de Dados através de Encadeamento

cabeça da lista

	chave	dados satélites		chave	dados satélites
1	10		1	10	4
2	19		2	19	0
3	13		3	13	2
4	12		4	12	3
5	7		5	7	1

Antes da classificação

Após a classificação

- Somente acesso sequencial aos registros ordenados!

15

## Classificação de Dados por Vetor Indireto de Ordenação (VIO)

	chave	dados satélites		chave	VIO
1	10		1	7	5
2	19		2	10	1
3	13		3	12	4
4	12		4	13	3
5	7		5	19	2

- Acesso sequencial, por pesquisa binária e direto, mas sempre por via indireta

16

## Métodos de Classificação de Dados

- Classificação por Inserção
- Classificação por Trocas
- Classificação por Seleção
- Classificação por Distribuição de Chaves
- Classificação por Intercalação
- Classificação por Cálculo de Endereços

17

18