

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
DEPARTAMENTO DE INFORMÁTICA APLICADA

DISCIPLINA: CLASSIFICAÇÃO E PESQUISA DE DADOS - Semestre 2004/1
CÓDIGO: INF01124
PRÉ-REQUISITO: INF01120 - Técnicas de Construção de Programas
CARGA HORÁRIA: 4 h/s
NATUREZA DAS AULAS: Expositivas
CRÉDITOS: 4
PROFESSOR: MANUEL MENEZES DE OLIVEIRA NETO

SÚMULA

Métodos de Classificação de Dados. Introdução à Análise da Complexidade de Algoritmos. Métodos de Armazenamento e Pesquisa de Dados em Tabelas. Técnicas de Organização de Arquivos e Compactação de Arquivos.

OBJETIVOS

Capacitar o aluno na seleção e análise de algoritmos de classificação de dados, de pesquisa de dados em tabelas e de técnicas de organização de arquivos e compactação de arquivos.

CONTEÚDO PROGRAMÁTICO

- 1 - Métodos de Classificação de Dados.
 - 1.1 - Introdução.
 - 1.2 - Formas de apresentação do resultado de uma classificação.
 - 1.3 - Famílias de métodos de classificação.
 - 1.4 - Método da Inserção Direta.
 - 1.5 - Método dos Incrementos Decrescentes (Shellsort).
 - 1.6 - Método da Bolha (Bubblesort) e suas variações .
 - 1.7 - Método da Partição e Troca (Quicksort).
 - 1.8 - Método da Seleção Direta.
 - 1.9 - Método da Seleção em Árvore (Heapsort).
 - 1.10 - Método da Distribuição de Chave.
 - 1.11 - Método da Intercalação (Mergesort).
 - 1.12 - Método de Cálculo de Endereços (Hashsort).
 - 1.13 - Intercalação de arquivos classificados.
- 2 - Introdução à Análise da Complexidade de Algoritmos.
 - 2.1 - Complexidade dos algoritmos de classificação.
 - 2.2 - Complexidade dos algoritmos sobre pesquisa de tabelas.
- 3 - Pesquisa de Dados em Tabelas.
 - 3.1 - Introdução.
 - 3.2 - Pesquisa Sequencial.
 - 3.3 - Pesquisa Binária.
 - 3.4 - Cálculo de Endereços.
 - 3.4.1 - Introdução.
 - 3.4.2 - Tratamento de Colisões.
 - 3.4.2.1 - Endereçamento Aberto.

- 3.4.2.1.1 - com Busca Linear.
- 3.4.2.1.2 - com Realeatorização.
- 3.4.2.1.3 - Desempenho.
- 3.4.2.2 - Encadeamento.
- 3.4.2.2.1 - Com Tabela de Descritores.
- 3.4.2.2.2 - Com Encadeamento Direto.
em Área Separada.
com Fusão de Listas.
- 3.4.2.3 - Alocação em Blocos.
- 3.5 - Hashing Dinâmico.
- 3.5.1 - Introdução.
- 3.5.2 - Hashing Estendível.
- 3.5.3 - Realocação de tabelas.
- 3.6 - Funções de Aleatorização.
- 3.6.1 - Compressão de Chaves Alfanuméricas.
- 3.6.2 - Método da Divisão.
- 3.6.3 - Método da multiplicação.
- 3.6.4 - Método do Quadrado.
- 3.6.5 - Método da Dobradura.
- 3.6.6 - Método da conversão de base.
- 3.6.7 - Funções Dependentes da Distribuição.
- 3.7 - Pesquisa em árvore.
- 3.7.1 - Árvore binária de pesquisa.
- 3.7.2 - Árvores-B.

4 - Organização de Arquivos.

- 4.1 - Arquivos seqüenciais.
- 4.2 - Arquivos indexados.
- 4.3 - Arquivos invertidos.
- 4.4 - Arquivos de texto.

5 - Compactação de Arquivos.

- 5.1 - Introdução.
- 5.2 - Racionalização do projeto de arquivos.
- 5.3 - Supressão de caracteres repetidos.
- 5.4 - Codificação de itens.
- 5.5 - Compressão de dados alfanuméricos.
- 5.5.1 Huffman.
- 5.7 - Compressão de dados multimídia.

PROCEDIMENTOS DIDÁTICOS

As aulas serão de natureza expositiva, utilizando-se slides e quadro. Serão utilizadas listas de exercícios e realizados trabalhos de implementação e um projeto final.

SISTEMA DE AVALIAÇÃO

O aluno será avaliado com base no desempenho nas provas, trabalhos e no projeto final, bem como por sua participação em aula. As provas, trabalhos e projeto final serão avaliados com nota entre 0.0 e 10.0. Conforme regulamento da Universidade, a frequência às aulas é obrigatória.

Ao longo do semestre, serão realizados:

- i. Duas provas, P1, na metade do semestre, e P2, prova final. P1 corresponderá a 25% da nota final; P2, a 35% da nota final;
- ii. Pelo menos dois trabalhos teórico-práticos (TPs) envolvendo exercícios e/ou implementação de algoritmos. A soma de todos os TPs corresponderá a 15% da nota final;
- iii. Um projeto final da disciplina, a ser realizado em grupos de dois estudantes, representando 20% da nota final.

Além disso, será atribuída nota pela participação (NP) em sala de aula, o que representará 5% da nota final.

Trabalhos e exercícios em sala de aula poderão ser realizados sem aviso prévio, e serão avaliados para compor parte da TP e, conseqüentemente, da nota final. A realização dos trabalhos é obrigatória, mesmo que o aluno obtenha bons resultados nas provas.

A média geral (MG) será obtida por meio da seguinte fórmula:

$$MG = 0,25 * P1 + 0,35 * P2 + 0,15 * TP + 0,2 * PF + 0,05 * NP$$

A conversão da MG para conceitos é feita por meio da seguinte tabela:

9,0 ≤ MG = 10,0 : conceito A (aprovado).
7,5 ≤ MG < 9,0 : conceito B (aprovado).
6,0 ≤ MG < 7,5 : conceito C (aprovado).
4,0 ≤ MG < 6,0 : sem conceito (recuperação).
0,0 ≤ MG < 4,0 : conceito D (reprovado).

Observações

1 - Somente serão calculadas as médias gerais daqueles alunos que tiverem, ao longo do semestre, obtido um índice de freqüência às aulas igual ou superior a 75 % das aulas previstas. Aos que não satisfizerem este requisito, será atribuído o conceito FF (Falta de Freqüência).

2 - Para poder realizar a prova de recuperação, o aluno deve ter realizado as duas provas (P1 e P2), ter entregue pelo menos dois dos três trabalhos práticos (TPs) e o projeto final (PF). Os que não se enquadrarem nesta situação receberão conceito D.

RECUPERAÇÃO

Os alunos cujas médias gerais forem inferiores a 6,0 (seis) e maiores ou iguais a 4,0 (quatro) e que satisfizerem as condições 1 e 2 acima, poderão prestar prova de recuperação, a qual versará sobre toda a matéria da disciplina.

Serão considerados aprovados na recuperação os alunos que obtiverem um aproveitamento de no mínimo 60 % da prova. A estes será atribuído o conceito C. Aos demais, o conceito D.

Não há recuperação das provas P1 e P2 por não comparecimento, exceto nos casos previstos na legislação (saúde, parto, serviço militar, convocação judicial, luto etc, devidamente comprovados).

BIBLIOGRAFIA

- 1 - AZEREDO, P.A. Métodos de Classificação de Dados e Análise de suas Complexidades. Editora Campus, RJ, 1995.
- 2 - KNUTH, D.: The Art of Computer Programming: Sorting and Searching. 2nd Edition. Vol. 3. Addison-Wesley, Reading, Mass, 1998.
- 3 - CORMEN, T., LEISERSON, C. E RIVEST, R.: Introduction to Algorithms. The MIT Press. Cambridge, Massachusetts, 1990.
- 4 - SANTOS, C.S. e AZEREDO, P. A. Tabelas: Organização e Pesquisa. Série Livros Didáticos, Editora Sagra Luzzato, Porto Alegre, 2001.
- 5 - VELOSO, P.A.S; SANTOS, C.S; AZEREDO, P.A; FURTADO, A.L.: Estruturas de Dados. Editora Campus, Rio de Janeiro, 1985.
- 6 - FURTADO, A.L. e SANTOS, C.S. dos. Organização de Banco de Dados. Editora Campus, Rio de Janeiro, 1988.
- 7 - SZWARCFITER, Jayme L. e MARKENZON, Lilian. Estrutura de Dados e seus Algoritmos. Rio de Janeiro: LTC, 1994.