

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

DISCIPLINA: CLASSIFICAÇÃO E PESQUISA DE DADOS - Semestre 2009/1
CÓDIGO: INF01124
PRÉ-REQUISITO: INF01203 – Estruturas de Dados
CARGA HORÁRIA: 4 h/s
ETAPA ACONSELHADA: 3º Semestre
NATUREZA DAS AULAS: Expositivas
CRÉDITOS: 4
PROFESSORES: JOSÉ PALAZZO MOREIRA DE OLIVEIRA (TURMA A)
MANUEL MENEZES DE OLIVEIRA NETO (TURMA B)
JOSÉ VALDENI DE LIMA e LEANDRO K. WIVES (TURMA C)

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 Seqüencial.
 - 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 - Funções Dependentes da Distribuição.
- 3.7 - Pesquisa em árvore.
 - 3.7.1 - Árvore binária de pesquisa.
 - 3.7.2 - Árvores digitais.
 - 3.7.3 - Árvores-B.

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

- 5 - Compressão 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 seqüências.
 - 5.6 - Códigos de comprimento variável.
 - Huffman.
 - 5.7 – Compressão de Dados Multimídia: Compressão de Imagens

PROCEDIMENTOS DIDÁTICOS

As aulas serão de natureza expositiva, utilizando-se canhão de projeção, slides e quadro. Será realizado um projeto/trabalho final envolvendo especificação e implementação. Até 20% das aulas poderão ser realizadas em ambiente de Ensino a Distância (Moodle). Complementarmente, serão utilizadas listas de exercícios, trabalhos e atividades para auto-estudo e avaliação em sala de aula ou no ambiente Moodle (extra classe).

Cópias das transparências apresentadas em aula serão disponibilizadas (quatro por página) em arquivos PDF após as aulas, no ambiente Moodle. Esse material deve ser utilizado como auxílio ao estudo e como complementação aos apontamentos realizados em aula. Espera-se

que os alunos tenham uma participação ativa nas aulas, realizando anotações, procurando compreender os conteúdos e, através de perguntas, resolvam suas dúvidas.

SISTEMA DE AVALIAÇÃO

O aluno será avaliado com base no desempenho nas provas, exercícios, trabalhos, atividades e no trabalho final, bem como por sua participação em aula. As provas, trabalhos, exercícios, atividades 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 considerados para a avaliação:

- i. Duas provas: P1, na metade do semestre, e P2, como prova final. P1 corresponderá a 25% da nota final; P2 a 35% da nota final;
- ii. Listas de exercícios e tarefas (LET) serão disponibilizados em aula e/ou no ambiente Moodle, e serão contabilizados. A soma de todos os exercícios e tarefas corresponde a 15% da nota final;
- iii. Um trabalho final (TF), a ser realizado em grupos de até dois estudantes, representando 20% da nota final;
- iv. Além disso, será atribuída nota pela participação (NP) em sala de aula, calculada em função da percentagem de aulas assistidas, o que representará 5% da nota final;
- v. A falta de mais do que 25% das aulas caracteriza o conceito final FF.

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

$$MG = 0,25 * P1 + 0,35 * P2 + 0,20 * TF + 0,15 * LET + 0,05 * NP$$

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), podendo passar para conceito C em caso de aprovação ou conceito D em caso de reprovação na recuperação.

0,0 ≤ MG < 4,0 : conceito D (reprovado).

Faltas > 25% : conceito FF (reprovado).

Observações

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

2 - Para poder realizar a prova de recuperação, o aluno deve ter realizado as duas provas (P1 e P2), ter entregado o trabalho final (TF) e ter realizado mais de 2/3 das listas de exercícios e tarefas (LET). Além disso, o(a) aluno(a) deverá ter nota igual ou superior a 6,0 em pelo menos uma das duas provas. Os que não se enquadrarem nessa situação receberão conceito D.

3 - O TF será desenvolvido em diversas etapas, a serem estabelecidas no início das aulas e disponibilizadas no Moodle. O não cumprimento das datas de entrega do TF, nas suas diferentes etapas, implica em nota 0.0 (zero) para essa atividade.

4 – As notas das provas e trabalhos serão disponibilizadas em até 15 dias depois de sua realização.

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.

CRONOGRAMA TENTATIVO DE ATIVIDADES

Aula	Conteúdo
1	Apresentação da Disciplina; Introdução à Classificação; Definições
2	Sistemas de Arquivos, Algoritmos em Memória e em Disco, Avaliação de Desempenho
3	Classes de Métodos de Classificação. Algoritmos em memória e em disco: avaliação de desempenho
4	Introdução à Análise de Complexidade de Algoritmos; Notações Theta, O e Omega.
5	Classificação de dados por Inserção : Direta, com Busca Binária; Métodos dos Incrementos Decrescentes
6	Classificação de dados por Troca : Bubblesort
7	Classificação de dados por Troca: QuickSort
8	Classificação de dados por Troca: QuickSort (Continuação)
9	Classificação de dados por Seleção: Seleção Direta e Heapsort
10	Classificação de dados por Seleção : Heapsort (Continuação)
11	Classificação por Intercalação: MergeSort em memória principal e secundária
12	Classificação em Tempo Linear $O(n)$: Counting Sort, Radix Sort e Bucket Sort
13	Funções de Cálculo de Endereço e Tratamento de Colisões
14	Redimensionamento de Tabelas e Hashing Dinâmico; Funções Dependentes de Distribuição
15	1a Avaliação (P1)
16	Discussão sobre os Trabalhos Finais
17	Pesquisa de Dados em Tabelas; Operações sobre Tabelas
18	Árvores Binárias de Pesquisa
19	Árvores AVL
20	Árvores B e B+
21	Árvores Patrícias
22	Organização de Arquivos: Arquivos Sequenciais e Arquivos Sequenciais Indexados
23	Organização de Arquivos: Arquivos Indexados, Diretos e Arquivos Invertidos
24	Racionalização de Projetos de Arquivos: Introdução a Normalização
25	Compressão de Arquivos: Méodos de Compressão e Codificação de Huffman
26	Compressão de Dados Multimidia: Compressão de Imagens
27	Revisão para Prova Final
28	2ª Avaliação (P2)
29	Apresentação de Trabalhos Finais - Parte I
30	Apresentação de Trabalhos Finais - Parte II