

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

DISCIPLINA:	CLASSIFICAÇÃO E PESQUISA DE DADOS
SEMESTRE:	2008/1
CÓDIGO:	INF01124
PRÉ-REQUISITO:	INF01203 – Estruturas de Dados
CARGA HORÁRIA:	4 h/s
NATUREZA DAS AULAS:	Expositivas dialogadas
CRÉDITOS:	4
PROFESSOR:	José Palazzo Moreira de Oliveira (Turma A, segundas e quartas às 13:30)

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 e implementação 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 e quadro. Serão disponibilizadas listas de exercícios para auto-estudo e auto-avaliação e será realizado trabalho especificação e implementação em um projeto final.

Cópias das transparências apresentadas em aula serão disponibilizadas (quatro por página) em arquivos PDF após as aulas. Este 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 e procurando compreender os conteúdos e, através de perguntas, resolverem as suas dúvidas.

SISTEMA DE AVALIAÇÃO

O aluno será avaliado com base no desempenho nas provas, participação em trabalhos desenvolvidos ao longo do semestre e no projeto final, bem como por sua participação em aula. As provas e o projeto

final serão avaliados com nota entre 0.0 e 10.0, os trabalhos durante o semestre serão contabilizados. 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, prova final sobre toda a matéria do semestre. P1 corresponderá a 30% da nota final e P2 a 40% da nota final;
- ii. Um trabalho prático (TP) em um total de 20% da nota final;
- iv. A resolução das listas de exercícios (LE) serão contabilizadas (número de exercícios resolvidos entregues) e não avaliadas. O objetivo é o auto-estudo e a auto-avaliação por parte dos alunos. Exercícios (outro componente da LE) adicionais poderão ser solicitados, total de 5% da nota final
- v. 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.
- vi. A falta de mais do que 25% das aulas caracteriza o conceito final FF.

A realização do trabalho prático é obrigatória, mesmo que o(a) aluno(a) obtenha bons resultados nas provas. A média geral (MG) será obtida por meio da seguinte fórmula:

$$MG = 0,30 * P1 + 0,40 * P2 + 0,20 * TP + 0,05 * LE + 0,05 * NP$$

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
0,0 = MG < 4,0 : conceito D (reprovado)
Não entrega do trabalho prático (completo) : 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 este requisito, será atribuído o conceito FF (Falta de Frequência).

2 - A entrega do trabalho prático, isto inclui cada uma das etapas previstas e a entrega do artigo final completo e consistente implica em reprovação, independentemente do resultado das provas.

3 - Para poder realizar a prova de recuperação, o aluno deve ter realizado as duas provas (P1 e P2), ter mais de 50% de participação nas listas de exercícios (LE) e ter entregue o trabalho individual completo. 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 nesta situação terão 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 três condições 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 será atribuído 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

Métodos de Classificação de Dados e Análise de suas Complexidades, Paulo A. Azeredo, Editora Campus, RJ, 1995.

Tabelas: Organização e Pesquisa, Clésio S. Santos e Paulo A. Azeredo, Série Livros Didáticos, Editora Sagra Luzzato, Porto Alegre, 2001.

Organização de Banco de Dados, Antônio L. Furtado e Clésio S. Santos e Paulo A. Azeredo, Editora Campus, Rio de Janeiro, 1988.

LEITURA COMPLEMENTAR

KNUTH, D.: The Art of Computer Programming: Sorting and Searching. 2nd Edition. Vol. 3. Addison-Wesley, Reading, Mass, 1998.

CORMEN, T., LEISERSON, C. E RIVEST, R.: Introduction to Algorithms. The MIT Press. Cambridge, Massachusetts, 1990.

Material disponível na Web principalmente o material disponível na Wikipedia e suas referências (o aluno deve desenvolver espírito crítico no estudo deste material devido a edição cooperativa do mesmo).

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

DISCIPLINA: CLASSIFICAÇÃO E PESQUISA DE DADOS - Semestre 2008/1

CÓDIGO: INF01124

PRÉ-REQUISITO: INF01203 – Estruturas de Dados

CARGA HORÁRIA: 4 h/s

NATUREZA DAS AULAS: Expositivas

CRÉDITOS: 4

PROFESSORES: MANUEL MENEZES DE OLIVEIRA NETO (TURMA B)

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 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 até 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:

$$NP \cdot PF \cdot TP \cdot P \cdot P \cdot MG \cdot 05 \cdot 0 \cdot 2 \cdot 0 \cdot 15 \cdot 0 \cdot 2 \cdot 35 \cdot 0 \cdot 1 \cdot 25 \cdot 0 + + + + =$$

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 frequência às aulas igual ou superior a 75 % das aulas previstas. Aos que não satisfizerem este 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 entregue pelo menos dois dos três trabalhos práticos (TPs) e o projeto final (PF). 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 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.