

# Organização de arquivos: Outras técnicas

SCC0607 – Estrutura de Dados III

Anderson Canale Garcia

Baseado no material dos professores:

Leandro C. Cintra

M.C.F de Oliveira

Rosane Minghim

# Sumário

- Modelos abstratos de dados
- Metadados
- Compressão de dados
- Compactação

# Modelos abstratos de dados

- Objetivo: focar no conteúdo da informação, ao invés do seu formato físico
- Informações atuais tratadas pelos computadores (som, imagens, etc.)
  - Não se ajustam bem à metáfora de dados armazenados como sequências de registros separados em campos
  - Melhor pensar como objetos que representam som, imagens, etc.
    - Maneiras próprias para manipulação

# Modelos abstratos de dados

- Termo: modelo abstrato de dados
  - Dado não precisa ser visto da forma como está armazenado
  - Visão dos dados orientada à aplicação, e não ao meio de armazenamento
- Quais as implicações?

# Sumário

- Modelos abstratos de dados
- **Metadados**
- Compressão de dados
- Compactação

# Metadados

- São dados que descrevem os dados primários em um arquivo
- Geralmente segue um padrão "acordado" pela comunidade de usuários
- Exemplos
  - FITS (Flexible Image Transport System)
    - Imagens de astronomia
    - Utiliza o formato ASCII para o cabeçalho e o formato binário para os dados primários
  - Dublic Core e MARC
    - Usado em bibliotecas e repositórios digitais

# Metadados

- Vantagens
  - Conteúdo "auto-explicável"
    - Torna viável acesso por terceiros
  - Portabilidade
    - Define-se um padrão para o tipo de arquivo e todos geram/acessam da mesma forma
    - PDF, PS, HTML, TIFF...
  - Permite conversão entre padrões
- Tem desvantagem?

# Sumário

- Modelos abstratos de dados
- Metadados
- **Compressão de dados**
- Compactação



# Compressão de dados

- Envolve a codificação da informação de modo que o arquivo ocupe menos espaço
  - Transmissão mais rápida
  - Processamento sequencial mais rápido
- Técnicas gerais ou específicas (p. ex.: voz, imagem ou texto)
- Técnicas reversíveis *vs.* Irreversíveis
- Variedade enorme de técnicas

# Compressão de dados – algumas técnicas

- Redução de redundância
- Omissão de sequências repetidas
- Códigos de tamanho variável

# Redução de redundância

- Exemplo:
  - Códigos de estado, armazenados na forma de texto: 2 bytes
  - Mas como existem 27 estados (Brasil), pode-se armazenar os estados em 5 bits
  - Pode-se, então, guardar a informação em 1 byte e economizar 50% do espaço
- Desvantagens?
  - Legibilidade, codificação/decodificação...

# Omissão de sequências repetidas

- Técnica que omite ou substitui sequências de dados repetidas para economizar espaço
- Exemplo: em um arquivo, substituir "aaaaaa" por "6a"
- Muito comum em compressão de imagens e textos

# Código de Huffman

- Exemplo de código de tamanho variável
  - Ideia: valores mais frequentes são associados a códigos menores
- No código ASCII: 1 byte por caracter (fixo)
  - 'A' = 65 (8 bits)
  - Cadeia 'ABC' ocupa 3 bytes

# Código de Huffman

- Se letras que ocorrem com frequência tem códigos menores, as cadeias tendem a ficar mais curtas
  - Requer informação sobre a frequência de ocorrência de cada símbolo a ser codificado
  - Muito usado para codificar texto
- Exemplo
  - Alfabeto: {A, B, C, D}
  - Frequência:  $A > C > B = D$
  - Possível codificação:
    - $A = 0, B = 1\ 1\ 0, C = 10, D = 111$
    - Cadeia: A B A C C D A
    - Código: 0 1 1 0 0 1 0 1 1 1 0

**Codificação não pode ser ambígua!**

Ex.:  $A = 0, B = 0\ 1, C = 1$

A C B A = 0 1 0 1 0

É possível decodificar?

# Sumário

- Modelos abstratos de dados
- Metadados
- Compressão de dados
- **Compactação**

# Compactação

- Consiste na busca por regiões do arquivo que não contém dados e posterior recuperação desses espaços perdidos
- Os espaços vazios são provocados, por exemplos, pela **eliminação de registros**