AEDsI – Aula 10
Arquivos

Universidade Federal de Minas Gerais

Primeiro Semestre de 2017

DCC
DEPARTAMENTO DE
CIÊNCIA DA COMPUTAÇÃO

UFMG

[1] Baseado nas aulas de MC102 da Prof. Islene Calciolari Garcia

UFMG

Roteiro

UFMG

Arquivos

Características

- Podem armazenar grande quantidade de informação.
- Dados são persistentes (gravados em disco).
- Acessso aos dados pode ser não seqüencial (acesso direto a registros em um banco de dados).
- Acesso à informação pode ser concorrente (mais de um programa ao mesmo tempo).

UFMG

Nomes e extensões

- Arquivos são identificados por um nome.
- O nome de um arquivo pode conter uma extensão que indica o conteúdo do arquivo.

Algumas extensões

arq.txt	arquivo texto simples
arq.c	código fonte em C
arq.pdf	*portable document format*
arq.html	arquivo para páginas WWW *(hypertext markup language)*
arq*	arquivo executável (UNIX)

UFMG

Tipos de arquivos

Arquivos podem ter o mais variado conteúdo, mas do ponto de vista dos programas existem apenas dois tipos de arquivo:

Arquivo texto: Armazena caracteres que podem ser mostrados diretamente na tela ou modificados por um editor de textos simples. Exemplos: código fonte C, documento texto simples, páginas HTML.

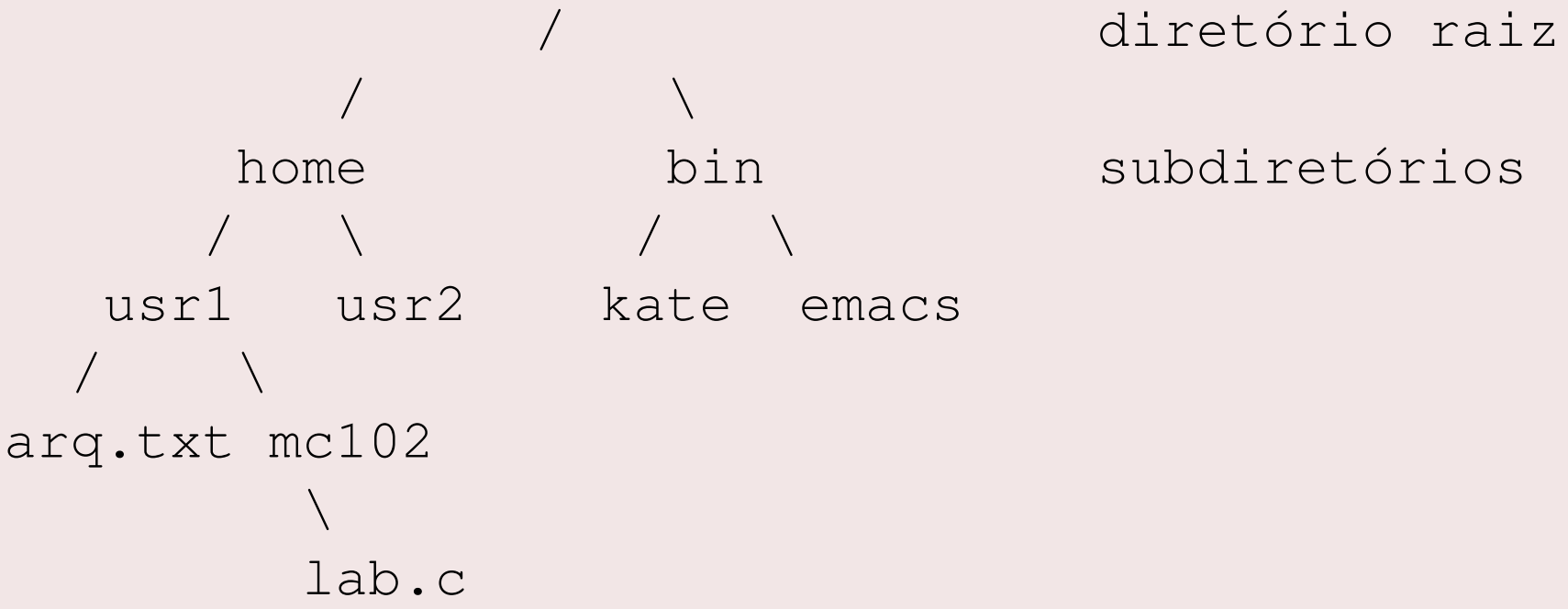
Arquivo binário: Seqüência de bits sujeita às convenções dos programas que o gerou, não legíveis diretamente. Exemplos: arquivos executáveis, arquivos compactados, documentos do Word.

UFMG

Diretório

- Também chamado de pasta.
- Contém arquivos e/ou outros diretórios.

Uma hierarquia de diretórios



UFMG

Caminhos absolutos ou relativos

O nome de um arquivo pode conter o seu diretório, ou seja, o caminho para encontrar este arquivo a partir da raiz. Os caminhos podem ser especificados de duas formas:

Caminho absoluto: descrição de um caminho desde o diretório raiz.

```
/bin/emacs
/home/usr1/arq.txt
```

Caminho relativo: descrição de um caminho desde o diretório corrente.

```
arq.txt
mc102/lab.c
```

Para ver qual é o diretório corrente, use o comando `pwd`. Para mudar de diretório, use o comando `cd`.

UFMG

Atributos de arquivos

Além do nome, arquivos possuem vários outros atributos:

- Nome do arquivo
- Proprietário do arquivo
- Datas de criação, alteração e acesso
- Tamanho em bytes
- Permissão de acesso

Para ver estes atributos, use os comandos `ls -l` e `stat`.

UFMG

Permissão de acesso

Existem três níveis de controle: proprietário, grupo e todos.

```
$ ls -l
-rw-r----- 1 jose alunos  545 Nov  8  2005 cp.c
drwxr-xr-x 2 jose alunos 4096 Jun  6 14:54 mc102/
```

- r: leitura
- w: escrita
- x: execução para arquivos, permissão de entrada para diretórios

UFMG

Abrindo um arquivo para leitura

- Antes de acessar um arquivo, devemos abri-lo com a função `fopen()`.
- Em caso de erro a função retorna `NULL`.
- A função `perror()` obtém e exibe uma mensagem explicativa.

Abrindo o arquivo `teste.txt`

```
if (fopen("teste.txt", "r") == NULL)
  perror("Erro ao abrir o arquivo.\n");
else
  printf("Arquivo aberto para leitura.\n");
```

Veja o exemplo em `fopen-r.c`.

UFMG

Lendo dados de um arquivo

- Não basta chamar a função `fopen()`, temos que pegar o seu valor de retorno (um apontador para *stream*).
- Para ler dados do arquivo, usamos a função `fscanf()`, semelhante à função `scanf()`.
- Para fechar o arquivo usamos a função `fclose()`.

Lendo dados do arquivo `teste.txt`

```
FILE *f = fopen ("teste.txt", "r");
while (fscanf(f, "%c", &c) != EOF)
  printf("%c", c);
fclose(f);
```

Veja o exemplo em `fscanf.c`.

UFMG

Escrevendo dados em um arquivo

- Para escrever em um arquivo, ele deve ser aberto de forma apropriada.
- Usamos a função `fprintf()`, semelhante a função `printf()`.

Copiando dois arquivos

```
FILE *fr = fopen ("teste.txt", "r");
FILE *fw = fopen ("saida.txt", "w");
while (fscanf(fr, "%c", &c) != EOF)
  fprintf(fw,"%c", c);
fclose(fr);
fclose(fw);
```

Veja o exemplo em `fprintf.c`.

UFMG

fopen

Um pouco mais sobre a função `fopen()`.

```
FILE* fopen(const char *caminho, char *modo);
```

Modos de abertura de arquivo

modo	operações	ponto no arquivo
r	leitura	início
r+	leitura e escrita	início
w	escrita	início
w+	leitura e escrita	início
a	escrita	final
a+	leitura	início
	escrita	final

UFMG

Lendo um vetor de um arquivo

```
FILE  *fr;
int i, n, *v;

fr = fopen ("v-in.txt", "r");
fscanf(fr, "%d", &n);  /* Dimensão do vetor */
v = (int *) malloc (n * sizeof(int));
for (i = 0; i < n; i++)
  fscanf(fr, "%d", &v[i]);
fclose(fr);
```

Veja o exemplo em `le_vetor.c`.

UFMG

Escrevendo um vetor em um arquivo

```
FILE  *fw = fopen ("v-out.txt", "w");
fprintf(fw, "%d\n", n); /* Dimensão do vetor */
for (i = 0; i < n; i++)
  fprintf(fw, "%d\n", v[i]);
fclose(fw);
```

Veja o exemplo em `le_vetor.c`.

UFMG

Lendo uma matriz de um arquivo

- Para usar alocação dinâmica, uma solução é criar um vetor linear de dimensão `nlin * ncol`.

```
int *v = (int *)
         malloc (nlin * ncol * sizeof(int));
for (i = 0; i < nlin * ncol; i++)
  fscanf(fr, "%d", &v[i]);
```

Veja o exemplo em `le_matriz.c`.

UFMG

Escrevendo uma matriz em um arquivo

- Para gravar uma matriz, usamos a idéia inversa, ou seja: `mat[i][j] = v[i*ncol + j]`.

```
for (i = 0; i < nlin; i++)
  for (j = 0; j < ncol; j++)
   fprintf(fw, "mat[%d][%d] = %d\n",
           i, j, v[i*ncol + j]);
```

Veja o exemplo em `le_matriz.c`.

UFMG

Lendo uma matriz de um arquivo

- Uma outra forma de fazer alocação dinâmica consiste em criar `nlin` vetores de `ncol` inteiros.

```
int **v = (int **) malloc(nlin * sizeof(int*));
for (i = 0; i < nlin; i++)
  v[i] = (int *) malloc(ncol * sizeof(int));
for (i = 0; i < nlin; i++)
  for (j = 0; j < ncol; j++)
    fscanf(fr, "%d", &v[i][j]);
```

Veja o exemplo em `le_matriz2.c`.

UFMG

Argumentos para o main

- Como já vimos, o bloco `main` é uma função.
- Esta função recebe argumentos da linha de comando.

```
int main (int argc, char* argv[]) {
  FILE *fr, *fw
  if (argc < 3) {
    printf("Uso: %s <origem> <destino>\n",
            argv[0]);
    return 1;
  }
  fr = fopen (argv[1], "r");
  fw = fopen (argv[2], "w");
```

Veja o exemplo em `cp.c`.

UFMG

Motivação

- Variáveis `int` ou `float` têm tamanho fixo na memória. Por exemplo, um `int` ocupa 4 bytes.
- Representação em texto precisa de um número variável de dígitos (`10, 5.673, 100.340`), logo de um tamanho variável.
- Armazenar dados em arquivos de forma análoga a utilizada em memória permite:
  - Reduzir o tamanho do arquivo.
  - Realizar busca não seqüencial.

UFMG

fread e fwrite

- As funções `fread` e `fwrite` permitem a leitura e escrita de blocos de dados.
- A idéia é semelhante a alocação de memória dinâmica.
- Devemos determinar o número de elementos a serem lidos ou gravados e o tamanho de cada um.

```
size_t fread(void *ptr, size_t size,
             size_t nmemb, FILE *stream);

size_t fwrite(const void *ptr, size_t size,
              size_t nmemb, FILE *stream);
```

Veja os exemplos em `vetor.c` e `copiar.c`.

UFMG

Acesso não seqüencial

- Fazemos o acesso não seqüencial usando a função `fseek`.
- Esta função altera a posição de leitura/escrita no arquivo.
- O deslocamento pode ser relativo ao:
  - início do arquivo (SEEK_SET)
  - ponto atual (SEEK_CUR)
  - final do arquivo (SEEK_END)

```
int fseek(FILE *stream, long offset, int whence);
```

Veja o exemplo em `fseek.c`.

UFMG

Registros

- Um arquivo pode armazenar registros (como um banco de dados).
- Isso pode ser feito de forma bem fácil se lembrarmos que um registro, como qualquer variável em C, tem um tamanho fixo.
- O acesso a cada registro pode ser direto, usando a função `fseek`.
- A leitura ou escrita do registro pode ser feita usando as funções `fread` e `fwrite`.

Veja o exemplo em `registro.c`.

UFMG

Exercício

Merge

- Escreva um programa que leia dois arquivos de inteiros ordenados e escreva um arquivo cuja saída é um único arquivo ordenado.
  - Vale a pena colocar o conteúdo dos arquivos de entrada em dois vetores?
  - Escreva duas versões deste programa, uma para arquivos texto e outra para arquivos binários.

UFMG

Exercício

Ordenação

- Escreva um programa que lê uma série de linhas de um arquivo texto e escreve um arquivo contendo estas linhas ordenadas.
  - Como você faria para armazenar estas linhas em memória?
  - Qual algoritmo de ordenação você acha mais adequado?