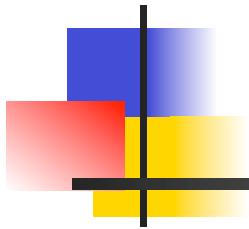


Organização de Arquivos



Cristina D. A. Ciferri

Thiago A. S. Pardo

Leandro C. Cintra

M.C.F. de Oliveira

Moacir Ponti Jr.



Por que Organizar Arquivos?

- Considere o seguinte *stream* (fluxo) de bytes
 - AmesJohn123 MapleStillwaterOK74075MasonAlan90 EastgateAdaOK74820
- Perguntas:
 - quais dados são armazenados?
 - onde começa e onde termina cada dado?
 - como recuperar um dado específico?

Necessidade de organização adequada de dados em arquivos!



Formas de Organização

- Organização em campos
 - menor unidade lógica de informação em um arquivo
- Organização em registros
 - conjunto de campos agrupados, os quais estão logicamente associados a uma mesma entidade

Os conceitos de campo e de registro correspondem a **ferramentas conceituais**, que não necessariamente existem no sentido físico



Métodos para Organização em Campos

- Forçar todos os campos para um **tamanho** (comprimento) **fixo**
- Começar cada campo com um **indicador de tamanho** (comprimento)
- Colocar **delimitadores** entre campos
- Usar expressões (*tags*)
“keyword=value” para identificar cada campo e seu conteúdo



Campos com Tamanho Fixo

- Cada campo ocupa no arquivo um tamanho fixo, pré-determinado
- Exemplo
 - nome: *string* de 12 caracteres (12 bytes)
 - rua: *string* de 10 caracteres (10 bytes)
 - número: inteiro (4 bytes)
 - cidade: *string* de 20 caracteres (20 bytes)



Campos com Tamanho Fixo

- Vantagens

- facilidade na pesquisa, desde que o tamanho do campo facilita a sua recuperação
- indicado para situações nas quais o comprimento dos campos é fixo ou apresenta pouca variação



Campos com Tamanho Fixo

- Desvantagens

- espaço alocado (e não usado) aumenta desnecessariamente o tamanho do arquivo (desperdício de espaço de armazenamento)
- solução inapropriada quando se tem uma grande quantidade de dados com tamanho variável
- dados podem precisar ser truncados



Campos com Indicador de Tamanho

- Tamanho de cada campo em bytes
 - armazenado imediatamente **antes** do dado
- Exemplo
 - tamanho: número inteiro (4 bytes)



Campos com Indicador de Tamanho

- Vantagem
 - economia de espaço de armazenamento, mesmo com a necessidade de se gastar alguns bytes adicionais para guardar o tamanho dos campos
 - dados não precisam ser truncados
- Desvantagem
 - dificuldade na pesquisa



Campos Separados por Delimitadores

- Caractere(s) especial(ais) que não fazem parte do domínio do dado
 - escolhido(s) para ser(em) inserido(s) ao final de cada campo (ou seja, armazenado(s) imediatamente **depois** do dado)
 - exemplos: /, tab, #
- Exemplo
 - delimitador: | (caractere de 1 *byte*)



Campos Separados por Delimitadores

- Vantagem
 - economia de espaço de armazenamento
- Desvantagens
 - dificuldade na pesquisa
 - necessidade de escolha de um delimitador que não pertence ao domínio dos dados



Uso de *Tags*

- Expressão “keyword=value”
 - colocada imediatamente **antes** do campo
 - possui semântica que explica o significado do campo
- Geralmente usada em conjunto com outro método para campos (ex. delimitador, indicador de tamanho)



Uso de *Tags* + Campos Separados por Delimitadores

- Exemplo
 - delimitador: | (caractere de 1 *byte*)
 - tags:
 - nome =
 - endereco =
 - numero =
 - cidade =



Uso de *Tags*

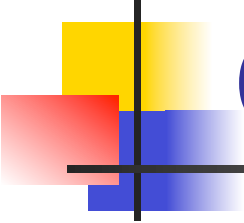
Expressão “keyword=value”

- Vantagens

- campo fornece informação semântica sobre si
- fica mais fácil identificar o conteúdo do arquivo
- fica mais fácil identificar campos “vazios”
 - *keyword* não aparece

- Desvantagem

- as *keywords* podem ocupar uma porção significativa do arquivo (desperdício de espaço de armazenamento)



Métodos para Organização em Registros

- Forçar todos os registros para um **tamanho fixo**
- Forçar todos os registros para conterem um **número fixo de campos**
- Começar cada registro com um **indicador de tamanho**
- Usar **índice** para manter informação de endereçamento
- Colocar **delimitadores** entre registros



Registros de Tamanho Fixo

- Todos os registros têm o mesmo número de bytes
- Pode-se ter:
 - registros de tamanho fixo com campos de tamanho fixo
 - registros de tamanho fixo com campos de tamanho variável

Um dos métodos mais comuns
de organização de arquivos



Registros de Tamanho Fixo com Campos de Tamanho Fixo

- Exemplo
 - Registros de tamanho fixo
 - tamanho de 46 bytes
 - Campos de tamanho fixo
 - nome: *string* de 12 caracteres (12 bytes)
 - rua: *string* de 10 caracteres (10 bytes)
 - número: inteiro (4 bytes)
 - cidade: *string* de 20 caracteres (20 bytes)



Registros de Tamanho Fixo com Campos de Tamanho Variável

- Exemplo
 - Registros de tamanho fixo
 - tamanho de 46 bytes
 - Campos de tamanho variável
 - delimitador: | (caractere de 1 byte)



Registros de Tamanho Variável com Número Fixo de Campos

- Cada registro contém um número fixo de campos
 - o tamanho do registro, em bytes, é variável
 - necessidade de ser usado em conjunto com outro método para campos (ex. delimitador, indicador de tamanho)
- Exemplo
 - delimitador: | (caractere de 1 *byte*)



Indicador de Tamanho para Registros

- Tamanho de cada registro em bytes
 - armazenado imediatamente antes do registro
- Os campos devem ser separados por outro método para campos (ex. delimitador, indicador de tamanho)

Método muito utilizado para manipular registros de tamanho variável



Indicador de Tamanho para Registros

- Exemplo
 - indicador de tamanho de registro
 - número inteiro de 4 bytes



Usar um Índice para Endereçamento

- **Arquivo secundário**
 - mantém informações sobre o endereço do primeiro *byte* de cada registro
 - indica o deslocamento (*byte offset*) de cada registro relativo ao início do arquivo
- Necessidade de ser usado em conjunto com outro método, como delimitadores



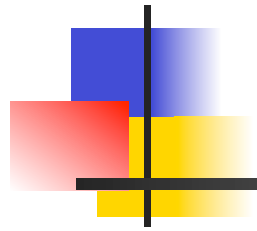
Usar um Índice para Endereçamento

- Vantagem
 - flexibilidade
- Desvantagens
 - necessidade de acessar dois arquivos diferentes
 - manutenção da consistência do índice



Registros Separados por Delimitadores

- Separar os registros com delimitadores
 - o delimitador de registro é colocado ao final do registro
 - necessidade de ser usado em conjunto com um método para campos (ex. delimitador, indicador de tamanho)
- Exemplo
 - delimitador de registro: #
 - delimitador de campo: |



Discussões



Observação

- Nenhum dos métodos descritos é apropriado para todas as situações
- Escolha do método depende
 - da natureza dos dados
 - para o que eles serão usados



Como Escolher a Organização de um Arquivo

- Análises

- arquivo pode ser dividido em campos?
- os campos são agrupados em registros?
- registros têm tamanho fixo ou variável?
- como separar os registros?
- como identificar o espaço utilizado e o "lixo"?

- Escolha

- depende, entre outras coisas, do que dados vão ser armazenados no arquivo



Tamanho dos Registros: Exemplo com Decisão Fácil

- Arquivo de controle de vendas
 - número do comprador: 8 bytes
 - data no formato DDMMAA: 6 bytes
 - número do item: 4 bytes
 - quantidade vendida: 4 bytes
 - valor da venda: 8 bytes
- Campos de tamanho fixo: **30 bytes**



Tamanho dos Registros: Exemplo com Decisão Fácil

- Registros de **30 bytes**
 - páginas de disco de 4KB (4.096 bytes)
 - número de registros por página: **136,53**
- Registros de **32 bytes**
 - páginas de disco de 4KB (4.096 bytes)
 - número de registros por página: **128**

escolhido

tamanho do registro deve se encaixar no
tamanho da página de disco



Tamanho dos Registros: Exemplo com Decisão Difícil

- Campos com tamanho muito variável
 - nome
 - endereço
- Abordagens simplistas para o tamanho do registro
 - 1 soma do **tamanho máximo** de cada campo
 - 2 soma do **tamanho mínimo** de cada campo



Tamanho dos Registros: Exemplo com Decisão Difícil

- Todos os campos de tamanho fixo
 - vantagem
 - simplicidade na pesquisa
 - desvantagens
 - problemas dos tamanhos máximo e mínimo



Tamanho dos Registros: Exemplo com Decisão Difícil

- Todos os campos de tamanho variável
 - vantagem
 - pode-se aplicar o efeito do **tamanho médio**
 - *nomes mais longos em geral não aparecem no mesmo registros que os endereços mais longos*
 - desvantagem
 - mais dificuldade na pesquisa por campos que não sofrem muita variação



Tamanho dos Registros: Exemplo com Decisão Difícil

- Decisão interessante
 - campos cujos dados possuem pouca variabilidade: **campos de tamanho fixo**
 - campos cujos dados possuem grande variabilidade: **campos de tamanho variável**

usar uma organização em campos adequada às características dos dados armazenados no arquivo



Organização Híbrida de Campos e Registros

- Campos
 - tamanho fixo
 - tamanho variável com indicador de tamanho
- Registros de tamanho variável
 - separados por delimitadores
 - delimitador: #



Organização Híbrida de Campos e Registros

- Exemplo
 - campos de tamanho fixo
 - número: inteiro (4 bytes)
 - telefone: *string* de 13 caracteres (13 bytes)
 - campos de tamanho variável
 - nome
 - rua
 - cidade

Organização Híbrida de Campos e Registros

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
123	1	6	-	3	3	7	3	-	3	3	7	3	b	5	M	A	R	I	A	5									
30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59
R	U	A	b	1	10	S	A	O	b	C	A	R	L	O	S	#	255	1	9	-	1	1	3						
60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89
3	-	1	1	3	3	b	4	J	O	A	O	5	R	U	A	b	A	9	R	I									
90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119
O	b	C	L	A	R	O	#	56	1	6	-	3	3	6	1	-	3	3	6	1	b	5	P						
120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149
E	D	R	O	6	R	U	A	b	X	V	10	S	A	O	b	C	A	R	L	O	S	#							



Observação

Ver programas em C no livro texto que ilustram a criação de arquivos com as diferentes organizações estudadas!