

Arquivos - Organização

Parte 2 - Registros

Prof.: Leonardo Tórtoro Pereira
leonardop@usp.br

*Material baseado em aulas dos professores: Elaine Parros Machado de Souza, Gustavo Batista, Robson Cordeiro, Moacir Ponti Jr., Maria Cristina Oliveira e Cristina Ciferri.

O Que Já Vimos?

Organização em Campos

- Estruturas de Organização de Campos:
 - ◆ Comprimento fixo
 - ◆ Indicador de comprimento
 - ◆ Delimitadores
 - ◆ Uso de tags

Organização em Campos

(a)	Maria	Rua 1	123	São Carlos
	João	Rua A	255	Rio Claro
	Pedro	Rua 10	56	Rib. Preto

(b) 05Maria05Rua 10312310São Carlos
04João05Rua A0325509Rio Claro
05Pedro06Rua 10025610Rib. Preto

(c) Maria|Rua 1|123|São Carlos|
João|Rua A|255|Rio Claro|
Pedro|Rua 10|56|Rib. Preto|

(d) Nome=Maria|Endereço=Rua 1|Número=123|Cidade=São Carlos|
Nome=João|Endereço=Rua A|Número=255|Cidade=Rio Claro|
Nome=Pedro|Endereço=Rua 10|Número=56|Cidade=Rib. Preto|

Como escalar isso para vários campos?

Organização em Registros!

Organização em Registros

- Registro:
 - ◆ Um conjunto de campos agrupados
- Arquivo organizado em registros
 - ◆ Nível de organização mais alto

Organização em Registros

- Assim como os campos, um registro é uma ferramenta conceitual
 - ◆ Está associado ao arquivo lógico e não à organização física
 - ◆ Outro nível de organização imposto aos dados para preservar significado semântico

Organização em Registros

→ Estruturas de Organização de Registros:

◆ Tamanho fixo

- Campos de tamanho fixo
- Campos de tamanho variável

◆ Tamanho variável

- Número pré-determinado de campos
- Uso de delimitadores
- Indicador de tamanho
- Uso de índice

Registros de Tamanho Fixo

Registros de Tamanho Fixo

Registro de tamanho fixo e campos de tamanho fixo:

Maria	Rua 1	123	São Carlos
João	Rua A	255	Rio Claro
Pedro	Rua 10	56	Rib. Preto

Registro de tamanho fixo e campos de tamanho variável:

Maria		Rua 1		123		São Carlos		← Espaço vazio →
João		Rua A		255		Rio Claro		← Espaço vazio →
Pedro		Rua 10		56		Rib. Preto		← Espaço vazio →

Registros de Tamanho Fixo

- Assume que todos os registros ocupam o mesmo número de bytes:
 - ◆ Campos podem ou não ter tamanho fixo
- É um dos métodos mais comuns de organização de arquivos
 - ◆ Simples
- Desvantagem?
 - ◆ Desperdício de memória secundária – fragmentação

Registros e Campos Fixos

Registros e Campos Fixos

→ Exemplo

- ◆ Registros de tamanho fixo – 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 e Campos Fixos

→ Exemplo em C:

```
struct Pessoa {  
    char nome[12];  
    char rua [10];  
    int  numero;  
    char cidade[20];  
};
```

Registros e Campos Fixos

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
M	A	R	I	A	b	b	b	b	b	b	b	R	U	A	b	1	b	b	b	b	b		123			S	A	O	b

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
C	A	R	L	O	S	b	b	b	b	b	b	b	b	b	b	J	O	A	O	b	b	b	b	b	b	b	b	R	U

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
A	b	A	b	b	b	b	b		255			R	I	O	b	C	L	A	R	O	b	b	b	b	b	b	b	b	b

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
b	b	P	E	D	R	O	b	b	b	b	b	b	b	R	U	A	b	X	V	b	b	b	b		56		S	A	

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
O	b	C	A	R	L	O	S	b	b	b	b	b	b	b	b	b	b												

Registros Fixos e Campos Variáveis

Registros Fixos e Campos Variáveis

→ Exemplo

- ◆ Registros de tamanho fixo – 46 bytes
- ◆ Campos de tamanho variável
 - Delimitador “|” de 1 byte

Registros Fixos e Campos Variáveis

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	
M	A	R	I	A		R	U	A	b	1		123					S	A	O	b	C	A	R	L	O	S		b	b	
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	
b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	J	O	A	O		R	U	A	b	A		255			
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	
		R	I	O	b	C	L	A	R	O		b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	
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	
b	b		P	E	D	R	O		R	U	A	b	X	V		56					S	A	O	b	C	A	R	L	O	S
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	
	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b														

Registros Variáveis e # Fixo de Campos

Registros Variáveis e # Fixo de Campos

- Ao invés de especificar que cada registro contém um número fixo de bytes, podemos especificar um número fixo de campos
 - ◆ O tamanho do registro é variável em bytes
 - ◆ ex.: campos separados por delimitadores:

Maria|Rua 1|123|São Carlos|João|Rua A|255|Rio Claro|Pedro|Rua 10|56|Rib. Preto|

4 campos

4 campos

4 campos

Registros Variáveis e # Fixo de Campos

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
M	A	R	I	A		R	U	A	b	1		123					S	A	O	b	C	A	R	L	O	S		J	O

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
A	O		R	U	A	b	A		255				R	I	O	b	C	L	A	R	O		P		E	D	R	O	

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
R	U	A	<i>b</i>	X	V		56					S	A	O	<i>b</i>	C	A	R	L	O	S								

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

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

Registros de Tamanho e # Campos Variáveis

Registros de Tamanho e # Campos Variáveis

- Separar registros com delimitadores
 - ◆ Análogos aos de fim de campo
 - Porém outro caractere deve ser utilizado
 - Delimitador de campos pode ser mantido
- Vantagens:
 - ◆ Permite número variável de campos
 - ◆ início do arquivo não é mais a única referência

Maria|Rua 1|123|São Carlos|#João|Rua A|255|Rio Claro|#Pedro|Rua 10|56|Rib. Preto|#



Registros de Tamanho e # Campos Variáveis

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
M	A	R	I	A		R	U	A	b	1		123					S	A	O	b	C	A	R	L	O	S		#	J

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
O	A	O		R	U	A	b	A		255					R	I	O	b	C	L	A	R	O		#	P	E	D	R

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
O		R	U	A	b	X	V		56					S	A	O	b	C	A	R	L	O	S		#				

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

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

Registros de Tamanho Variável e Indicador de Tamanho

Registros de Tamanho Variável e Indicador de Tamanho

- Indicador de Tamanho:
 - ◆ Indicador que precede o registro
 - Fornece o seu tamanho total
 - Campos separados por delimitadores
- No início de cada registro:
 - ◆ Sabe-se onde termina aquele registro
 - ◆ Possível acessar diretamente o registro seguinte
 - ◆ Estratégia bastante usada para arquivos com registros de tamanho variável

Registros de Tamanho Variável e Indicador de Tamanho

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
28				M	A	R	I	A		R	U	A	b	1		123					S	A	O	b	C	A	R	L	O

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
S		26				J	O	A	O		R	U	A	b	A		255					R	I	O	b	C	L	A	R

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
O		255				P	E	D	R	O		R	U	A	b	X	V		56					S	A	O	b	C	A

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
R	L	O	S																										

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

Registros de Tamanho Variável e Índice de Endereçamento

Registros de Tamanho Variável e Índice de Endereçamento

- Índice de Endereçamento:
 - ◆ Armazena o byte offset
 - ◆ Deslocamento relativo ao início do arquivo
 - ◆ Também usado para calcular o tamanho dos registros.
 - ◆ Campos separados por delimitadores
- Desvantagens?
 - ◆ Arquivo de índice
 - ◆ Manutenção de consistência

Registros de Tamanho Variável e Índice de Endereçamento

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
M	A	R	I	A		R	U	A	<i>b</i>	1		123					S	A	O	<i>b</i>	C	A	R	L	O	S		J	O

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	
A	O		R	U	A	<i>b</i>	A		255				R	I	O	<i>b</i>	C	L	A	R	O				P	E	D	R	O	

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
R	U	A	<i>b</i>	X	V		56					S	A	O	<i>b</i>	C	A	R	L	O	S								

Arquivo do índice para endereçamento

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
00					28					54																			

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, de que dados vão ser armazenados no arquivo

Tamanho dos Registros: Exemplo com Decisão Fácil

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

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

Tamanho dos Registros: Exemplo com Decisão Difícil

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

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
1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0

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	#							

Referências

- M. J. Folk, B. Zoellick and G. Riccardi. File Structures: An object-oriented approach with C++, Addison Wesley, 1998.
- Códigos da aula estão no repositório <https://github.com/LeonardoTPereira/Alg2/tree/main/Aula%2003>