

# 1

## INTRODUCCION: ARCHIVOS Y ESTRUCTURAS DE ARCHIVOS

- 1.1 Almacenamiento primario y secundario 2
- 1.2 Nada es gratis 3
- 1.3 Archivos 5
- 1.4 Estructuras de archivos *versus* estructuras de datos 6
- 1.5 Un conjunto de herramientas conceptuales 7
- Resumen 8
- Términos clave 9

# 2

## OPERACIONES FUNDAMENTALES PARA EL PROCESAMIENTO DE ARCHIVOS

- 2.1 Archivos físicos y archivos lógicos 12
- 2.2 Apertura y creación de archivos 13
- 2.3 Cierre de archivos 17
- 2.4 Lectura y escritura 18
- 2.5 Detección del fin del archivo 22
- 2.6 Localización 22
  - 2.6.1 Localización en C 23
  - 2.6.2 Localización en Pascal 25

# INDICE GENERAL

- 2.7 Caracteres inesperados en archivos 27**
- Resumen 28
- Términos clave 29
- Ejercicios 31
- Lecturas adicionales 33
- El programa *agrega* en Pascal 34


**3**

## **DISPOSITIVOS DE ALMACENAMIENTO SECUNDARIO Y SOFTWARE DE SISTEMAS: CONSIDERACIONES DE DESEMPEÑO**

- 3.1 Discos 39**
  - 3.1.1 La organización de los discos 39
  - 3.1.2 Estimación de las capacidades y necesidades de espacio 42
  - 3.1.3 Organización por sectores 43
  - 3.1.4 Organización por bloques 48
  - 3.1.5 Sobrecarga por datos usados para control 50
  - 3.1.6 El costo de un acceso a disco 52
- 3.2 Cinta magnética 56**
  - 3.2.1 Organización de datos en cintas 56
  - 3.2.2 Estimación de requerimientos de longitud de cinta 58
  - 3.2.3 Estimación de los tiempos de transmisión de datos 60
  - 3.2.4 Aplicaciones de las cintas 62
- 3.3 Otros tipos de almacenamiento 62**
- 3.4 El almacenamiento como una jerarquía 65**
- 3.5 El viaje de un byte 65**
  - 3.5.1 El administrador de archivos 67
  - 3.5.2 El buffer de E/S 69
  - 3.5.3 El byte sale de la memoria RAM: el procesador de E/S 71
  - 3.5.4 El byte llega al disco: el controlador del disco 73
- 3.6 Manejo de buffers 74**
- Resumen 77
- Términos clave 79
- Ejercicios 83
- Lecturas adicionales 87

## 4

**CONCEPTOS FUNDAMENTALES DE ESTRUCTURAS DE ARCHIVOS**

- 4.1 Un archivo como secuencia de bytes 90
- 4.2 Estructuras de campos 92
  - 4.2.1 Método 1: Fijar la longitud de los campos 93
  - 4.2.2 Método 2: Comenzar cada campo con un indicador de longitud 94
  - 4.2.3 Método 3: Separar los campos con delimitadores 94
- 4.3 Lectura de una secuencia de campos 95
- 4.4 Estructuras de registros 96
  - 4.4.1 Método 1: Hacer los registros de una longitud predecible 97
  - 4.4.2 Método 2: Comenzar cada registro con un indicador de longitud 99
  - 4.4.3 Método 3: Usar un segundo archivo para mantener información sobre las direcciones 99
  - 4.4.4 Método 4: Colocar un delimitador al final de cada registro 100
- 4.5 Una estructura de registros que usa un indicador de longitud 100
- 4.6 Mezcla de números y caracteres: uso de un vaciado hexadecimal 102
- 4.7 Lectura de registros de longitud variable de un archivo 106
- 4.8 Extracción de registros por llave: formas canónicas para llaves 106
- 4.9 Una búsqueda secuencial 109
- 4.10 Evaluación del desempeño de la búsqueda secuencial 109
- 4.11 Mejora del desempeño de la búsqueda secuencial: manejo de registros en bloques 111
- 4.12 Acceso directo 113
- 4.13 Elección de una estructura y una longitud de registro 115
- 4.14 Registros de encabezado 119
- 4.15 Acceso y organización de archivos 119
- Resumen 121
- Términos clave 123
- Ejercicios 125
- Lecturas adicionales 130
- Programas en C 131
- Programas en Pascal 146

**5****MANTENIMIENTO DE ARCHIVOS Y  
ELIMINACION DE REGISTROS**

- 5.1** Mantenimiento de archivos 164
- 5.2** Compactación del almacenamiento 165
- 5.3** Panorama de la eliminación de registros de longitud fija 167
- 5.4** Realización de la eliminación de registros de longitud fija 171
- 5.5** Eliminación de registros de longitud variable 173
- 5.6** Fragmentación del almacenamiento 177
- 5.7** Estrategias de colocación 181
- Resumen 183
- Términos clave 185
- Ejercicios 187
- Lecturas adicionales 189

**6****LOCALIZACION RAPIDA DE DATOS EN UN  
ARCHIVO: INTRODUCCION A LA CLASIFICACION  
Y A LA BUSQUEDA BINARIA**

- 6.1** Localización de datos en los archivos ya desarrollados 193
- 6.2** Búsqueda por conjetura: búsqueda binaria 193
- 6.3** Clasificación de un archivo de disco en memoria RAM 196
- 6.4** Algoritmo para la clasificación en memoria RAM 201
- 6.5** La función de clasificación: `clasif_shell()` 202
- 6.6** Limitaciones de la búsqueda binaria y de la clasificación en memoria RAM 203
- 6.7** Clasificación por llave 206
  - 6.7.1** Descripción del método 206
  - 6.7.2** Limitaciones del método de clasificación por llave 208
  - 6.7.3** Otra solución: ¿Por qué molestarse en escribir de nuevo el archivo? 209
- 6.8** Registros fijos 211
- Resumen 212
- Términos clave 214
- Ejercicios 215
- Lecturas adicionales 217

Programas en C 218  
 Programas en Pascal 222


**7**
**INDIZACION**

- 7.1 ¿Qué es un índice? 228
- 7.2 Índice simple de un archivo con entradas secuenciales 229
- 7.3 Operaciones básicas en un archivo indizado con entradas secuenciales 233
- 7.4 Índices demasiado grandes para almacenarse en memoria 237
- 7.5 Indización para proporcionar acceso mediante varias llaves 239
- 7.6 Extracción de información mediante combinaciones de llaves secundarias 244 ↗
- 7.7 Mejora de la estructura secundaria de índices: listas invertidas 246
  - 7.7.1 Primer intento de solución 247
  - 7.7.2 Una mejor solución: ligar la lista de referencias 249
- 7.8 Índices selectivos 253
- 7.9 Enlace (*Binding*) 253
- Resumen 255
- Términos clave 257
- Ejercicios 258
- Lecturas adicionales 262


**8**
**PROCESAMIENTO SECUENCIAL  
 COORDINADO Y CLASIFICACION DE  
 ARCHIVOS GRANDES**

- 8.1 Un modelo para la realización de procesos secuenciales coordinados 265
  - 8.1.1 Correspondencia de nombres en dos listas 265
  - 8.1.2 Intercalación de dos listas 270
  - 8.1.3 Resumen del modelo de procesamiento secuencial coordinado 272
- 8.2 Aplicación del modelo a un programa de libro mayor 275
  - 8.2.1 El problema 275
  - 8.2.2 Aplicación del modelo al programa de libro mayor 278

- 8.3 Extensión del modelo para incluir la intercalación múltiple 284**
- 8.4 La intercalación como forma de clasificación de archivos grandes en disco 287**
  - 8.4.1 Patrones de intercalación de varios pasos 291
  - 8.4.2 Incremento en las longitudes de las porciones mediante el uso de selección por reemplazo 294
  - 8.4.3 Longitud promedio de las porciones para la selección por reemplazo 297
  - 8.4.4 Costo del uso de la selección por reemplazo 299
  - 8.4.5 Configuraciones de disco 302
  - 8.4.6 Resumen de la clasificación en disco 302
- 8.5 Clasificación de archivos en cinta magnética 303**
  - 8.5.1 Intercalación balanceada 304
  - 8.5.2 Intercalación en varias fases 306
- 8.6 Paquetes de clasificación e intercalación 309**
  - Resumen 309
  - Términos clave 312
  - Ejercicios 315
  - Lecturas adicionales 318

## 9

### **ARBOLES B Y OTRAS ORGANIZACIONES DE ARCHIVOS ESTRUCTURADAS EN FORMA DE ARBOL**

- 9.1 Introducción. La invención de los árboles B 322**
- 9.2 Planteamiento del problema 325**
- 9.3 Árboles binarios de búsqueda como solución 325**
- 9.4 Árboles AVL 329**
- 9.5 Árboles binarios paginados 332**
- 9.6 El problema con la construcción descendente de los árboles paginados 335**
- 9.7 Árboles B: construcción ascendente 337**
- 9.8 División y promoción 337**
- 9.9 Algoritmos para la búsqueda e inserción en árboles B 342**
- 9.10 Nomenclatura de árboles B 353**
- 9.11 Definición formal de las propiedades de los árboles B 354**
- 9.12 Profundidad de la búsqueda en el peor caso 355**
- 9.13 Eliminación, redistribución y concatenación 357**

- 9.13.1 Redistribución 361
- 9.14 Redistribución durante la inserción: una forma de mejorar la utilización del almacenamiento 362**
- 9.15 Árboles B\* 363**
- 9.16 Manejo de páginas en buffers: árboles B virtuales 365**
  - 9.16.1 Reemplazo LRU 366
  - 9.16.2 Reemplazo según la altura de la página
  - 9.16.3 Importancia de los árboles B virtuales 369
- 9.17 Colocación de la información asociada con la llave 370**
- 9.18 Registros y llaves de longitud variable 371**
- Resumen 373**
- Términos clave 375**
- Ejercicios 377**
- Lecturas adicionales 380**
- Programa en C para insertar llaves en un árbol B 382**
- Programa en Pascal para insertar llaves en un árbol B 390**

## LA FAMILIA DE LOS ARBOLES B<sup>+</sup> Y EL ACCESO A LOS ARCHIVOS SECUENCIALES INDIZADOS

### 10

- 10.1 Acceso secuencial indizado 398**
- 10.2 Mantenimiento de un conjunto de secuencias 399**
  - 10.2.1 Uso de bloques 400
  - 10.2.2 Elección del tamaño del bloque 403
- 10.3 Adición de un índice simple al conjunto de secuencias 404**
- 10.4 El contenido del índice: separadores en lugar de llaves 405**
- 10.5 Árboles B<sup>+</sup> de prefijos simples 409**
- 10.6 Mantenimiento de árboles B<sup>+</sup> de prefijos simples 410**
  - 10.6.1 Cambios localizados en bloques individuales del conjunto de secuencias 410
  - 10.6.2 Cambios que involucran varios bloques en el conjunto de secuencias 412
- 10.7 Tamaño de bloque del conjunto índice 415**
- 10.8 Estructura interna de los bloques del conjunto índice: árbol B de orden variable 416**
- 10.9 Carga de un árbol B<sup>+</sup> de prefijos simples 420**
- 10.10 Árboles B<sup>+</sup> 424**

### **10.11 Los árboles B, B<sup>+</sup> y B<sup>+</sup> de prefijos simples en perspectiva 426**

**Resumen 430**

**Términos clave 432**

**Ejercicios 434**

**Lecturas adicionales 439**

## **11**

### **DISPERSION (HASHING)**

#### **11.1 Introducción 442**

11.1.1 ¿Qué es la dispersión? 444

11.1.2 Colisiones 445

#### **11.2 Un algoritmo simple de dispersión 447**

#### **11.3 Funciones de dispersión y distribuciones de registros 451**

11.3.1 Distribución de registros entre direcciones 451

11.3.2 Algunos otros métodos de dispersión 453

11.3.3 Predicción de la distribución de los registros 454

11.3.4 Predicción de las colisiones en un archivo lleno 459

#### **11.4 ¿Cuánta memoria adicional debe usarse? 460**

11.4.1 Densidad de empaquetamiento 460

11.4.2 Predicción de colisiones para diferentes densidades de empaquetamiento 461

#### **11.5 Resolución de colisiones mediante saturación progresiva 465**

11.5.1 Funcionamiento de la saturación progresiva 465

11.5.2 Longitud de búsqueda 467

#### **11.6 Almacenamiento de más de un registro por dirección: compartimientos 470**

11.6.1 Efectos de los compartimientos en el desempeño 471

11.6.2 Aspectos de la realización 476

#### **11.7 La operación de eliminación 479**

11.7.1 Marcas de inutilización para eliminaciones 480

11.7.2 Implicaciones de las marcas de inutilización para las inserciones 481

11.7.3 Efectos de las eliminaciones y adiciones en el desempeño 482

#### **11.8 Otras técnicas de resolución de colisiones 483**

11.8.1 Dispersión doble 483

11.8.2 Saturación progresiva encadenada 484

11.8.3 Encadenamiento con un área de saturación separada 486

11.8.4 Tablas de dispersión: reconsideración de la indización 488

11.8.5 Dispersión extensible 488

**11.9 Patrones de acceso a los registros 490**

Resumen 491

Términos clave 495

Ejercicios 498

Lecturas adicionales 504

**Apéndice A Tabla de códigos de caracteres ASCII 508**

**Apéndice B Funciones de cadenas en Pascal:  
*herramientas.prc* 509**

**Apéndice C Introducción a C 514**

**Apéndice D Comparación de unidades de disco 570**

**Bibliografía 573**

**Vocabulario técnico bilingüe\* 587**

**Indice de materias 579**