

Fundamentos de Organización de Datos

Clase 5

Agenda

Indices

- Definición
- Operaciones básicas

Ejemplo

Indices secundarios

- Características

Búsqueda de datos - Indices

Búsqueda de información:

- debemos minimizar el número de accesos

Secuencial (poco eficiente)

Binaria (muy costosa)

Estructuras auxiliares

Búsqueda de datos - Indices

Ejemplo

Las últimas págs. de un libro suelen contener un índice (tabla que contiene una lista de temas y los n.º de pag. mejor alternativa donde pueden buscar un tema a lo largo del libro en forma

Búsqueda de datos - Índices

Otro ejemplo: encontrar libros en una biblioteca (por autor, título o tema)

- **Alternativa 1:** disponer 3 copias de cada libro y 3 edificios de biblioteca separados.
- Edificio 1: libros clasificados por autor,
- Edificio 2: libros clasif por titulo,

El uso de índices proporciona varios caminos de acceso a un archivo

Indices □ definición

Herramienta

para encontrar registros en un archivo.

Consiste de un campo de

Tabla

que opera con un procedimiento que acepta referencia que indica donde

Estructura de

datos. Cada registro dentro del archivo de datos tiene una entrada usada para decrementar el tiempo de acceso a un archivo que permite la rápida

Indice □ Definición

Índice: equivale a índice temático de un libro

Característica fundamental

(tema
,
#hoja
{clave
,

NRR/d
Estruct
instanci
ura
d en
mas
bytes)
simpl
e es

Permite
arbol
imponer
orden en
un archivo
sin que
realmente
este se
r

Indices □ Ejemplo

Dir. Reg.	Cía	Nº ID	Título	Compositores	Artista
32	LON	2312	Romeo y Julieta	Prokofiev	Maazel
77	RCA	2626	Cuarteto en Do...	Beethoven	Julliard
132	WAR	23699	Touchstone	Corea	Corea
167	ANG	3795	Sinfonía Nº 9	Beethoven	Giulini
211	COL	38358	Nebraska	Springsteen	Springsteen
256	DG	18807	Sinfonía Nº 9	Beethoven	Karajan
300	MER	75016	Suite el Gallo...	Rymsky-Korsakov	Leinsdorf
353	COL	31809	Sinfonía Nº 9	Dvorak	Bernstein
396	DG	139201	Concierto para Violín	Beethoven	Ferras
422	FF	245	Good News	Sweet Honey in..	Sweet Honey in..

Indices □ ejemplo

Llave primaria: cía grabadora + N° de identificación de la cía

- Forma canónica: cía en mayúsculas + N° identificación
- No se puede hacer búsqueda binaria sobre el archivo ya que tiene reg. de longitud variable (no se puede usar en NRR como medio de acceso)

Dos Archivos: índice y datos

- Se construye un índice: llave de 12 caracteres (alineada a izq. y completada con blancos) más un campo de referencia (dir. del primer byte del registro correspondiente)
- Estructura del índice: archivo ordenado de reg. de long fija (puede hacerse búsqueda binaria).
- En memoria
- Más fácil de manejar que el arch. de datos

Indices ☐ ejemplo

Llave	Ref
ANG3795	167
COL31809	353
COL38358	211
DG139201	396
DG18807	256
FF245	442
LON2312	32
MER75016	300
RCA2626	77
WAR23699	132

Dir. de registro

Registro de Datos

32	LON 2312 Romeo y Julieta Prokofiev...
77	RCA 2626 Cuarteto en Do...
132	WAR 23699 Touchstone Corea...
167	ANG 3795 Sinfonía N°9 Beethoven...
211	COL 38358 Nebraska Springsteen...
256	DG 18807 Sinfonía N° 9 Beethoven...
300	MER 76016 Suite El gallo de Oro Rimsky...
353	COL 31809 Sinfonía N°9 Dvorak...
396	DG 139201 Concierto para violín Beethoven...
422	FF 245 Good News Sweet Honey in the....

Indices □ como implantarlos?

Operaciones básicas en un archivo indizado

- Índice en memoria (búsqueda binaria + rápida, comparada con archivos clasificados)
- Crear los archivos (el índice y el archivo de datos se crean vacíos, solo con registro cabecera)
- Cargar el índice en memoria (se supone que cabe, ya que es lo suficientemente pequeño. Se almacena en un arreglo)
- Reescritura del archivo de índice (cambios □ reescribir)

Indices □ como implantarlos?

Agregar nuevos registros

- Implica agregar al archivo de datos y al archivo de índices
- Archivo de datos: copiar al final (se debe saber el NRR (fija) o distancia en bytes (variable) para el índice)
- Índice ordenarse con cada nuevo elemento en forma

Eliminar un registro

- Arch. datos □ Cualquier técnica de las vistas para reutilizar el espacio
- Arch. índices □ se quita la entrada (ó se podría marcar como borrado).

Indices □ como implantarlos?

Actualización de registros

- Sin modificar la clave (que pasa con el índice?)
 - Si el registro no cambia de longitud, se almacena en la misma posición física, el índice “no se toca”.
 - Si el reg. cambia de longitud (se agranda) y se reubica en el arch. de datos
 - se debe guardar la nueva posición inicial en el índice
 - Si se trata de long. Fija, no hay que hacer mas actividad
- Modificando la clave (que sucede?)
 - Se modifica el archivo de datos
 - Se debe actualizar y reorganizar el archivo de índices
 - Cómo simplificar □ Modificar = Eliminar + Agregar (ya vistos)

Indices □ Resumen

- Ventajas
 - Se almacena en memoria principal
 - Permite búsqueda binaria
 - El mantenimiento es menos costoso
- Desventajas
 - Si no caben en memoria RAM?
 - Reescritura del archivo de índices?
 - Persistencia de datos

Indices □ Persistencia de Datos

- **Índices grandes para entrar en memoria □ Soluciones**
 - Uso de Árboles
 - Organización por Dispersión (hashing)

Indices secundarios

Índices Secundarios

No
sería
natura
En su
lugar
souetta
se un
dalla
como
porta
ment
ave
un
om os
e am
ariva
perien
che cen
e q
de
reco si
rivel
lo el p
dir
la
ave ar
scu h
arig i
uede

Indices secundarios

Indice de	Compositores
Llave Secundaria	Llave Primaria
BEETHOVEN	ANG3795
BEETHOVEN	DG139201
BEETHOVEN	DG18807
BEETHOVEN	RCA2626
COREA	WAR23699
DVORAK	COL31809
PROKOFIEV	LON2312
RIMSKY-KORSAKOV	MER75016
SPRINGSTEEN	COL38358
SWEET HONEY....	FF245

Indices secundarios

Problemas: la repetición de información

- El arch. de índices se debe reacomodar con cada adición, aunque se ingrese una clave secundaria ya existente, dado que existe un 2do orden por la clave primaria.
- Misma clave varias ocurrencias, en distintos registros
 - Se desperdicia espacio
 - Menor posibilidad de que el índice quepa en memoria

Indices secundarios

Soluciones

BEETHOVEN ANG3795 DG139201 DG18807 RCA2626

- Arreglo: clave + vector de punteros con ocurrencias

- Al agregar un nuevo reg. de una clave existente no se debe recomendar nada > solo recomendar el vector de ocurrencias

Problema: elección del tamaño del vector.

- Tamaño fijo
 - Puede haber casos en que sea insuficiente
 - Puede haber casos que sobre espacio, provocando fragmentación interna

- Mejora: clave + lista de punteros con ocurrencias

Indices secundarios

Listas invertidas:

Archivos en los que una llave secundaria lleva a un conjunto de una o más claves.
No se pierde espacio (no primarias □ lista de referencias de claves primarias)

Si se agrega un elem. a la lista

□ no se necesita una reorganización completa

Indices secundarios

Listas invertidas

Organización
física

Operaciones

Archivos
secundarios
Marcas o
referencias
Agregar un
nuevo
eliminar
consolidar
borrar
reemplazar
Actualizar
Relaciones
Concurrencias
dependencias
en la lista
invertida

Indices secundarios

NRR	Archivo de índice secundario	
0	BEETHOVEN	3
1	COREA	2
2	DVORAK	7
3	PROKOFIEV	10
4	RIMSKY-KORSAKOV	6
5	SPRINGSTEEN	4
6	SWET HONEY IN...	9

NRR	Arch de listas de llaves primarias	
0	LON2312	-1
1	RCA2626	-1
2	WAR23699	-1
3	ANG3795	8
4	COL38358	-1
5	DG18807	1
6	MER76016	-1
7	COL31809	-1
8	DG139201	5
9	FF245	-1
10	ANG36193	0

Indices secundarios

- **Ventajas**

- El único reacomodamiento en el arch. índice -> al agregar o cambiar un nombre
- Borrar o añadir grabaciones para un compositor -> sólo cambiar el archivo de listas
- Como el reacomodamiento es a bajo costo se podría almacenar el arch. índice en mem. secundaria , liberando RAM

- **Desventaja**

- el arch. de listas es conveniente que esté en memoria ppal. porque podría haber muchos desplazamientos en disco □ costoso si hay muchos índices secundarios