

Base de Datos 1



Índices sobre archivo binarios Un caso Practico

Responsable: Prof. Cherencio Guillermo

JTP: Prof. Romero Juan Carlos

Colaboradores: Prof. Racker José

Prof. Perello Mario

T.P. 1B - Consideraciones



- Estructura del archivo de Datos del TP 1B
 - CUIT (Big Long) (Entero de 8 bytes) (campo clave)
 - RazonSocial (String de 50)
 - Domicilio (String 50)
 - Localidad (String 30)
 - TE (Big Long)
 - Celular (Big Long)
 - Mail (String 100)

T.P. 1B - Consideraciones



- Implementación
 - Se deben mantener 2 archivos: Archivo de Datos y el Archivo de Indices.
 - **Archivo de Datos**: archivo binario con la estructura solicitada mas al menos un campo adicional para marcar los eliminados.
 - **Archivo de Indices**: archivo binario con la estructura adecuada para poder mantener en disco un índice de tipo denso representado mediante un árbol binario de búsqueda.

T.P. 1B - Consideraciones



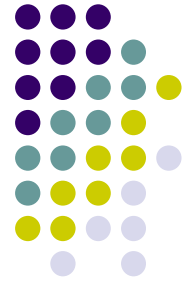
- Implementación
 - **Archivo de Datos**: Se puede crear de antemano el archivo con el universo de claves o bien se lo puede ir haciendo crecer a medida que se insertan datos.
 - Las *eliminaciones son marcadas como borradas*, es decir NO se elimina físicamente el registro, es un borrado lógico.
 - Si el archivo es generado de antemano se debe tener una marca al **ultimo registro utilizado** para luego insertar el nuevo registro cuando hay una inserción o alta de cliente.
 - Tanto para las eliminaciones como para las modificaciones *primero se busca la clave en el archivo de índices* y luego se apunta al registro de datos.

T.P. 1B - Consideraciones



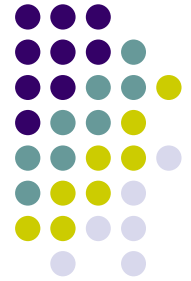
- Implementación
 - **Archivo de Índices:** Se puede crear de antemano el archivo con el universo de claves o bien se lo puede ir haciendo crecer a medida que se insertan datos.
 - Las eliminaciones son marcadas como borradas, es decir NO se elimina físicamente el registro, es un borrado lógico (*teniendo en cuenta la lógica de la eliminación en árboles binarios de búsqueda mas algunas consideraciones especiales que detallamos mas adelante*).
 - Si el archivo es generado de antemano es mas sencillo ya que para agregar una clave solo hay que ver de que hoja del árbol se cuelga la nueva clave (*se puede tener un apuntador al ultimo registro utilizado para ir rápidamente a la próxima posición disponible para insertar*).
 - Si el archivo crece a medida que agregamos claves también es necesario tener una *ubicación del ultimo registro o EOF* para poder poner rápidamente una clave nueva.

T.P. 1B - Consideraciones



- Implementación
 - **Estructura del Archivo de Índices:**
 - Clave (CUIT)
 - Apuntador al Hijo Izquierdo
 - Apuntador al Hijo Derecho
 - Apuntador al Registro de Datos
 - Marca de Eliminado

T.P. 1B - Consideraciones



- Consideraciones Especiales

- Archivo de Índices:

- La **raíz** siempre debe estar el registro **CERO**. Esto se debe contemplar especialmente en la *eliminación* cuando la raíz podría ser eliminada por tener *un solo hijo* (a diferencia de la implementación del árbol en memoria)

T.P. 1B - Consideraciones



- Proceso de Inserción (Alta de un cliente)
 - Buscar la clave en el archivo de índices, si NO existe entonces es posible agregar el cliente.
 - Agregar el registro en el archivo de datos (al final del archivo o en la próxima posición libre disponible a la cual se tenga referencia)
 - Agregar la clave en un hoja del árbol del archivo de índice.

T.P. 1B - Consideraciones



- Proceso de Inserción (Agregar clave archivo de índice)
 - Ubicarse en el registro CERO (raíz). Recuperar el CUIT.
 - Compara CUIT a agregar, si es menor tomar el Hijo Izquierdo y apuntar a ese registro (Hacer Seek y recuperar clave). Caso contrario hacer lo mismo por el lado derecho (hijo derecho). Iterar hasta llegar a una hoja.
 - Cuando llegamos a una Hoja (Apuntador del hijo izquierdo que NO apunta a nada, por ejemplo marcado como "-1" o derecho), debemos tomar el proximo registro disponible y agregar el registro en el indice y colgar el apuntador correspondiente del padre (similar a lo que se hace en memoria pero enlazando o encadenando registros de un archivo)

T.P. 1B - Consideraciones



- Proceso de Eliminacion (Eliminar una clave archivo de índice)
 - Ubicarse en el registro CERO (raíz). Recuperar el CUIT.
 - Compara CUIT a buscar, si es menor tomar el Hijo Izquierdo y apuntar a ese registro (Hacer Seek y recuperar clave). Caso contrario hacer lo mismo por el lado derecho (hijo derecho). Iterar hasta encontrar la clave o bien NO se encuentra el cliente a eliminar.
 - Si ubicamos el cliente tener en cuenta:
 - Si es la raíz y es el único registro, se elimina directamente.
 - Si es un nodo con un solo hijo o es una hoja, se marca como borrado (*siempre que NO sea la raíz*)
 - Si es un nodo con 2 hijos validos, entonces debemos ubicar la clave a reemplazar al igual que el proceso en memoria y reacomodar todos los apuntadores tal cual es en memoria pero con los registros del archivo.

T.P. 1B - Consideraciones



- Listado en In-Orden de las claves
 - Al igual que en memoria se debe realizar un recorrido en **IN-ORDEN** para obtener el listado de clientes ordenados por su CUIT.
 - El recorrido puede ser “**recursivo**” o “**iterativo**”, en ambos casos la lógica es la misma que para el recorrido en memoria solo que en vez de apilar direcciones de memoria (punteros) se *apilan enteros que apuntan a registros del archivo*.
 - Por cada clave recorrida se debe realizar el “Seek” + “Read” al archivo de datos para listar el registro.
 - El recorrido finaliza cuando la pila de la recursividad o la pila del recorrido iterativo este vacía y no quede nada por procesar.

T.P. 1B – Ejemplo de Inserciones



- Supongamos insertar consecutivamente las siguiente claves:
 - 20-7789123-6,
 - 27-12345678-9,
 - 20-2154730-1,
 - 30-70754877-3,
 - 30-45678901-2
- Todas las clave son usadas sin los guiones, representado el CUIT.
- Asumiendo que partimos del archivo vacío, en la próxima PPT se muestra ambos archivos como quedarían

T.P. 1B – Ejemplo de Inserciones



Archivo de Indice					Archivo de Datos	
Reg. A. Ind	HI (Nro Reg)	CLAVE	R.A.D.	HD (Nro Reg)	Nro Reg.	CUIT
0	2	20-7789123-6	0	1	0	20-7789123-6
1	-1	27-12345678-9	1	3	1	27-12345678-9
2	-1	20-2154730-1	2	-1	2	20-2154730-1
3	4	30-70754877-3	3	-1	3	30-70754877-3
4	-1	30-45678901-2	4	-1	4	30-45678901-2
		(0)				
(2)		20-7789123-6		(1)		
20-2154730-1				27-12345678-9	(3)	
				(4)	30-70754877-3	
				30-45678901-2		

T.P. 1B – Ejemplo de Eliminacion



Archivo de Indice					Archivo de Datos	
Reg. A. Ind	HI (Nro Reg)	CLAVE	R.A.D.	HD (Nro Reg)	Nro Reg.	CUIT
0	2	20-7789123-6	0	1	0	20-7789123-6
1	-1	27-12345678-9	1	3	1	27-12345678-9
2	-1	20-2154730-1	2	-1	2	20-2154730-1
3	4	30-70754877-3	3	-1	3	30-70754877-3
4	-1	30-45678901-2	4	-1	4	30-45678901-2
		(0)				
(2)		20-7789123-6		(1)		
20-2154730-1				27-12345678-9		(3)
				(4)		30-70754877-3
				30-45678901-2		

Borramos la raiz