

**Ejercicio 1:**

USR ([] , s) = s

USR (t: r , s) = if t ∈ r || t ∈ s then USR (r,s) else t: USR (r, s)

Definición usando operadores estudiados.

USR (r, s) = v (r ++ s)

**Operador físico:**

Primero concateno r y s: r ++ s

ordeno r ++ s removiendo duplicados

estimación de costo:

para r ++ s :

$b_r + b_s$  transferencia de bloques

$b_r + b_s$  accesos a bloques (cuando todos los bloques de r y s no están contiguos)

para ordenación removiendo duplicados: peor caso: no hay duplicados es el costo de la ordenación. Tomamos u = r++ s.

$b_u * (2 * \text{techo}(\log_{M-1}(b_u / M)) + 1)$  TB

$2 * \text{techo}(b_u / M) + b_u * (2 \text{techo}(\log_{M-1}(b_u / M)) - 1)$  AB

**Ejercicio 2:**

1. Cantidad de claves de búsqueda:

$$4096 = N * \text{tamaño nombre} + (N+1) * 8 = N * 40 + (N + 1) * 8 = 48 N + 8$$

$4088 = 48 N$ ,  $N = 85.16$  Luego  $N = 86$  claves de búsqueda por nodo.

Altura del árbol B+

$$\begin{aligned} 2. \text{ Altura de árbol B+} &= h = \text{techo} (\log_{43} (10000)) \\ &= \text{techo} (\log_{10} (10000) / \log_{10} (43)) \\ &= \text{techo} (4 / 1,63) = \text{techo} (2,45) = 3 \end{aligned}$$

3. Si un nombre no se repite más de 15 veces, como el índice es primario en nombre, la tabla usuarios esta ordenada por nombre.

Una fila de usuarios contiene: 40 (nombre) + 6 (uid) + 8 (puntero) = 54 B

En un bloque de usuarios entran =  $4096/54 = 75,85 = 75$  tuplas.

Luego en el peor caso hay que acceder a dos bloques para encontrar todos los usuarios de un mismo nombre.

El costo se compone de:  $(h + 2)$  transferencias de bloques.

Ejercicio 3:

	persona	bibliotecario	Trabaja_en
Tamaño	1000	400	1500
Costo	0	0	0
Mejor plan	persona	bibliotecario	Trabaja_en

	{persona, bibliotecario}	{persona, trabaja_en}	{bibliotecario, trabaja_en}
Tamaño	400	1500	800
Costo	0	0	0
Mejor plan	Bibliotecario $\bowtie$ persona	Persona $\bowtie$ trabaja_en	Bibliotecario $\bowtie$ trabaja_en

$F_s(\text{bibliotecario.DNI} == \text{persona.DNI}, \text{bibliotecario}, \text{persona})$

$$= 1/\max(1000, 400) = 1/1000$$

$F_s(\text{persona.DNI} == \text{trabaja_en.DNI}, \text{persona}, \text{trabaja_en})$

$$= 1/\max(1000, 750) = 1/1000$$

$F_s(\text{bibliotecario.DNI} == \text{trabaja_en.DNI}, \text{bibliotecario}, \text{trabaja_en})$

$$= 1/\max(400, 750) = 1/750$$

Pasemos a  $n = 3$

Costo  $\{\{\text{persona}, \text{bibliotecario}\}, \{\text{trabaja_en}\}\}$

$$= \text{costo } (\text{bibliotecario} \bowtie \text{persona}) \bowtie \text{trabaja_en}$$

$$= \text{tamaño bibliotecario} \bowtie \text{persona} = 400$$

Costo  $\{\{\text{persona}, \text{trabaja_en}\}, \{\text{bibliotecario}\}\}$

$$= \text{costo } (\text{persona} \bowtie \text{trabaja_en}) \bowtie \text{bibliotecario}$$

$$= \text{tamaño persona} \bowtie \text{trabaja_en} = 1500$$

Costo  $\{\{\text{bibliotecario}, \text{trabaja_en}\}, \{\text{persona}\}\} =$

$$= \text{costo } (\text{bibliotecario} \bowtie \text{trabaja_en}) \bowtie \text{persona}$$

= tamaño bibliotecario  $\bowtie$  trabaja\_en = 800

$\{\{persona, bibliotecario\}, \{trabaja\_en\}\}$  da el menor costo

Entonces el mejor plan es (bibliotecario  $\bowtie$  persona)  $\bowtie$  trabaja\_en

Calculemos el tamaño:

$$|(bibliotecario \bowtie persona) \bowtie trabaja\_en|$$

$$= |\text{bibliotecario} \bowtie \text{persona}| * |\text{trabaja\_en}| *$$

$$\text{fs}((\text{bibliotecario} \bowtie \text{persona}). \text{DNI} == \text{trabaja\_en}. \text{DNI}, \text{bibliotecario}, \text{persona})$$

$$= 400 * 1500 / \max(1000, 400) = 400 * 1500 / 1000 = 40 * 15 = 600$$

	{bibliotecario, persona, trabaja_en}
Tamaño	600
Costo	400
Mejor plan	(bibliotecario $\bowtie$ persona) $\bowtie$ trabaja_en

#### Ejercicio 4:

1. para cada término: una tupla de elemento de lista de ocurrencias asociada a un término tiene identificador de documento (donde aparece el término) y frecuencia de documento o TF-IDF. Esta información me permite calcular la relevancia de cada documento que satisface la consulta y ordenar los resultados por relevancia.
2. Solución:
  - a. Buscar el vocabulario de la consulta.
  - b. Calcular el vector de la consulta.
  - c. Para cada término se busca la información de documentos en el índice invertido ( $S'i$  para término  $i$ ).
  - d. Calcular los  $S_i$  (a partir de  $S'i$ ).
  - e. Evaluar la expresión de consulta booleana aplicando los operadores. Se obtiene así un conjunto de documentos.
  - f. Calcular similitud de vectores de cada documento arrojado por la búsqueda y el vector de la consulta.