**CONSIGNA 2C 2019 3Op**

5) Dada la siguiente colección de documentos:

D1: AZUL MARRON AZUL ROSA

D2: VERDE AZUL AMARILLO ROJO BLANCO

D3: VERDE NEGRO BLANCO

D4: BLANCO NEGRO

D5: VERDE VERDE NEGRO VERDE

Construya un índice invertido, indicando el paso a paso en la construcción del mismo y seleccionando un método de almacenamiento de punteros y del léxico que considere adecuado para este caso, justificando su elección. En el índice construído, resuelva la consulta por frase "VERDE NEGRO", explicando el paso a paso en su resolución y contabilice la cantidad de accesos necesarios para su resolución. (\*\*\*) (15 pts)

**RESOLUCIÓN**

Para construir el índice invertido primero se deben tomar los términos ordenados alfabéticamente, y otorgarles un formato del estilo:

Termino: Dn: Freq (Pos 1, …, Pos m), con m=Freq

Con las posiciones decodificadas como distancias respecto de la anterior.

Entonces, de esta forma se permiten consultas por proximidad, y construido queda:

* Amarillo: D2:1(3)
* Azul: D1:2(1,2), D2:1(2)
* Blanco: D2:1(5), D3:1(3), D4:1(1)
* Marron: D1:1(2)
* Negro: D3:1(2), D4:1(2), D5:1(3)
* Rojo: D2:1(4)
* Rosa: D1:1(4)
* Verde: D2:1(1), D3:1(1), D5:3(1,1,2)

Teniendo la lista de términos puedo construir el código utilizando léxico concatenado, y codificación gamma para los documentos.

El léxico concatenado queda:

Y la codificación Gamma:

La tabla finalmente queda:

|  |  |  |
| --- | --- | --- |
|  | Término | Documentos |
| 0 | 0 | 0 |
| 1 | 8 | 7 |
| 2 | 12 | 20 |
| 3 | 18 | 37 |
| 4 | 24 | 42 |
| 5 | 29 | 59 |
| 6 | 33 | 68 |
| 7 | 37 | 75 |

Índice invertido construido

Para resolver el query de “Verde Negro” debo primero buscar a Verde en el índice con búsqueda binaria:

1. Acceso a Marrón en disco, no es el que busco (1 acceso a índice, y 1 a disco).
2. Acceso a Rojo en disco, no es el que busco (1 acceso a índice, y 1 a disco).
3. Acceso a Rosa en disco, no es el que busco (1 acceso a índice, y 1 a disco).
4. Acceso a Verde, si es (1 acceso a índice, y 1 a disco).
5. Habiendo encontrado a verde, accedemos a la posición donde se encuentran los documentos (1 lectura de disco).

Finalmente, los documentos que corresponden con el query de verde serán:

D2:1(1), D3:1(1), D5:(1,2,4)

Procediendo con el mismo query para Negro:

1. Acceso a Marrón (Ya está en memoria), no es.
2. Acceso a Rojo (Ya está en memoria), no es.
3. Acceso a Negro (1 acceso a índice, y 1 a disco), si es
4. Realizamos 1 lectura en disco para encontrar los documentos a los cuales pertenece.

El resultado del query entonces para negro es:

D3:1(2), D4:1(2), D5:1(3)

Entonces, buscamos los documentos que tengan verde en una posición n, y negro en n+1.

Viendo el resultado se concluye que D2 Y D4 quedan descartados por no estar en ambos resultados. Ahora, viendo D3, es un resultado positivo, pues Verde esta en la pos 1, y Negro en la pos 2. Y para D5, también es positivo, pues verde esta en la pos 2, y negro en la pos 3.

Respuesta final: D3 y D5 cumplen el Query

Accesos a memoria totales: 5 índice + 7 Disco = 12 Lecturas.