

- a) Resolver la consulta “aquel arbusto azul” indicando detalladamente cada paso. Informar la cantidad de accesos a disco necesarios para la resolución de la misma.

Observamos claramente que el léxico está en front coding parcial, por lo tanto vamos a tener que formar las palabras con ayuda de las anteriores, a menos que tengan 0 letras compartidas con su anterior.

Formamos las palabras:

Para poder formar las palabras realizamos búsqueda binaria en la estructura del índice invertido.

Aquel: Empezamos con la búsqueda binaria la cual nos dirige a la posición 3. La cual nos indica que debemos leer desde el léxico 20 hasta el 24 (**1 acceso a disco**). Como no comparte letras con el anterior podemos formar la palabra completa. La cual es “Aquel” Al ser la palabra buscada averiguamos cuales son los docs que contienen dicha palabra. Accedo al puntero de documentos (**1 acceso a disco**) y caigo en la posición 14 la cual sólo contiene un “1”.

Por lo tanto la palabra aquel solo está contenida en el **documento 1**

Arbusto: Como en el caso anterior arrancamos en la mitad de la tabla, en la cual se encuentra la palabra “aquel” que ya conocemos. Al no ser la palabra buscada volvemos a buscar en la nueva mitad de la tabla (mitad inferior). Ahora estamos posicionados en nuestra nueva posición, que nos indica que debemos leer desde el léxico 29 hasta el 32 (**1 acceso a disco**) y para saber cuál palabra es debemos descifrar la palabra anterior, ya que comparte 3 chars. La palabra anterior nos indica que debemos leer desde el léxico 25 hasta el 28 (**1 acceso a disco**), y está también comparte chars con la palabra anterior, pero dicha palabra es “aquel” por lo tanto no debemos volver a acceder a dicha posición. Con esto podemos formar la palabra “árbol” (léxico a partir del 25) y con esta información podemos formar “arbusto” que es la palabra buscada.

Cómo encontramos la palabra arbusto accedo al puntero de documentos (**1 acceso a disco**) en la posición 21. Obtenemos los docs 1 y 3, ya que están codificados en distancias y primero tenemos un “1”, luego un “2” por lo tanto el doc 3 tiene distancia 2 con el documento 1. La palabra arbusto está contenida en los **documentos 1 y 3**.

Azul: Como en los casos anteriores arrancamos en la mitad de la tabla, en la cual se encuentra la palabra “aquel” que ya conocemos. Al no ser la palabra buscada volvemos a buscar en la nueva mitad de la tabla (mitad inferior). Ahora estamos posicionados en nuestra nueva posición, que nos indica que debemos leer desde el léxico 29 hasta el 32, es la palabra “arbusto” que ya conocemos. Al no ser la palabra buscada volvemos a partir a la mitad la tabla y caemos en la posición del léxico 33 (**1 acceso a disco**) y observamos que es la palabra buscada. Por lo tanto accedemos a los punteros de documentos (**1 acceso a disco**), en la posición 25, observamos que la palabra azul está en los **documentos 1, 3 y 4**

En total tenemos 7 accesos a disco.

Y para saber qué documentos contienen las 3 palabras hacemos la intersección de documentos. Obtenemos que **el documento 1 contiene las 3 palabras**.

- b) Para la consulta rankeada "aquel arbusto azul" determinar el TF.IDF de cada documento resultado de la búsqueda e indicar cómo quedaría el orden de los documentos resultado de la misma.

Hablamos de documentos 1, 3 y 4. Pero también hay que tener en cuenta que existe un documento 2. Pero al no tener ninguno de los términos no es tomado en cuenta en el ranking

Como no tenemos frecuencias de cada término asumimos que aparecen solo 1 vez. Debemos calcular el $IDF = \log(N+1/ft_i)$ para cada término.

Calculamos:

Aquel : $\log(5/1) = 0,7$

Arbusto: $\log(5/2) = 0,4$

Azul: $\log(5/3) = 0,22$

Rankeamos los documentos:

doc1: $1 * 0,7 + 1 * 0,4 + 1 * 0,22 = 1,3$

doc3: $1 * 0,4 + 1 * 0,22 = 0,62$

doc4: $1 * 0,22 = 0,22$

El ranking queda:

1ro: Doc 1, 2do: doc 3 y 3ro doc4