

EDA – Preguntas Lab 6

Andrés Mugnier, 201729994

David Burgos ,201818326

Parte 1:

- Teniendo en cuenta cada uno de los requerimientos ¿Cuántos índices implementaría en el Reto? y ¿Por qué?

Req1 – El mejor índice para este requerimiento es por el año de nacimiento de los artistas debido a que se puede ir uno por uno en el rango y se encuentran directamente.

Req2 – En este caso el agrupar por fechas es poco eficiente debido a que es poco común que varias obras compartan una misma fecha específica, se podría utilizar solo el año para no tener una cantidad excesiva de mapas individuales

Req3 - La mejor opción de índice es el medio de la obra ya que se pueden agrupar varias obras en un solo mapa y encontrar una cantidad específica de obras del artista

Req4 - en este caso es muy útil el clasificar los artistas por ID debido a que se tienen que recorrer las obras y esta es la forma más fácil de acceder a sus artistas

Req5 - Aquí se pueden organizar las obras con índice de su departamento para encontrar las obras a mover más fácil

Req6 - se necesita cambiar un poco la carga de datos, pero si se añade un valor de cuantas obras tiene se puede organizar por este criterio el mapa

- Según los índices propuestos ¿en qué caso usaría Linear Probing o Separate Chaining en estos índices? y ¿Por qué?

Linear – R2, R4, R6 – como la llave varía demasiado entre elementos es más fácil adoptar esta medida

Separate – R1, R3, R5 - En este caso nos importa recorrer los elementos que cumplan cierta condición rápidamente, por eso el separate chaining es más efectivo.

- Dado el número de elementos de los archivos MoMA, ¿Cuál sería el factor de carga para estos índices según su mecanismo de colisión? (Artistas, Large) -

En los linear buscamos un factor de 0.7 lo que equivale a 15224 / 21751 (primo de 21748)

Y en los separate buscamos un factor de 3 (o hasta más dependiendo del caso) lo que equivale a 15224 / 5077 (primo de 5074)

Parte 2:

Se repitieron las pruebas tres veces y se obtuvieron los siguientes resultados:

Con Linear Probing y un factor de carga de 0.5 se obtuvo un tiempo de 7520.84 milisegundos.

Con Separate Chaining y un factor de carga 4 se obtuvo un tiempo de 7515.63 milisegundos.

- ¿Qué diferencias en el tiempo de ejecución notan al ejecutar la carga los datos al cambiar la configuración de Linear Probing a Separate Chaining?

Viendo los tiempos presentador anteriormente, se puede ver que al cambiar de Linear Probing con un factor de carga de 0.5 a Separate Chaining con un factor de carga, se redujo el tiempo en 5.21 milisegundos. Es decir, Separate Chaining fue más rápido por muy poco.

Parte 3:

Se repitieron las pruebas y se obtuvieron los siguientes resultados:

Tiempos obtenidos con Linear Probing con diferentes factores de carga			
Factor de carga	0.2	0.5	0.8
Tiempo de ejecución (miliseg)	7760.42	7520.84	7520.83

Tiempos obtenidos con Separate Chaining con diferentes factores de carga			
Factor de carga	2	4	8
Tiempo de ejecución (miliseg)	7541.66	7515.63	7546.87

- ¿Qué configuración de ADT Map escogería para el índice de técnicas o medios?, especifique el mecanismo de colisión, el factor de carga y el numero inicial de elementos.

Como se obtuvieron mejores tiempos utilizando Separate Chaining, utilizaremos Separate Chaining en ambos casos. Para el caso del índice de medios, con el código obtuvimos que hay 21251 medios distintos en el archivo large, por lo que si queremos tener un factor de carga cercano a 1 (es lo ideal) podemos iniciar el tamaño de la tabla en el primo más cercano a 21251 que es 21269, obteniendo así un factor de carga de 0.99.

Hay 21251

- ¿Qué configuración de ADT Map escogería para el índice de nacionalidades?, especifique el mecanismo de colisión, el factor de carga y el numero inicial de elementos

Como se obtuvieron mejores tiempos utilizando Separate Chaining, utilizaremos Separate Chaining en ambos casos. Para el caso del índice de nacionalidades, con el código obtuvimos que hay 116 nacionalidades distintas en el archivo large, por lo que si queremos tener un factor de carga cercano a 1 (es lo ideal) podemos iniciar el tamaño de la tabla en el primo más cercano a 116 que es 127, obteniendo así un factor de carga de 0.91.