

OBSERVACIONES LAB 6

Nicholas Barake 202020664

Jesed Domínguez 202011992

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

Se piensan implementar un total de tres mapas, dos que son los originales de artista y obras cuando se cargan los datos por primera vez, y el tercero siendo el de nacionalidades. No se crean mapas para los requerimientos de las técnicas por artista, el transporte, ni el más prolífico, por lo que se planea retornar listas. Pero se tiene en mente crear copias de los mapas iniciales para estos requerimientos, para evitar “dañarlos” en la creación de las listas.

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

Se utilizará el método de Separate Chaining para todos los índices, por lo que se considera que es más eficiente de memoria, ya que no conocemos bien el número de datos en cada archivo de datos. Pero esto puede ir cambiando a medida de que se aumenta el tamaño de los archivos.

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

Como los archivos MoMA se saben que contienen muchos datos, en especial los más grandes, el factor de carga seguramente es mayor que uno. Ya que se está intentando preservar memoria y se está utilizando Chaining como método de solución de colisiones, se guardarán varios datos en un espacio más pequeño. Esto significa que las listas van a ser un poco más largas.

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

Al cargar los dos archivos más grandes, al parecer los tiempos de ejecución de la carga de datos no son muy distintos (o no lo suficientemente distintos para no ser intercambiables entre sí). Aunque sean muy similares, si se puede distinguir que con Linear Probing la carga toma un poco menos que con Chaining. Esto es, una diferencia de 100 ms en el archivo de 80pct y una diferencia de 1000 ms en el archivo large.

Notas:

- Para PROBING: entre más pequeño el factor de carga, más se demora.
- Para CHAINING: el factor de carga no afecta mucho el tiempo de carga.
- Para AMBOS: entre más grande los archivos, más se demora. (obvio)

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

Se escogería el Probing ya que mostró tiempos de ejecución más rápidos al cambiar el factor de carga, el mejor nos resultó 0.7. Además, como hay muchas más llaves en el mapa de técnicas, tendría más sentido solucionar la colisión revisando cuales técnicas ya están. La cantidad inicial sería un poquito menos que el tamaño del archivo de las obras, ya que no muchas de las técnicas se repiten.

f) ¿Qué configuración de ADT Map escogería para el índice de nacionalidades?, especifique el mecanismo de colisión, el factor de carga y el número inicial de elementos.

Escogeríamos el Chaining debido a que los buckets hacen que sea óptimo al agregar una misma llave, el factor de carga que usaríamos sería 4.0 y el número inicial de elementos sería 198 ya que esos son la cantidad de países más algunas llaves que pueden ser ‘Unknown’.