

OBSERVACIONES DEL LA PRACTICA

Estudiante 1: Sergio Iván Rincón Carvajal Cod 201914107

Estudiante 2: Luis Ernesto Tejón Rojas Cod 202113150

a)

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

- **Requerimientos 1 y 2:** Para los dos primeros requerimientos no vale la pena implementar índices con tablas de Hash. Esto se debe a que en estos requerimientos se requiere de consultar no un valor específico si no un rango de años, para el cual la manera en la que ya están hechos los requerimientos, haciendo uso de listas organizadas, puede no tener la menor complejidad en la carga de datos, pero sí una complejidad muy baja en la consulta ($2\log n$), por lo que sería difícil bajar la complejidad temporal, por lo menos sin hacer uso de árboles.
- **Requerimiento. 3:** La idea para este requerimiento consiste en generar un índice de autores en donde, a su vez, cada autor contiene un subíndice que indica la cantidad de obras por medio y otros datos asociados convenientes para el requerimiento y una lista de obras organizadas por medio. Lo anterior se haría con el fin de encontrar rápidamente el medio más empleado y las obras de este mismo.
- **Requerimiento 4:** En la implementación original de este requerimiento se usaban diccionarios normales de Python para llevar el conteo del número de obras por nacionalidad y el número de nacionalidades en general, naturalmente lo mejor será cambiar estos diccionarios por TAD map para así hacer uso de las librerías dadas con lo que estos mapas pasarían a cumplir la misma función que hacían antes los diccionarios.
- **Requerimiento 5:** En este caso se crearía un índice por departamentos, en donde cada departamento contuviese un índice pequeño con información del precio calculada en la

carga de datos, y una lista de obras del departamento organizada por antigüedad. El objetivo de esto sería obtener rápido el precio de transporte junto con las obras más antiguas.

- **Requerimiento 6:** Para el bono aprovecharemos el merge que hicimos a la lista de artistas por 'BeginDate', el cual se usó para dar un rango de artistas por año en el requerimiento 1, esto haciendo uso de dos búsquedas binarias, luego haremos un mapa de artistas agregando cada obra de ese artista a una lista que será el valor asociado al key del artista en el mapa.

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

Linear Probing es ideal cuando se busca gran velocidad, pero se ocupa una buena cantidad de espacio, lo que podría ser contraproducente con muchos datos. Por su parte, Separate Chaining puede ser un poco menos rápido que Linear Probing, pero garantiza salvaguardar cierta memoria. En este caso, y como la cantidad de datos no es tan grande y tanto los departamentos, nacionalidades y autores no son tantos, se puede pensar en implementar Linear Probing en todos los casos. Esto se debe a que se está buscando la mejor eficiencia temporal sin tener en cuenta la memoria. Ya si se hablase de un índice de obras y una cantidad de datos muy grande, quizá si podría ser ideal salvaguardar memoria usando Separate Chaining.

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

En este caso el factor de carga depende de si la configuración para lidiar con las colisiones es Linear Probing o Separate Chaining. En el caso de emplear una configuración Linear Probing en el índice de medios (clasificado por los IDs de los artistas), el factor de carga inicial es $15220/30011 = 0.507$ ya que se busca que este valor tienda a 0.5. Por otra parte, al emplear

separate chaining el factor de carga es $15220/3821 = 3.9$ cuyo valor tiende a 4. La idea es que las listas encadenadas en separate chaining no queden tan largas para evitar hacer demasiadas comparaciones en un Hit o un Miss. Cabe resaltar que el factor no depende del todo del número de elementos, ya que se establece de tal manera que la búsqueda sea eficiente (menor o igual a 0.5 en linear probing y entre 2 y 8 máximo en separate chaining.)

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

Ambientes de pruebas

	Máquina 1	Máquina 2
Procesadores	AMD Ryzen 5 3550H with Radeon Vega Mobile Gfx 2.10 GHz	Intel(R) Core(TM) i7-6700HQ CPU @ 2.60GHz, 2.61GHz
Memoria RAM (GB)	16GB	12GB
Sistema Operativo	Windows 10 home 64-bits	Windows 10 home 64-bits

Tabla 1. Especificaciones de las máquinas para ejecutar las pruebas de rendimiento.

Maquina 1

Resultados

Porcentaje de la muestra [pct]	Linear Probing[sec]	Separate Chaining[sec]
small	0.11	0.12
large	127.92	143.08

Tabla 2. Comparación de tiempos de carga de datos.

Maquina 2

Resultados

Porcentaje de la muestra [pct]	Linear Probing[sec]	Separate Chaining[sec]
small	0.63	0.57
large	136.77	150.07

Tabla 2. Comparación de tiempos de carga da datos.

- d) ¿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.

Dadas las pruebas de tiempo se ve que es ligeramente más veloz haciendo uso del ADT Map tipo Linear Probing, además en ninguno de los casos de excedió de la cantidad de memoria disponible en la maquina, por lo que lo más lógico es elegir el tipo de mapa que de mayo velocidad.

Factor de carga: 0.5

Número de elementos inicial para medios: 30000

El número inicial de elementos se escogió multiplicando por 2 el número de parejas llave valor que esperabamos guardar, y el factor de carga se eligió 0.5 por mostrar bueno resultados y por la teoría vista en clase.

- e) ¿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.

Al igual que con los medios, y gracias a que no se excede la cantidad de memoria ocupada se eligió nuevamente el mapa tipo Linear Probing con el mismo factor de carga '0.5' y para elegir el número de elementos se siguió la misma regla de escoger como número inicial de elementos el doble del número esperado de parejas llave valor a guardar en el mapa-

Factor de carga: 0.5

Número de elementos inicial para nacionalidades: 120