

Observaciones laboratorio 6

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

- Para los requerimientos 1 y 2, se podría hacer un mapa usando como llaves los años de nacimiento y las fechas de adquisición para luego mirar llave por llave si estas están dentro del rango, pero esto en verdad sería menos eficiente que ordenar un arreglo con los artistas o las obras; podría hacerse la implementación por mapas, pero sería menos eficiente que la hecha por listas en el reto 1. Para el requerimiento 3, yo implementaría en el catálogo un índice que tenga como llaves los nombres de los artistas y como valores sus diccionarios para disminuir la complejidad de buscar el id de un artista, además de un índice que tenga como llaves los id's de los artistas de las obras y como valores todas las obras que tengan ese artista entre sus artistas, para así encontrar la lista de obras del artista que se busca, y con esta hacer un mapa auxiliar con un índice que use como llaves los medios usados por el artista. Para el requerimiento 4, usaremos el mismo índice que ordena las obras por sus artistas. Así, para cada llave en el índice de artistas, se buscará la nacionalidad de ese artista y se agregarán todas las obras de la lista del id de ese artista a la nacionalidad correspondiente en un nuevo índice que tiene como llaves las nacionalidades. Para el requerimiento 5, vamos a hacer un índice que tenga como llaves los departamentos del museo y como valores todas las obras pertenecientes a esos departamentos. Por último, para el requerimiento 6 usaremos el mismo índice que tiene como llaves los artistas y para cada artista haremos un mapa auxiliar que tiene como llaves los diferentes medios.
- Así tendríamos que habría que implementar 4 índices en el catálogo del reto:
 - *Por id de los artistas con sus diccionarios como valores
 - *Por nombre del artista con una lista con sus obras como valor
 - *Por nacionalidades que tiene como valor la lista de obras
 - *Por departamentos del museo

- Además, estarían los índices para hacer el requerimiento 1 y 2 por medio de mapas aunque no sea más eficiente que usar listas: estos serían el índice por año de nacimiento de los autores y el índice por fecha de adquisición de las obras.
- Además, para los requerimientos 3 y 6 se crearán mapas fuera del catálogo que tienen como llaves los medios y como valores listas de obras, pero solo de un artista específico por mapa.
- Usar todos estos índices implementarían una solución con tiempos de respuesta muy bajos, pero que ocupa mucha memoria implementando todos los índices en el catálogo. Para una solución equilibrada entre gastos de espacio y tiempo, tendríamos que quitar algunos índices que nos dan una mejora en el tiempo.

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

- Ya que linear probing es más rápido que separate chaining, pero desperdicia más memoria, en este caso solo usaríamos separate chaining para los índices que no necesiten demasiadas búsquedas y que tengan muchas nuevas inserciones, por lo que podemos ahorrar memoria. En el resto de índices usaremos linear probing para buscar la solución más rápida.
- En este caso usaríamos separate chaining en:
 - *Índice por id de los artistas con sus diccionarios como valores
 - *Índice por nombre del artista con una lista con sus obras como valor
- Usaríamos Linear Probing en:
 - *Índice por nacionalidades que tiene como valor la lista de obras
 - *Índice por departamentos del museo
 - *Índices auxiliares por medio para las obras de un artista específico
- Si usáramos los índices por año de nacimiento del autor y por fecha de adquisición de la obra, tendríamos que iterar por todas las llaves en los requerimientos, por lo que usaríamos linear probing para hacerlo más r

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

- Los factores de carga para los índices (suponiendo que las colisiones sean uniformes) serían:
 - *Índice por nombre del artista con una lista con sus obras como valor: el factor de carga óptimo sería 4, de forma que hubiera 4 veces más artistas que el número de buckets en el mapa.
 - *Índice por nacionalidades que tiene como valor la lista de obras: el factor de carga sería el número de posiciones que hay que recorrer en los índices antes de encontrar dónde poner una nueva llave, sería el número de nacionalidades sobre la capacidad dada para el arreglo.
 - *Índice por departamentos del museo: el número de obras para un el número total de departamentos.
 - *Índices auxiliares por medio para las obras de un artista específico: sería el número de medios sobre la capacidad del mapa.
 - *Índice por año de nacimiento de autores: el factor de carga, tal como en los otros lineal probing, sería el número de años sobre la capacidad del mapa.
 - *Índice por fecha de adquisición de las obras: el número de fechas en las que hubo una adquisición sobre la capacidad del mapa.
 - *Índice por id de los artistas con sus diccionarios como valores: en este caso, el factor sería el número de artistas sobre el número de buckets del arreglo, sería óptimo que hubiese unos 4 id de artistas por bucket.

Tiempos de carga de datos large con Separate Chaining:

Factor de Carga	Tiempo de Carga promedio
4.00	5359.375ms
2.00	5318.125ms
8.00	5593.75ms

Tiempos de carga de datos large con Linear Probing:

Factor de Carga	Tiempo de Carga promedio
0.5	5262.5ms
0.20	5487.75ms
0.80	5500.0ms

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

Como se esperaba, el sondeo lineal tiene tiempo un poco más bajos que el encadenamiento separado, pues para hacer los índices de la carga de datos, se usan funciones que buscan sobre el mapa y la búsqueda es más rápida en el sondeo lineal porque no tiene que buscar linealmente en un bucket como en el encadenamiento.

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.

Como medios es un índice del que vamos a iterar sobre todas sus llaves, vamos a usar un mapa que resuelva las colisiones por sondeo lineal para priorizar la velocidad y que no tenga que buscar los elementos en buckets, además que el número de técnicas no es tan alto como para que gaste mucha memoria haciendo sondeo lineal. El factor de carga será de 0.5, ya vimos que para factores más altos gasta más tiempo buscando un objeto por los clusters más grandes, y para valores más bajos tiene que hacer mucho rehash, además de que deja más espacio sin usar. El número inicial de elementos sería el número de medios que hay en los datos, yo usé 30, mas este tamaño se arreglará con el rehash si hace falta.

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.

Como nacionalidades es un índice del que también vamos a iterar sobre todas sus llaves para encontrar la lista de tamaño mayor, vamos a usar un mapa que resuelva las colisiones por sondeo lineal para priorizar la velocidad y que no tenga que buscar los elementos en buckets, además que el número de nacionalidades no es tan alto como para que gaste mucha memoria haciendo sondeo lineal. El factor de carga será el mismo del índice por medios, pues serán mapas muy parecidos, con un número no muy alto de llaves que se asocia con una lista larga de obras. El

número inicial de elementos debería ser el número de nacionalidades que hay en los datos, en este caso diremos 45, mas este tamaño se arreglará con el rehash si hace falta.