

OBSERVACIONES DE LA PRACTICA

Valeria Caro Ramírez Cod 202111040

Sofia Velasquez Marin Cod 202113334

- a) Teniendo en cuenta cada uno de los requerimientos ¿Cuántos índices implementaría en el Reto? y ¿Por qué?
- b) Según los índices propuestos ¿en qué caso usaría Linear Probing o Separate Chaining en estos índices? y ¿Por qué?
- c) Dado el número de elementos de los archivos MoMA, ¿Cuál sería el factor de carga para estos índices según su mecanismo de colisión?
- d) ¿Qué diferencias en el tiempo de ejecución notan al ejecutar la cargar los datos al cambiar la configuración de Linear Probing a Separate Chaining?
- 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 numero inicial de elementos.
- f) ¿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.

Respuestas:

- a) Req 1: Hacer un índice por fecha de nacimiento (Llave: BeginDate, Valor: lista de artistas que nacieron en ese año).
Req 2: Hacer un índice por fecha de creación (Llave: Date, Valor: lista de obras que se crearon en ese año).
Req 3: Hacer 2 índices, uno por el nombre del artista (Llave: DisplayName, Valor: ConstituentID) y el otro por el id del artista y su valor seria otro índice por la técnica de las obras (Llave: ConstituentID, Valor: TAD Map -> (Llave: Medium, Valor: lista de obras del artista que se crearon con dicha técnica)).
Req 4: Hacer un índice por la Nacionalidad de los artistas de la obra (Llave: Nationality, Valor: lista de obras que tienen un artista con dicha nacionalidad).
Req 5: Hacer un índice por departamento (Llave: Department, Valor: lista de obras que pertenecen a dicho departamento).
Req 6: Se usaría los índices de los requerimientos 1 y 3.
- b) Debido al tamaño de los archivos 'large' Separate Chaining sería la mejor forma de manejar las colisiones y no usar tanta memoria.
- c) Linear Probing: el factor de carga seria 0.5 no es menor porque especialmente para el archivo de artworks se gastaría mucha memoria.
Separate Chaining: el factor de carga seria entre 4.0 y 8.0 para tener buckets muy grandes, pero tampoco muy pequeños para que ayude a reducir el consumo de memoria.

d)

Separate Chaining

Factor de carga	Tiempo de ejecución promedio [ms]
2.0	30234.375
4.0	31531.25
8.0	40062.5

Linear Probing

Factor de carga	Tiempo de ejecución promedio [ms]
0.2	34296.875
0.5	29640.625
0.8	34539.45

Tiempo promedio con Separate Chaining: 33942.71 ms

Tiempo promedio con Linear Probing: 32825,65 ms

La diferencia entre los tiempos de ejecución es de 1117,06 ms donde usar Linear Probing tiene un tiempo de ejecución menor que Separate Chaining, pero la diferencia en los tiempos es muy mínima.

- e) Se contaron cuantas técnicas diferentes había en el archivo large y se encontró que hay 21190. Utilizaríamos Separate Chaining con un factor de carga de 4.0 y el numero inicial de elementos seria 5309 (*Siguiente numero primo de $(21190 \div 4)$*) y así no consumir demasiada memoria si lo hiciéramos con Linear Probing con un factor de 0.5 el número inicial de elementos seria 42391 (*Siguiente numero primo de (21190×2)*) para evitar el rehash.
- f) Se contaron cuantas nacionalidades diferentes había en el archivo large y se encontró que hay 121. Utilizaríamos Linear Probing con un factor de carga de 0.5 y el numero inicial de elementos seria 251 (*Siguiente numero primo de (121×2)*) Así evitaríamos que se hiciera rehash.