

Laboratorio 6

Ronald Pardo | 202111309 | r.diazp@uniandes.edu.co | ISIS 1225 (4)

Paso 1

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

INDICE	MAPTYPE	LOADFACTOR	JUSTIFICACION DE ELECCION DE INDICE
catalog['artists map'] : TAD MAP(ConstituentID -> DisplayName, Gender, BeginDate, ArtworkNumber, MediumNumber, EndDate, Nationality, Mediums : TAD MAP(), TopMedium , TopMedium artworks)	Linear probing	0.5	Usare esta estructura porque tengo una TAD LIST con la que hago los recorridos y obtengo los nombres, adicionalmente, esto me servira para guardar los artistas de las obras.
catalog['artworks map'] : TAD MAP(ObjectID --> ConstituentID, DateAcquired, CreditLine, Title, Artists : TAD LIST(DisplayName), Classification, Medium, Dimensions, Date)	Linear probing	0.5	Usare esta estructura porque tengo una TAD LIST donde almaceno los medios y las obras de los artistas las guardo por medio de sus ObjectID's, asi es facil acceder a ellas en los requerimientos con el mapa.
catalog['artists with ids'] : TAD MAP(ConstituentID : DisplayName)	Separate Chaining	4	Usare estas estructuras combinadas porque por medio de ella podre hacer conversiones de datos de forma rapida sin acceder a mucha informacion no necesaria.
catalog['ids with artists'] : TAD MAP(DisplayName : ConstituentID)	Separate Chaining	4	

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

La razón por la cual decidí usar Linear Probing en todos los índices (excepto por `catalog['artists with ids']` y `catalog['ids with artists']`) es porque mi máquina --no tengo pareja-- tiene 8gb de ram y usualmente el tipo de colisiones por Separate Chaining (que denotare por las siglas SC) consumen aproximadamente más memoria y no resultan ser eficientes en sus búsquedas¹. Debido a que en las colisiones tipo SC hacen uso de Linked Lists para almacenar buckets lo que las hace dependientes de su función de Hash y los pointers que contiene cada elemento en el map agregan memoria, asimismo, el hecho de que se usen Linked Lists para los buckets implica que en ciertos casos la memoria se encuentre en una distancia no cercana entre los objetos, lo cual provoca que hacer una instrucción de búsqueda pueda tomar más tiempo en ciertos casos. Adicionalmente la razón por la que no use Linear probing en los índices mencionados inicialmente es que estos dos índices serán 2 mapas con solo 2 llaves y sus valores no consumirán tanta memoria, así que estoy dispuesto a sacrificar su memoria para obtener buenos tiempos con ellos.

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

El factor de carga para estos índices será 0.5, ya que con este valor se mantendrá una proporción en el uso de memoria y en la complejidad durante la búsqueda de elementos. Adicionalmente mi máquina cuenta con una cantidad de ram limitada, lo cual explica porque prefiero usar PROBING debido a su menor uso de memoria.

- D. ¿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?

Índice de nacionalidad:

Linear probing (loadfactor = 0.5) = 15.625 ms

Separate Chaining (loadfactor = 0.5) = 0.0 ms

Índice de medios:

Linear probing (loadfactor = 0.5) = 15.625 ms

Separate Chaining (loadfactor = 0.5) = 0.0 ms

Diferencias: Como se puede ver claramente, por medio del uso de Separate Chaining se obtuvo tiempos más bajos, lo cual evidencia porque Separate Chaining tiene complejidades bajas en comparación a Linear probing, a pesar de que ocupe más memoria.

¹ Esta información la recupere del siguiente documento proporcionado por la Universidad de Texas:
<https://www.cs.utexas.edu/~mitra/csSpring2017/cs313/lectures/hash.html>

- 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.

Usare un ADT Map con un mecanismo de collision de Linear Probing y con un factor de 0.5, adicionalmente, el numero de elementos iniciales lo obtuve por medio de pruebas con la libreria ADT para encontrar el numero de medios en el archivo large, y obtuve el numero 21251, por lo cual usare este numero a la hora de crear el mapa y asi evitar re-hash.

- 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.

Usare un ADT Map con un mecanismo de collision de Linear Probing y con un factor de 0.5, adicionalmente, el numero de elementos iniciales lo obtuve por medio de pruebas con la libreria ADT para encontrar el numero de medios en el archivo large, y obtuve el numero 64570, por lo cual usare este numero a la hora de crear el mapa y asi evitar re-hash.