

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

Un mapa es un tipo de estructura de datos de búsqueda rápida de claves que ofrece un medio flexible de indexación en sus elementos individuales. Los índices de los elementos de un mapa se denominan claves.

Requerimiento 1: 1 mapa

Un mapa, donde la llave sea el año de nacimiento y el valor sea una lista de diccionarios, donde cada uno de estos lleva la información requerida del artista que nace en dicho año (info: Artists.csv).

El número de índices depende de la cantidad de años que se encontraron en dicho rango pedido por el usuario (es decir el número de años donde se identificó que uno o mas artistas nació). Ese será la cantidad de índices.

Requerimiento 2: 1 mapa

Un mapa, donde la llave sea cierta fecha de adquisición y el valor sea una lista de diccionarios, donde cada uno de estos lleva la información requerida de la obra de arte que pertenece a dicho año de adquisición por parte del museo (info: Artworks.csv).

El número de índices depende de la cantidad de fechas de adquisición que se encontraron en dicho rango pedido por el usuario (es decir el número de años donde se identificó que una o mas obras fueron adquiridas). Ese será la cantidad de índices.

Requerimiento 3: 2 mapas

Un mapa, donde la llave sea el nombre del artista y el valor sea un segundo mapa que contenga diccionarios donde la llave, sea el tipo de técnica y el valor sea una lista de las obras que pertenezcan a esa técnica.

Ej: {llave: nombre_artista; valor: {llave: tipo_tecnica ; valor : [obras]}}

En este requerimiento como haremos uso de dos mapas, entonces para el primero de ellos el número de índices dependerá del número total de artistas identificados en el archivo csv, mientras que para el segundo mapa el número de índices será dependiendo de número total de técnicas identificadas. Esos serán el numero de índices correspondientes para cada mapa.

Requerimiento 4: 3 mapas

Mapa 1

Llave: ConstituentID

Valor: lista de obras de ese ID

El número de índices depende del número de obras del archivo csv. Ese será la cantidad de índices.

Mapa 2

Llave: ConstituentID

Valor: lista de autores de ese ID

El número de índices depende del número de autores en total del archivo csv. Ese será la cantidad de índices.

Mapa 3

Llave: Nacionalidad

Valor: lista de las obras dicha nacionalidad

El número de índices depende de la cantidad de nacionalidades encontradas en el archivo csv. Ese será la cantidad de índices.

Requerimiento 5:

Ej: {llave: departamento; valor: {llave: obras ; valor : [información de las obras: título, artista, dimensiones, etc...]} }

Esta estructura contiene un primer mapa donde el número índices depende de la número total de departamentos, mientras que en el segundo mapa su número de índices depende la número total de obras identificadas por dicho departamento.

El número de índices se determino teniendo en cuenta la cantidad de datos que se necesitaban para cada requerimiento y estos son variables para cada uno y para cada caso de prueba. Tomando en mente, que estamos manejando la estructura de mapas se organizó de una forma eficiente la totalidad de los datos como se observan en las explicaciones anteriores.

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

Teniendo en cuenta el punto anterior, para el requerimiento 1 y 2 usaríamos Linear Probing, ya que es una estructura de datos para mantener una colección de pares clave-valor y buscar el valor asociado con una clave dada.

En el requerimiento 3, 4 y 5 creemos que el mas acertado es separate chaining, ya que como van a haber diversos valores asociados a las misma llave (ej. Muchas obras en una técnica), entonces usamos chaining para que al haber una colisión queden todos estos valores en una misma lista asociada, de acuerdo al caso.

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

Si es linear probing su factor de carga es 0.5 y si es separate chaining su factor es 4.0.