
Laboratorio 6: Observaciones

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

Sería necesario implementar los siguientes índices para los artistas:

- 1- **ConstituentID**: permitirá acceder con facilidad a la información de un artista específico.
- 2- **BeginDate**: permitirá determinar con facilidad la información de interés de los artistas que nacieron dentro del rango (requerimiento 1).

Y para las obras:

- 1- **ObjectID**: permitirá acceder con facilidad a la información de una obra específica.
- 2- **Medium**: permitirá acceder fácilmente a las obras que fueron creadas usando una técnica y, por ende, usando también el índice **ObjectID**, acceder con facilidad a las obras de un artista que fueron creadas usando una técnica (requerimiento 3).
- 3- **Nationality**: permitirá acceder con facilidad a las obras cuyos autores son de cierta nacionalidad (requerimiento 4).
- 4- **DateAcquired**: permitirá determinar con facilidad la información de interés de las obra que fueron creadas dentro del rango (requerimiento 2).

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

Debido a que tienen una cantidad de llaves reducida, los siguientes índices se podrían implementar usando *Linear Probing* (ya que este sacrifica espacio para ganar tiempo):

- ➔ **BeginDate**.
- ➔ **Nationality**.
- ➔ **DateAcquired**.

Por otro lado, ebido a que tienen una cantidad de llaves muy grande (cercanas a la cantidad de artistas u obras totales del catálogo), los siguientes índices se podrían implementar usando *Separate Chaining* (ya que este sacrifica tiempo para ganar espacio):

- ➔ **ConstituentID**.
- ➔ **ObjectID**.
- ➔ **Medium**. (la mayoría de técnicas son únicas por obra).

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

Para los que se implementarán mediante *Separate Chaining*, se podría usar un factor de carga de 4 (el de defecto), ya que se conoce la cantidad de elementos que se deberán almacenar. Lo mismo sucede con los índices que se implementarán mediante *Linear Probing*, por lo que para estos se podría usar un factor de carga de 0,5 (el de defecto).

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*?

Para responder esta pregunta, se optó por realizar diversas pruebas para cada configuración; se midieron los tiempos de ejecución de la carga de datos para diferentes tamaños de los archivos y para una misma configuración. Los resultados se recopilaron en la siguiente tabla:

| Configuración | Tamaño Archivos | Tiempo de Ejecución (msg) |
|---------------|-----------------|---------------------------|
| Chaining, 4.0 | small | 62,500 |
| | 5pct | 359,375 |
| | 10pct | 687,500 |
| | 20pct | 1.328,125 |
| | 30pct | 2.843,750 |
| | 50pct | 3.453,125 |
| | 80pct | 5.187,500 |
| Probing, 0.5 | small | 62,500 |
| | 5pct | 406,250 |
| | 10pct | 703,125 |
| | 20pct | 1.406,250 |
| | 30pct | 2.140,625 |
| | 50pct | 3.390,625 |
| | 80pct | 5.250,000 |

Tabla 1: primera prueba.

Estos resultados permiten evidenciar que, para la mayoría de los tamaños de los archivos, la implementación de encadenamiento separado fue más eficiente que la implementación de *linear probing* (ambos usando el factor de carga por defecto). Aun así, la diferencia de tiempo no es muy grande o considerable; ambas implementaciones son bastante eficientes para cargar los datos.

Sin embargo, considerando los resultados se podría especular que, conforme aumenta la cantidad de datos que hay que cargar, podría ser un poco más eficiente la implementación de *linear probing*, ya que los tiempos de ejecución fueron menores con los archivos de tamaño *30pct* y *50pct*. Aun así, no se dispone de una cantidad de observaciones lo suficientemente grande como para poder afirmar esto con certeza.

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.

Para responder esta pregunta, se optó por realizar diversas pruebas para cada configuración; se midieron los tiempos de ejecución de la carga de datos para diferentes tamaños de los archivos y para una misma configuración. Los resultados se recopilaron en la siguiente tabla:

| Configuración | Tamaño Archivos | Tiempo de Ejecución (msg) |
|---------------|-----------------|---------------------------|
| Chaining, 2.0 | small | 62,500 |
| | 50pct | 3.406,250 |
| | 80pct | 5.406,250 |
| Chaining, 8.0 | small | 78.125,000 |
| | 50pct | 3.390,625 |
| | 80pct | 5.984,375 |
| Probing, 0.2 | small | 93,750 |
| | 50pct | 4.265,625 |
| | 80pct | 7.718,750 |
| Probing, 0.8 | small | 46,875 |
| | 50pct | 3.703,125 |
| | 80pct | 5.296,875 |

Tabla 2: segunda prueba.

Considerando estos resultados, para cargar los datos del archivo de tamaño *80pct* se deberá escoger la implementación de *Linear Probing* con un factor de carga de 8.0 y un número inicial de elementos de 146100/0.8.

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.

Considerando los resultados de la tabla 2, para cargar los datos del archivo de tamaño *80pct* se deberá escoger la implementación de *Linear Probing* con un factor de carga de 8.0 y un número inicial de elementos de 200/0.8.