

Sábado 9 de octubre, 2021

Integrantes:

- Andrés Castro - 201920805 - a.castrol@uniandes.edu.co
- Nicolas Merchan Cuestas - 202112109 - n.merchan@uniandes.edu.co

Observaciones – Laboratorio 6

Nota: Se utilizan los archivos -small para el presente laboratorio. Existen 768, 1948, 383 y 49 artistas, obras de arte, técnicas y nacionalidades en los archivos -small, respectivamente.

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

Dados los conjuntos de datos proporcionados, consideramos pertinente utilizar 1543, 3907, 768 y 101 índices para los artistas, las obras de arte, las técnicas y las nacionalidades, respectivamente. La decisión en cuestión se da en base a una priorización de mayor eficiencia en la búsqueda de datos que de uso de espacio. Ello se debe, a su vez, a que los equipos utilizados para crear y ejecutar el programa solicitado disponen de espacio en memoria principal suficiente para ejecutar el programa de tal forma. En ese orden de ideas, se profundizará sobre los mecanismos de utilizados para dicho fin en el siguiente apartado.

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

Consideramos que el uso del mecanismo *Linear Probing* es propicio para ordenar los elementos analizados. En ese orden de ideas, la localización de un elemento dado en un TAD lista puede realizarse con una complejidad constante, lo cual es ideal a la hora de comparar los elementos de `Artists-utf8-small.csv` y `Artworks-utf8-small.csv` en los requerimientos del Reto 2. Finalmente, se utiliza el número de índices igual al número primo más cercano al doble del tamaño de los datos en cuestión, de acuerdo a un factor de carga igual a 0.5.

3. 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 del mecanismo *Linear Probing* en el caso particular del reto 2 sería igual a 0.5. Ello se debe a que se busca cumplir con los requerimientos teóricos de la implementación del mecanismo de colisión seleccionado.

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

Se utilizó el mecanismo *Linear Probing* con un factor de carga de 0.5. Así mismo, se utilizaron 768 y 101 índices para los mapas referentes a las técnicas y nacionalidades, respectivamente. De ese modo, el programa tardó, en promedio, 1609.375 milisegundos en cargar todos los datos. Paralelamente, se utilizó el mecanismo *Separate Chaining* con un factor de carga de 4. Así mismo, se utilizaron 97 y 13 índices para los mapas referentes a las técnicas y nacionalidades, respectivamente. De ese modo, el programa tardó, en promedio, 1682.29 milisegundos en cargar todos los datos. En conclusión, el proceso de carga de datos es ligeramente más rápido cuando se hace uso del mecanismo *Linear Probing*.

Resultados Pruebas – Paso 4 (archivos -small)

1. Prueba *Separate Chaining* 1
 - Factor de carga: 2
 - Índices técnicas: 192
 - Índices nacionalidades: 29
 - Resultado: 1583.33 milisegundos
2. Prueba *Separate Chaining* 2
 - Factor de carga: 8
 - Índices técnicas: 53
 - Índices nacionalidades: 11
 - Resultado: 1713.54 milisegundos
3. Prueba *Linear Probing* 1
 - Factor de carga: 0.2
 - Índices técnicas: 1931
 - Índices nacionalidades: 251
 - Resultado: 1807.29 milisegundos
4. Prueba *Linear Probing* 2
 - Factor de carga: 0.8
 - Índices técnicas: 487
 - Índices nacionalidades: 67

- Resultado: 1630.2 milisegundos

5. ¿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.

En el caso del índice de técnicas elegiríamos una configuración de TAD lista en la cual se hiciese uso del mecanismo *Linear Probing*, un factor de carga de 0.5 y un número inicial de 768 elementos. La localización de un elemento dado en un TAD lista puede realizarse con una complejidad constante, lo cual es ideal a la hora de comparar los elementos de `Artists-utf8-small.csv` y `Artworks-utf8-small.csv` en los requerimientos del Reto 2. Así mismo, la ligera diferencia en tiempos de ejecución con el mecanismo *Separate Chaining* con un factor de carga de 2 y un número inicial de 192 elementos al momento de cargar los datos no compensa, a nuestro parecer, la complejidad constante de búsqueda por parte del mecanismo *Linear Probing*.

6. ¿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.

En el caso del índice de nacionalidades elegiríamos una configuración de TAD lista en la cual se hiciese uso del mecanismo *Linear Probing*, un factor de carga de 0.5 y un número inicial de 101 elementos. La localización de un elemento dado en un TAD lista puede realizarse con una complejidad constante, lo cual es ideal a la hora de comparar los elementos de `Artists-utf8-small.csv` y `Artworks-utf8-small.csv` en los requerimientos del Reto 2. Así mismo, la ligera diferencia en tiempos de ejecución con el mecanismo *Separate Chaining* con un factor de carga de 2 y un número inicial de 29 elementos al momento de cargar los datos no compensa, a nuestro parecer, la complejidad constante de búsqueda por parte del mecanismo *Linear Probing*.