

## Lab 6: Mecanismos de colisión

Camilo Garcia – 201728914

Daniel Vargas - 201822068

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

Implementaríamos 5 índices distintos en el reto, según el tipo de búsqueda que se deba hacer en cada requisito. Estos serían: Técnicas, Años, Fechas (DD/MM/AAAA), Nacionalidades, Departamentos. Haríamos esto para reducir el número de comparaciones y recorridos necesarios a la hora de buscar elemento por un parámetro dado, por ejemplo, para el de departamentos, en vez de recorrer y comparar todas las obras, puede buscar directamente en la lista de elementos dentro de la llave que se obtiene con el departamento en específico.

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

Para técnicas, años, nacionalidades, y departamentos, utilizaríamos probing, ya que son valores que tienden a repetirse, luego no habría una longitud tan larga de llaves, y el gasto de espacio adicional que implica linear probing no sería tan grave.

Para el mapa de fechas, donde los valores no tienden a repetirse y hay una muy amplia variedad, sería mejor utilizar separate chaining, ya que de lo contrario el gasto de espacio y las operaciones de rehash probablemente serían muy altos.

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 índices que utilicen probing, utilizaríamos un factor de carga de 0.5, para no gastar tanta memoria pero que las búsquedas igual sean rápidas.

Para los índices que utilicen chaining, utilizaríamos un factor de carga de 2.0, para acortar el número de elementos en cada bucket.

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?

Notamos que se demora un poco más en cargar los datos al utilizar separate chaining. Utilizando la muestra de 30%, los tiempos fueron:

CHAINING, FC(4.00) : 29031.25

PROBING, FC(0.5) : 28875.0

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.

Para el Map de técnicas, utilizaríamos el mecanismo de colisión probing, con un factor de carga de 0.5 y 5000 elementos iniciales (para la muestra de 30%). Esto lo decidimos teniendo en cuenta un balance entre rapidez de carga, y memoria utilizada. El tamaño lo decidimos al ver que existen varias obras con técnicas muy específicas, que es posible no se repitan, por lo que es mejor comenzar con un tamaño relativamente grande de elementos iniciales.

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.

Para el Map de técnicas, utilizaríamos el mecanismo de colisión probing, con un factor de carga de 0.5 y 100 elementos iniciales. Esto lo decidimos teniendo en cuenta un balance entre rapidez de carga, y memoria utilizada. El tamaño lo decidimos al ver que se repiten varios países, y a pesar de que se incremente el tamaño de artistas no va a incrementar drásticamente el número de países presentes.

Los resultados con la muestra de 30% fueron:

CHAINING, FC:2.00 : 28968.75

CHAINING, FC:8.00 : 28859.375

PROBING, FC: 0.2: 29203.125

PROBING, FC: 0.8: 28937.5