

# LABORATORIO No. 6: Mecanismos de Colisión

Valentina Perea Marquez – 202013095

Mateo Lopez Cespedes - 202014481

## Preguntas Preparatorias

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

Para los 6 requerimientos necesitaría 4 índices, uno para el requerimiento por fechas con los que solo se tomarían las llaves dentro del rango anual, uno por las técnicas se llamaría únicamente la llave con la técnica, otro con llave de la nacionalidad y por último el de los departamentos.

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

En todos los casos idealmente haría uso del Separate Chaining puesto que el mecanismo de colisión en buckets ahorra tiempo de búsqueda y de carga, añadiendo a una lista. Al contrario de un Linear Probing que en el colisiones debe hacer una búsqueda lineal hasta hallar un slot vacío y se pierde la relación de índice llave.

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

Considerando que el archivo de artistas contiene 15220 artistas y el de obras 138112 obras y que para los 4 índices se usará Separate Chaining, un factor de carga apropiado para aquellos índices que tomen sus datos del archivo de artistas sería de 1. Esto resultaría en una lista de 15220 posiciones en la que en cada posición hay un solo artista. Esto es considerando el peor de los casos en que el requerimiento use todos los artistas disponibles. Ahora, para los índices que usen el archivo de obras y considerando el peor caso en el que se usen todos los 138112 datos, un factor de carga apropiado sería de 8.

## Preguntas Factor de Carga

1.0. Tabla tiempo carga de datos de Mapas de Nacionalidad y Medios

CANTIDAD DE DATOS	LINEAR PROBING	SEPARATE CHAINING
.small	46.875 ms	31.25 ms
.20pct	7484.375 ms	4187.5 ms
.50pct	20906.25 ms	11484.375 ms
.80pct	28953.125 ms	15625.0 ms

.large	36234.375 ms	20046.875 ms
--------	--------------	--------------

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?

Linear Probing tarda mas en completar la carga de datos y a medida que crecen los datos va aumentando mas el tiempo que el Separate Chaining. Esto se debe a el mecanismo de colisión

CANTIDAD DE DATOS	LINEAR PROBING (2.00)	LINEAR PROBING (8.00)
.large	22421.875 ms	19890.625 ms
.small	46.875 ms	46.875 ms

CANTIDAD DE DATOS	SEPARATE CHAINING (0.2)	SEPARATE CHAINING (0.8)
.large	21375.0 ms	20609.375 ms
.small	46.875 ms	46.875 ms

d)¿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.

Usaria un Separate Chaining pues en primer lugar el archivo large contiene un gran numero de datos y si uso un linear probing se podría terminar con una complejidad de  $O(n)$  en una búsqueda lo cual no seria bueno para obtener resultados rápidos. Además, el mismo tamaño del archivo podría generar muchas colisiones lo cual tardaría mas tiempo en la carga de datos, por este motivo la solución de hash de un bucket en el Separate Chaining es la mejor para carga y búsqueda, además de que tiene un numero de slots más reducido.

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 esta, aunque por tiempo seria beneficioso un Separate Chaining, el numero es mucho menor que el de las obras y teniendo en cuenta que se agruparan por llaves de nacionalidad estas son mucho mas reducidas y por lo tanto es posible usar un mapa Probing, aunque es verdad que habrán colisiones el tamaño de las llaves es mucho más

reducido por lo que sería menor sobre todo si se implementa una buena función de Hash.