

Nicolás Palacios - 202021759

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

Lo que tengo planeado es hacer lo siguiente. La notación que utilizaré es {} para mapas y [] para listas.

```
Catalog = {Artists :  "id" : { Constituent id : {artista}}
              "Año" : { Año de Nacimiento : [{artistas}]
                    }

              {Artworks: "Año de adquisición: [{obras}]
                        "Nacionalidad" : [{obras}]
                        "Departamento" : [{obras}]
                        }
}
```

Ya que estoy solo, no voy a hacer el requerimiento 3, para el cual sería bueno tener un índice por nombre de artista. Por eso no lo incluí.

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

Sin haber hecho las pruebas, esperarí que Separate Chaining sea una mejor implementación para lo que estamos haciendo. Esto se debe a que con esta implementación nos ahorramos los grandes clusters de datos que se pueden generar con linear probing. Sin embargo, el hecho de que todos los datos estén juntos en linear probing puede hacer las consultas y las operaciones mucho más rápidos que con una lista enlazada. Probablemente lo mejor que se puede hacer es esperar a ver los resultados de las pruebas y con eso tomar una decisión.

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

Según lo visto en clase, se recomienda que para Separate Chaining el factor de carga sea 4. Esto es independiente de la cantidad de datos ya que habla de la relación que debe de tener el arreglo a la cantidad de datos. En este caso el arreglo debería de ser 4 veces menor a la cantidad de datos.

Por el lado de linear Probing, se recomienda que sea de 0.5, es decir, que el arreglo sea el doble de grande que la cantidad de datos.

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

110									
	A	B	C	D	E	F	G	H	
1									
2		Chaining (Promedio de 3 pruebas)			Probing (Promedio de 3 pruebas)				
3		2	4	8	0.2	0.5	0.8		
4	Carga	36531.25	37651	38166	38301	37923	37276		
5	medios	10.41	0	0	0	5.2	0		
6	nacionalidades	0	0	0	0	0	0		
7									
8									
9									
10	medio	Felt-tipped marker and pencil on paper							
11		Russian							
12									
13									
14									
15									

Esta es la tabla de resultados de las pruebas que llevé a cabo. La verdad, los resultados fueron muy similares para todas las implementaciones. Sin embargo, parece que Chaining en efecto es un poquito más rápido en general. Los resultados de las pruebas están en milisegundos y se utilizaron los archivos large. Puede que los requerimientos de prueba utilizados no sean los mejores ya que el de nacionalidades es $O(1)$ para ambas implementaciones, ya que solamente es ver el size del mapa, y el de medios son una cantidad muy pequeña de datos ya que hay muchos medios diferentes. Puede que sea necesario hacer las pruebas con los verdaderos requerimientos del reto para ver cual es mejor.

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.

Como fue dicho anteriormente, no hay una diferencia evidente entre las implementaciones. Sin embargo, para este requerimiento escogería el separate chaining con un factor de carga de 4. Parece que es el que logra hacerlo en el menor tiempo con el mejor tiempo de carga de datos. Para el número inicial de elementos, no sabemos cuantos medios hay, sin embargo sabemos que hay muchos. Por lo tanto, el número por defecto de 37 es un poco pequeño para 150 mil obras. Por lo tanto empezar aunque sea con 200 o más podría ser una buena idea.

e) ¿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.
Como fue dicho anteriormente, este requerimiento es $O(1)$ sin importar su implementación. Por lo tanto da igual. Sin embargo, separate chaining tuvo el menor tiempo de carga en general. Por lo tanto, estaría inclinado a usar ese. Con respecto a factor de carga, las pruebas arrojaron que 2 era más rápido que 4 por poco, sin embargo, la teoría recomienda 4. Finalmente, para el número inicial de datos, puede que los 37 por defecto estén bien. Ya que actualmente hay alrededor de 200 países en el mundo y no sabemos qué porcentaje estén representados en estos datos.