

LABORATORIO No. 6: Mecanismos de Colisión

Objetivos

Comprender la implementación del Tipo Abstracto de Datos Tabla de Símbolos no ordenadas (Map) y su uso para la solución de problemas.

Al finalizar este laboratorio el estudiante estará en capacidad de:

- Entender a usar las tablas de hash como estructura de datos y su impacto en los órdenes de crecimiento temporal.
- Familiarizarse con las estrategias de manejo de colisiones en tablas de hash
- Comparar los tiempos de respuesta de las tablas de hash cuando su factor de carga cambia

Fecha Límite de Entrega

Miércoles **13 de octubre** antes de la media noche (11:59 p.m.).

Preparación del Laboratorio

- Revisar el API del TAD Map ubicado en **DISClib\ADT\map.py**
- Revisar las estructuras de datos **chaininghashtable.py** y **probinghashtable.py** ubicado en **DISClib\DataStructures**
- Instalar las extensiones de VS Code para medir el consumo de recursos **Resource Monitor**, URL: [mutantdino.resourcemonitor](https://mutantdino.github.io/resource-monitor/)

Trabajo Propuesto

PASO 1: Responder preguntas preparatorias

En esta parte consideraremos el uso apropiado de los TAD Map respondiendo las siguientes preguntas:

- a) Teniendo en cuenta cada uno de los requerimientos ¿Cuántos índices implementaría en el Reto? y ¿Por qué?
- b) Según los índices propuestos ¿en qué caso usaría **Linear Probing** o **Separate Chaining** en estos índices? y ¿Por qué?
- c) Dado el número de elementos de los archivos MoMA, ¿Cuál sería el factor de carga para estos índices según su mecanismo de colisión?

PASO 2: Implementar modificaciones en el Reto No. 2

En esta sección evaluaremos el efecto del uso de diferentes mecanismos de colisión y distintos factores de carga en TAD Map en la implementación del reto.

Integre a su código del Reto No. 2 las siguientes instrucciones en la carga del catálogo del MoMA y recuerde que en el laboratorio anterior implementó un índice por medios o técnicas ("Medium") utilizando **TAD Map**.

Para ello siga las siguientes instrucciones.

- 1) Modifique la función de carga del catálogo de MoMA para incluir un nuevo índice por la nacionalidad de la obra ("Nationality") y que contenga una lista de todas sus obras asociadas.
- 2) Modifique el código del Reto No. 2 para incluir una opción que cuente el número total de obras de una nacionalidad utilizando el índice creado por la propiedad "Nationality"

Paso 3: Separate Chaining o Linear Probing

Configure el parámetro del método de manejo de colisiones para los dos índices implementados ("Medium" y "Nationality") como **Separate Chaining** (*maptypes= 'CHAINING'*) y con un factor de carga por defecto de 4.00.

Ahora, ejecute la carga de los datos MoMA y observe el tiempo de ejecución que requiere este proceso, anote los datos devueltos para usarlos posteriormente.

Termine el programa y modifique el código, cambie la configuración de ambos índices a **Linear Probing** (*maptypes= 'PROBING'*) con un factor de carga de 0.50.

Ejecute de nuevo la carga de datos en el catalogo y anote los datos del tiempo de ejecución de esta funcionalidad.

Al completar estas actividades, responda las siguientes preguntas:

- c) ¿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**?

Paso 4: Modificar el factor de carga

Ahora vuelva a configurar los índices en **Separate Chaining**, pero ahora cambie el factor de carga por defecto a 2.00. ejecute prueba y anote los resultados. Subsecuentemente ejecute una segunda prueba utilizando un factor de carga de 8.00.

Cambie las configuraciones de los índices a **Linear Probing**, pero ahora cambie el factor de carga por defecto a 0.20. ejecute prueba y anote los resultados. Subsecuentemente ejecute una nueva prueba utilizando un factor de carga de 0.80.

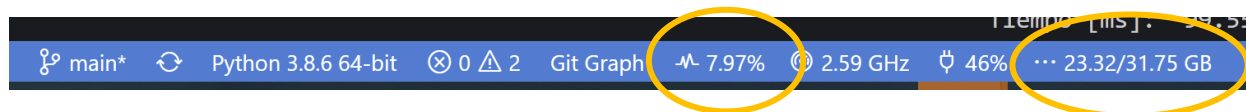
Al finalizar este procedimiento responda las siguientes preguntas:

- 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.
- 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.

Durante la ejecución del código recomendamos escoger el archivo de datos más apropiado a las capacidades de sus máquinas de trabajo. Y utilice **Resource Monitor** de **VS Code** para monitorear la memoria disponible de su computador y evitar bloqueos.

Recuerde que la carga de datos tiene un **máximo de duración de 15.0 minutos**, en caso de que esto suceda aborte el procesamiento y reduzca la cantidad de datos que su aplicación planea manejar.

Aborte la prueba en caso de que la memoria se acerque a su capacidad máxima o en caso de que su procesador este sobre cargado por mucho tiempo. Ambos datos se pueden ver en la barra inferior de información del IDE como se muestra a continuación.



PASO 4: Actualizar los repositorios

En el repositorio del Reto confirme los cambios con los comandos **Commit** y **Push** en la rama **main** local y de GitHub con el comentario *"Reto 2 – Avance en toma de datos"* antes de la fecha límite de entrega.

PASO 5: Revisar entregables de la practica

Finalmente, para realizar la entrega del laboratorio revise que sus entregables de la practica estén completos. Para ello, siga las siguientes indicaciones:

- 1) Acceso al profesor de laboratorio y los monitores de su sección a la organización del grupo.
- 2) **README** del repositorio con los datos completos de los integrantes del grupo (nombre completo, correo Uniandes y código de estudiante).
- 3) Enlace al repositorio GitHub **Reto2-G<<XX>>** con rama **Main** actualizada con el comentario *"Reto 2 – Avance en índices"* antes del límite de entrega.
- 4) Incluir en repositorio del reto en la carpeta **Docs** el documento **observaciones-lab6.pdf** con las respuestas a las preguntas de observación.
 - a) Teniendo en cuenta cada uno de los requerimientos ¿Cuántos índices implementaría en el Reto? y ¿Por qué?
 - b) Según los índices propuestos ¿en qué caso usaría **Linear Probing** o **Separate Chaining** en estos índices? y ¿Por qué?
 - c) Dado el número de elementos de los archivos MoMA, ¿Cuál sería el factor de carga para estos índices según su mecanismo de colisión?
 - 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**?
 - 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.
 - f) ¿Qué configuración de ADT Map escogería para el **índice de nacionalidades**?, especifique el mecanismo de colisión, el factor de carga y el número inicial de elementos.

PASO 9: Compartir resultados con los evaluadores

Envíe el **enlace (URL)** del repositorio por **BrightSpace** antes de la fecha límite de entrega.

Recuerden que cualquier documento solicitado durante la práctica debe incluirse dentro del repositorio GIT y solo se calificarán los entregables hasta el último ***COMMIT*** realizado previo a la media noche (11:59 PM) del 13 de octubre de 2021.