|  |  |  |
| --- | --- | --- |
|  |  |  |

FUNDAMENTOS DE ESTRUCTURAS DE DATOS

Y ANÁLISIS DE ALGORITMOS

**INFORME N°3**

**“Diccionarios, Tablas Hash vs Árboles de búsqueda”**

Junio 2023

**ERICH GERMÁN GRÜTTNER DÍAZ**

**Índice**

[1. Introducción 4](#_Toc133954985)

[2. Descripción de los algoritmos a ser comparados 5](#_Toc133954986)

[2.1 Ordenamiento 5](#_Toc133954987)

[2.1.1 Selection Sort **¡Error! Marcador no definido.**](#_Toc133954988)

[2.1.2 Merge Sort **¡Error! Marcador no definido.**](#_Toc133954989)

[2.1.3 Quick Sort **¡Error! Marcador no definido.**](#_Toc133954990)

[2.1.4 Sort Interno C++ **¡Error! Marcador no definido.**](#_Toc133954991)

[2.1.5 Tabla resumen algoritmos de ordenamiento a analizar en este informe **¡Error! Marcador no definido.**](#_Toc133954992)

[2.2 Multiplicación de matrices 6](#_Toc133954993)

[2.2.1 Algoritmo iterativo cúbico tradicional **¡Error! Marcador no definido.**](#_Toc133954994)

[2.2.2 Algoritmo iterativo cúbico optimizado para mantener la localidad de los datos **¡Error! Marcador no definido.**](#_Toc133954995)

[2.2.3 Algoritmo Strassen **¡Error! Marcador no definido.**](#_Toc133954996)

[3. Descripción de los datasets 9](#_Toc133954997)

[3.1 Datasets para algoritmos de ordenamiento **¡Error! Marcador no definido.**](#_Toc133954998)

[3.1.1 Inputs **¡Error! Marcador no definido.**](#_Toc133954999)

[3.1.2 Outputs **¡Error! Marcador no definido.**](#_Toc133955000)

[3.1.3 Archivos para graficar 13](#_Toc133955001)

[3.2 Datasets para multiplicación de matrices **¡Error! Marcador no definido.**](#_Toc133955002)

[3.1.1 Inputs **¡Error! Marcador no definido.**](#_Toc133955003)

[3.1.2 Outputs **¡Error! Marcador no definido.**](#_Toc133955004)

[3.1.3 Archivos para graficar **¡Error! Marcador no definido.**](#_Toc133955005)

[4. Resultados experimentales 15](#_Toc133955006)

[4.1 Medición rendimiento algoritmos de ordenamiento **¡Error! Marcador no definido.**](#_Toc133955007)

[4.1.1 Caso 1: Datos desordenados repetidos **¡Error! Marcador no definido.**](#_Toc133955008)

[4.1.2 Caso 2: Datos desordenados con valores únicos **¡Error! Marcador no definido.**](#_Toc133955009)

[4.1.3 Caso 3: Datos ordenados repetidos **¡Error! Marcador no definido.**](#_Toc133955010)

[4.1.4 Caso 4: Datos ordenados reversa (descendente) **¡Error! Marcador no definido.**](#_Toc133955011)

[4.1.5 Caso 5: Datos parcialmente ordenados **¡Error! Marcador no definido.**](#_Toc133955012)

[4.2 Medición rendimiento para multiplicación de matrices **¡Error! Marcador no definido.**](#_Toc133955013)

[4.2.1 Caso 1: Matrices cuadradas **¡Error! Marcador no definido.**](#_Toc133955014)

[4.2.2 Caso 2: Matrices cuadradas con tamaño potencia de 2 **¡Error! Marcador no definido.**](#_Toc133955015)

[4.2.3 Caso 3: Matrices rectangulares con matriz A (filas fijas, columnas variables) y matriz B (filas variables, columnas fijas) **¡Error! Marcador no definido.**](#_Toc133955016)

[4.2.4 Caso 4: Matrices rectangulares con matriz A (filas fijas, columnas variables) y matriz B (filas y columnas variables) **¡Error! Marcador no definido.**](#_Toc133955017)

[4.2.5 Caso 5: Matrices rectangulares con matriz A (filas y columnas variables) y matriz B (filas variables y columnas fijas) **¡Error! Marcador no definido.**](#_Toc133955018)

[5. Conclusiones 20](#_Toc133955019)

[5.1 Algoritmos de ordenamiento 20](#_Toc133955020)

[5.2 Multiplicación de matrices **¡Error! Marcador no definido.**](#_Toc133955021)

[6. Referencias 22](#_Toc133955022)

# 1. Introducción

Este informe tiene como objetivo presentar un análisis sobre las diferencias de funcionamiento, rendimiento e implementación de árboles de búsqueda versus el uso de tablas hash con hashing abierto y cerrado.

En la primera parte se describirán las estructuras a utilizar

A continuación será descrito el marco experimental, tanto los detalles de construcción como los diferentes datasets utilizados.

Además del análisis global entre estructuras, serán utilizadas diferentes claves para ver el comportamiento

Serán presentados gráficos y tablas de tiempo y ejecución

Se mostrará que el árbol tiene O(Log n)

Se mostrará que hashign cerrado tiene

Se mostrará que hashing abierto tiene

Finalmente, en el apartado de conclusiones se revisarán aquiellos puntos notables del análisis con indicaciones

**\*Sobre el dataset que es fijo y que es bueno para el árbol**

**\*Sobre hashing cuál es finalmente la diferencia entre abierto y cerrado, sus ventajas y desventajas**

# 2. Descripción de las estructuras de datos

## 2.1 Árbol binario de búsqueda

BST, o Binary Search Tree en inglés, se traduce como árbol de búsqueda binario en español. Es una estructura de datos jerárquica y ordenada que se utiliza para almacenar y organizar datos de manera eficiente, permitiendo una búsqueda rápida.

En un Binary Search Tree, cada nodo contiene un valor único y se divide en un subárbol izquierdo y uno derecho. La clave característica de un BST es que, para cada nodo, todos los valores en el subárbol izquierdo son menores que el valor del nodo, y todos los valores en el subárbol derecho son mayores. Esta propiedad permite realizar búsquedas de manera eficiente.

La estructura organizada de un BST se basa en el principio de "divide and conquer". Al realizar una búsqueda en un BST, se compara el valor buscado con el valor del nodo actual. Si es igual, se ha encontrado el elemento buscado. Si es menor, se sigue el subárbol izquierdo. Si es mayor, se sigue el subárbol derecho. Este proceso continúa hasta encontrar el valor deseado o llegar a una hoja, donde se determina que el valor no existe en el árbol.

La principal ventaja de los BST es que permiten búsquedas eficientes en tiempo logarítmico, lo que significa que el tiempo necesario para encontrar un elemento aumenta de manera gradual a medida que el tamaño del árbol crece. Además, los BST también permiten otras operaciones útiles como inserción, eliminación y recorrido ordenado de los elementos.

Sin embargo, es importante destacar que la eficiencia de un BST depende de su estructura y puede degradarse si el árbol está desequilibrado. En el peor de los casos, un árbol desequilibrado puede tener un rendimiento similar a una lista enlazada, lo que anula las ventajas de la búsqueda rápida. Por lo tanto, es necesario mantener el equilibrio del árbol para asegurar un rendimiento óptimo, para lo cual existen técnicas de balanceo, como el árbol AVL y el árbol rojo-negro.

En resumen, un Binary Search Tree (árbol de búsqueda binario) es una estructura de datos ordenada que permite una búsqueda eficiente en tiempo logarítmico. Es útil cuando se necesita realizar búsquedas frecuentes en una colección de elementos y se mantiene el equilibrio del árbol para asegurar un rendimiento óptimo.

https://es.wikipedia.org/wiki/Árbol\_binario\_de\_búsqueda

https://es.wikipedia.org/wiki/Árbol\_AVL

## 2.2 Tabla Hash con hashing abierto

Una tabla hash con hashing abierto es una estructura de datos que utiliza una función hash para almacenar y recuperar elementos de manera eficiente. En lugar de resolver las colisiones mediante el uso de listas enlazadas (hashing cerrado), en el hashing abierto, se emplea una estrategia que involucra explorar otras ubicaciones en la tabla hash cuando se produce una colisión.

La tabla hash en el hashing abierto consta de una serie de "ranuras" o "buckets", cada una de las cuales puede contener un elemento o estar vacía. Cada elemento se inserta en una ranura de acuerdo con el valor que produce la función hash aplicada a su clave. Si una ranura está ocupada, se utiliza una estrategia de exploración para buscar la siguiente ranura disponible.

Existen diferentes técnicas de exploración utilizadas en el hashing abierto, algunas de las cuales son:

1. Exploración lineal: Si se produce una colisión en una ranura, se examinan las ranuras consecutivas hasta encontrar una ranura vacía. La exploración se realiza de manera lineal, moviéndose secuencialmente a través de las ranuras.

2. Exploración cuadrática: En lugar de moverse linealmente, se utiliza una secuencia cuadrática para determinar las ranuras a explorar después de una colisión. Esto ayuda a evitar agrupamientos y distribuye los elementos colisionantes de manera más uniforme.

3. Exploración por doble hashing: En esta técnica, se utiliza una segunda función hash para determinar los desplazamientos entre las ranuras a explorar. La segunda función hash se aplica cuando ocurre una colisión y se calcula un nuevo valor de salto para buscar una nueva ranura.

Cada técnica de exploración tiene sus ventajas y desventajas, y su elección depende del contexto y los requisitos específicos del problema.

Es importante tener en cuenta que en el hashing abierto, la tabla hash debe tener suficiente capacidad para evitar un alto grado de colisiones, lo que podría afectar negativamente el rendimiento. Además, si la tabla se llena por completo, puede ser necesario redimensionarla para evitar la degradación del rendimiento.

En resumen, una tabla hash con hashing abierto es una estructura de datos que utiliza una función hash y estrategias de exploración para manejar colisiones. Permite almacenar y recuperar elementos de manera eficiente, explorando ranuras alternativas cuando se produce una colisión. La elección de la estrategia de exploración y el manejo adecuado de la capacidad de la tabla son aspectos importantes para garantizar un rendimiento óptimo.

Diagrama hashing abierto

## 2.3 Tabla Hash con hashing cerrado

Una tabla hash con hashing cerrado y uso de listas enlazadas combina dos técnicas para manejar colisiones en una tabla hash. En lugar de simplemente reemplazar el valor existente cuando se produce una colisión, se utiliza una lista enlazada para almacenar varios elementos con la misma posición hash.

El proceso de inserción en una tabla hash con hashing cerrado y listas enlazadas es el siguiente:

1. Se calcula el valor hash de la clave utilizando una función de hash. Esta función mapea la clave a un número entero, que se utilizará como índice en la tabla hash.

2. Se verifica si la posición calculada está ocupada. Si está vacía, se inserta el elemento en esa posición.

3. Si la posición está ocupada, se agrega el nuevo elemento al final de la lista enlazada asociada a esa posición.

En el caso de búsqueda en una tabla hash con hashing cerrado y listas enlazadas:

1. Se calcula el valor hash de la clave utilizando la misma función de hash.

2. Se accede a la posición correspondiente en la tabla hash.

3. Se recorre la lista enlazada asociada a esa posición y se busca la clave deseada. Si se encuentra, se devuelve el valor correspondiente. Si no se encuentra, se considera que la clave no está presente en la tabla.

La ventaja de utilizar listas enlazadas en caso de colisión es que se puede almacenar un número ilimitado de elementos con la misma posición hash. Sin embargo, el acceso a los elementos puede ser un poco más lento debido a la necesidad de recorrer la lista enlazada para buscar un elemento en particular.

En resumen, una tabla hash con hashing cerrado y uso de listas enlazadas es una estructura de datos que combina el hashing cerrado para asignar claves a posiciones en una tabla y el uso de listas enlazadas para almacenar múltiples elementos con la misma posición hash. Esto permite manejar colisiones y almacenar varios elementos en una misma posición de la tabla.

\*Diagrama hashing Cerrado

# 3. Código y documentación

El código de la aplicación construida se encuentra en el repositorio: <https://github.com/egruttner/FEDA-Informe3>

**3.1 Implementación de árbol binario**

Se utilizó la librería “AVL-Tree”, desarrollada por Kadir Emre Oto en 2018, <https://github.com/KadirEmreOto/AVL-Tree>

Según su autor, el AVL Tree que implementa esta librería corresponde a un árbol binario auto-balanceado, que garantiza un complejidad de tiempo O(Log N) para la inserción, borrado y búsqueda. Para mostrar el árbol y limpiarlo, en cambio, se garantiza una complejidad de tiempo de O(n).

Permite almacenar cualquier tipo de dato que permita la operación “mayor que (>)”. En particular para este informe se utilizarán long y string.

**3.2 Implementación de tabla hash con hashing abierto**

Se utilizó un código base generado por ChatGPT y luego modificado y adaptado a las necesidades de la tarea.

Fueron creadas clases separadas para manejar User\_Id y User\_Name, con sus tipos de datos propios (long y string, respectivamente).

Para el caso de string, la función transforma cada elemento del dato de entrada a int multiplicando por 31 (número primo). Con el resultado se realiza la división entera contra el tamaño de la tabla. Lo que es considerado un Hash de división.

Se utilizó una tabla de tamaño 10000.

**3.3 Implementación de tabla hash con hashing cerrado**

Se utilizó un código base generado por ChatGPT y luego modificado y adaptado a las necesidades de la tarea.

Fueron creadas clases separadas para manejar User\_Id y User\_Name, con sus tipos de datos propios (long y string, respectivamente).

Para el caso de string, toma cada elemento del dato, lo convierte a int y lo suma. Con el resultado se realiza la división entera contra el tamaño de la tabla. Lo que es considerado un Hash de división.

Para el manejo de colisiones se utilizó **“Sondeo Lineal”.**

Se utilizó una tabla de tamaño 30000.

**Wikipedia sondeo**

https://es.wikipedia.org/wiki/Tabla\_hash

**3.4 Medición de tiempos de ejecución**

Se utilizó la librería “Chrono” de STL, para medir la ejecución del código.

Cada tarea ejecutada fue medida en base a un promedio de ejecución de 10 intentos.

Cada ejecución se almacena en un archivo CSV con los tiempos para cada tamaño de dataset.

**3.5 Construcción del código**

Se realizó en base a **make**, lo que permite realizar ejecuciones programáticas en base a parámetros de entrada del programa.

Adicionalmente, se incorporó un archivo de tipo Bash para ejecutar otras tareas como limpieza y carga del programa.

**3.6 Generación de gráficos**

Finalmente, con los datos obtenidos en formato CSV, se utilizó un script en lenguaje Python, que utiliza la librería **MatPlotLib** y que permite generar gráficos customizados de acuerdo a las pruebas requeridas por el proyecto.

# 4. Datasets y diseño experimental

**4.1 Archivo base**

El proyecto se basa en procesar datos sobre el dominio de usuarios de Twitter. Para ello se dispuso del archivo **“universities\_followers\_2022.csv.zip”** el cual consiste en 29.245 usuarios de Twitter que son seguidores de la cuenta oficial de alguna universidad chilena en la red social.

**4.2 Archivos de input - Inserción**

Para realizar pruebas parciales, se crearon copias del archivo base de distinto tamaño y peso:

|  |  |  |
| --- | --- | --- |
| Archivo | Cantidad de registros | Peso |
| Input1.csv | 1000 | 85KB |
| Input2.csv | 5000 | 418KB |
| Input3.csv | 10000 | 816KB |
| Input4.csv | 15000 | 1,2MB |
| Input5.csv | 20000 | 1,7MB |

|  |
| --- |
| Inputx.csv |
| university;user\_id;user\_name;number\_tweets;friends\_count;followers\_count;created\_at |
| Dato 1 |
| … |
| Dato n |

**4.3 Archivos para realizar búsquedas**

|  |  |  |
| --- | --- | --- |
| Archivo | Cantidad de registros | Peso |
| Input1.csv | 100 | 8KB |
| Input2.csv | 200 | 17KB |
| Input3.csv | 300 | 25KB |
| Input4.csv | 400 | 34KB |
| Input5.csv | 500 | 42KB |

|  |
| --- |
| Inputx.csv |
| university;user\_id;user\_name;number\_tweets;friends\_count;followers\_count;created\_at |
| Dato 1 |
| … |
| Dato n |

Para el caso de datos no “encontrables” se aplicó el siguiente algoritomo

* Si se procesa User\_Id, se toma el dato desde el registro de datos “encontrables” y se le aumenta un número de su valor
* Si se procesa User\_Name, se toma el dato desde el registro datos “encontrables” y se le concantena una letra “X” para que no aparezca en el resultado

### 4.4 Archivos para graficar

Para facilitar la generación de gráficos para la observación del rendimiento de los algoritmos, se crearon archivos del tipo CSV con el siguiente formato:

|  |
| --- |
| nombre\_algoritmo\_ordenamiento\_results.csv |
| **Tamaño muestra, Tiempo[ms]** |
| **valor 1, tiempo 1** |
| **…** |
| **valor n, tiempo n** |

Estos archivos se pueden acceder en el repositorio en [[11](#once)]

4.4 Diseño experimental

Para la realización de las pruebas se utilizó un equipo MacbookPro con procesador M1 y 8Gb de memoria. El chip M1 tiene 8 núcleos (4 de alta eficiencia a 3.2 GHz + 4 de alto rendimiento a 2.0 GHz) y una velocidad de transferencia de 50Gb por segundo.

Forma de realizar las mediciones (ordenamiento y multiplicación de matrices)

* Generación de datasets
* Utilización de script bash para la llamada de cada dataset

Ejemplo, para la ejecución de script de multiplicación de matrices:

|  |
| --- |
| num\_datasets=5  for (( i=1; i <= $num\_datasets; ++i ))  do  make num\_dataset=$i  python3 plot.py csv/csv$i/standard\_results.csv  python3 plot.py csv/csv$i/transpose\_results.csv  python3 plot.py csv/csv$i/strassen\_results.csv  python3 plot\_dos.py csv/csv$i/standard\_results.csv csv/csv$i/transpose\_results.csv  python3 plot\_todos.py csv/csv$i/standard\_results.csv csv/csv$i/transpose\_results.csv csv/csv$i/strassen\_results.csv  done |

Este script ejecuta el llamado a make y luego muestra en pantalla los gráficos generados a partir de los archivos CSV producidos.

Para calcular el tiempo de ejecución, se utilizó la librería Chrono de C++:

|  |
| --- |
| long long execution\_time\_ms(Func function, const vector<int> &A, string alg) {  auto start\_time = std::chrono::high\_resolution\_clock::now();  function(A, alg);  auto end\_time = std::chrono::high\_resolution\_clock::now();  return std::chrono::duration\_cast<std::chrono::microseconds>(end\_time - start\_time).count();  } |

Donde “alg” corresponde al algoritmo a medir (ej: Quick sort)

Para graficar se usó la librería matplotlib de Python. Los gráficos para algoritmos de ordenamiento se encuentran en [[15](#quince)] y los gráficos asociados a la multiplicación de matrices se encuentran en [[16](#dieciseis)].

# 5. Resultados experimentales

**5.1 Mediciones de rendimiento usando Árbol de búsqueda**

Se realizaron ejecuciones separadas en base al índice utilizado. A nivel de código implicó instanciar dos tipos de árboles, uno que gestiona datos de tipo **long (User\_Id)**, y el otro que gestiona datos de tipo **string (User\_Name)**.

**5.1.1 Inserción usando User\_Id**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 1000 | 1376.8 | | 5000 | 5042 | | 10000 | 9140.7 | | 15000 | 14282.2 | | 20000 | 20312.2 | |  |

**5.1.2 Inserción usando User\_Name**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 1000 | 916.7 | | 5000 | 4960.9 | | 10000 | 10612.1 | | 15000 | 17048.8 | | 20000 | 23350.8 | |  |

**5.1.3 Búsqueda de datos cargados, usando User\_Id**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 100 | 253.2 | | 200 | 443.2 | | 300 | 673.6 | | 400 | 840.2 | | 500 | 1077 | |  |

**5.1.4 Búsqueda de datos cargados, usando User\_Name**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 100 | 916.7 | | 200 | 4960.9 | | 300 | 10612.1 | | 400 | 17048.8 | | 500 | 23350.8 | |  |

**5.1.5 Búsqueda de datos no cargados, usando User\_Id**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 100 | 916.7 | | 200 | 4960.9 | | 300 | 10612.1 | | 400 | 17048.8 | | 500 | 23350.8 | |  |

**5.1.6 Búsqueda de datos no cargados, usando User\_Name**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 100 | 916.7 | | 200 | 4960.9 | | 300 | 10612.1 | | 400 | 17048.8 | | 500 | 23350.8 | |  |

**5.2 Mediciones de rendimiento usando Tabla Hash con hashing abierto**

Se realizaron ejecuciones separadas en base al índice utilizado. A nivel de código implicó instanciar dos tipos de tablas, uno que gestiona datos de tipo **long (User\_Id)**, y el otro que gestiona datos de tipo **string (User\_Name)**.

**5.2.1 Usando User\_Id**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 1000 | 916.7 | | 5000 | 4960.9 | | 10000 | 10612.1 | | 15000 | 17048.8 | | 20000 | 23350.8 | |  |

**5.2.2 Usando User\_Name**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 1000 | 916.7 | | 5000 | 4960.9 | | 10000 | 10612.1 | | 15000 | 17048.8 | | 20000 | 23350.8 | |  |

**5.2.3 Búsqueda de datos cargados, usando User\_Id**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 100 | 916.7 | | 200 | 4960.9 | | 300 | 10612.1 | | 400 | 17048.8 | | 500 | 23350.8 | |  |

**5.2.4 Búsqueda de datos cargados, usando User\_Name**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 100 | 916.7 | | 200 | 4960.9 | | 300 | 10612.1 | | 400 | 17048.8 | | 500 | 23350.8 | |  |

**5.2.5 Búsqueda de datos no cargados, usando User\_Id**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 100 | 916.7 | | 200 | 4960.9 | | 300 | 10612.1 | | 400 | 17048.8 | | 500 | 23350.8 | |  |

**5.2.6 Búsqueda de datos no cargados, usando User\_Name**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 100 | 916.7 | | 200 | 4960.9 | | 300 | 10612.1 | | 400 | 17048.8 | | 500 | 23350.8 | |  |

**5.3 Medición de rendimiento de inserción en Tabla Hash con Hashing cerrado**

**5.3.1 Usando User\_Id**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 1000 | 916.7 | | 5000 | 4960.9 | | 10000 | 10612.1 | | 15000 | 17048.8 | | 20000 | 23350.8 | |  |

**5.3.2 Usando User\_Name**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 1000 | 916.7 | | 5000 | 4960.9 | | 10000 | 10612.1 | | 15000 | 17048.8 | | 20000 | 23350.8 | |  |

**5.3.3 Búsqueda de datos cargados, usando User\_Id**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 100 | 916.7 | | 200 | 4960.9 | | 300 | 10612.1 | | 400 | 17048.8 | | 500 | 23350.8 | |  |

**5.3.4 Búsqueda de datos cargados, usando User\_Name**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 100 | 916.7 | | 200 | 4960.9 | | 300 | 10612.1 | | 400 | 17048.8 | | 500 | 23350.8 | |  |

**5.3.5 Búsqueda de datos no cargados, usando User\_Id**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 100 | 916.7 | | 200 | 4960.9 | | 300 | 10612.1 | | 400 | 17048.8 | | 500 | 23350.8 | |  |

**5.3.6 Búsqueda de datos no cargados, usando User\_Name**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Tamaño muestra | Tiempo [ms] | | 100 | 916.7 | | 200 | 4960.9 | | 300 | 10612.1 | | 400 | 17048.8 | | 500 | 23350.8 | |  |

* **EXTRA**

**Cantidad de buckets**

**Cambio de función de Hashing**

# 6. Conclusiones

## 6.1 Árbol de búsqueda – AVL Tree­

El uso de un árbol balanceado y de la librería AVLTree, que garantizaba tiempo de complejidad O(Log n), hacía esperar de dsfdsfsdfsdf

Si bien esta expectativa se cumple a nivel individual, al compararse con el uso de tabla Hash (con hashing abierto y cerrado), pareciera dar la impresión de que solamente eso se cumple a niveles pequeños de datos

Gráfico

Pero si se hace un acercamiento al gráfico al comienzo de su ejecución, con menor cantidad de gráficos se puede observar la curva carcaterítiscasd

Gráfico con Zoom

Sobre User\_Id y User\_Name

Por la construcción de la librería y el rendimiento observado en los gráficos, se ve que el algoritmos funciona mejor cuando se trata de User\_Name, dato de tipo String

Este comportamiento se puede atribuir a las siguientes causas:

* Los largos de datos del dataset de User\_Name son menores que los largos de datos de User\_Id
* dsfdsfdsf

Comparación entre inserción y búsqueda – General

La inserción tiene una complejididad de tiempo de…

Y respecto a la búsqueda, se puede observar que tiene una complejidad de efefef

## 6.2 Tabla Hash con hashing cerrado

**Se espera una**

## 6.3 Tabla Hash con hashing abierto

## 6.4 Uso de claves User\_Id vs User\_Name en Árbol

## 6.5 Uso de claves User\_Id vs User\_Name en Tabla Hash con hashing abierto

## 6.6 Uso de claves User\_Id vs User\_Name en Tabla Hash con hashing cerrado

## 6.7 Comparación de inserción de datos usando clave User\_Id en Árbol, Tabla Hash con hashing abierto y Tabla Hash con hashing cerrado

## 6.8 Comparación de inserción de datos usando clave User\_Name en Árbol, Tabla Hash con hashing abierto y Tabla Hash con hashing cerrado

## 6.9 Comparación de búsqueda de datos usando clave User\_Id en Árbol, Tabla Hash con hashing abierto y Tabla Hash con hashing cerrado

## 6.10 Comparación de búsqueda de datos usando clave User\_Name en Árbol, Tabla Hash con hashing abierto y Tabla Hash con hashing cerrado

Extra

Comparación más o menos buckets

Pruebas con otras funciones de Hashing

# 7. Referencias

[1] Wikipedia, Selection sort. [En línea]. Disponible: https://en.wikipedia.org/wiki/Selection\_sort

[2] Wikipedia, Merge sort. [En línea]. Disponible: https://en.wikipedia.org/wiki/Merge\_sort

[3] Wikipedia, Quick sort. [En línea]. Disponible: https://es.wikipedia.org/wiki/Quicksort

[4] Wikipedia, Sort STL. [En línea]. Disponible: https://en.wikipedia.org/wiki/Sort\_(C%2B%2B)

[5] Medium, Algorithm Analysis & Time Complexity Simplified. [En línea]. Disponible: https://randerson112358.medium.com/algorithm-analysis-time-complexity-simplified-cd39a81fec71

[6] Wikipedia, Multiplicación de matrices. [En línea]. Disponible: https://es.wikipedia.org/wiki/Multiplicación\_de\_matrices

[7] De User:Bilou - Trabajo propio, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=1538693

[8] Github, Explicación del algoritmo de Strassen. [En línea]. Disponible: https://alu0100881677.github.io/DAA\_L2\_1\_Strassen/Strassen.html

[9] Github, Herramienta para la generación de datasets para algoritmos de ordenamiento. [En línea]. Disponible: https://github.com/egruttner/FEDA-informe1/tree/main/ordenamiento/code/datasets

[10] Github, Outputs de ejecución de algoritmos de ordenamiento. [En línea]. Disponible: https://github.com/egruttner/FEDA-informe1/tree/main/ordenamiento/code/output

[11] Github, Archivos para graficar. [En línea]. Disponible: https://github.com/egruttner/FEDA-informe1/tree/main/ordenamiento/code/csv

[12] Github, Herramienta para la generación de datasets para multiplicación de matrices. [En línea]. Disponible: https://github.com/egruttner/FEDA-informe1/tree/main/matrices/code/datasets

[13] Github, Outputs de ejecución de algoritmos de ordenamiento. [En línea]. Disponible: https://github.com/egruttner/FEDA-informe1/tree/main/matrices/code/output

[14] Github, Archivos para graficar. [En línea]. Disponible: https://github.com/egruttner/FEDA-informe1/tree/main/matrices/code/csv

[15] Github, Gráficos para algoritmos de ordenamiento. [En línea]. Disponible: https://github.com/egruttner/FEDA-informe1/tree/main/ordenamiento/gr%C3%A1ficos

[16] Github, Gráficos para multiplicación de matrices. [En línea]. Disponible: https://github.com/egruttner/FEDA-informe1/tree/main/matrices/gr%C3%A1ficos

[17] Github, Código fuente herramienta análisis algoritmos de ordenamiento. [En línea]. Disponible: https://github.com/egruttner/FEDA-informe1/tree/main/ordenamiento/code

[18] Github, Código fuente herramienta análisis algoritmos de ordenamiento. [En línea]. Disponible: https://github.com/egruttner/FEDA-informe1/tree/main/matrices/code