**Proyecto 1:**

**Organización de Archivos**

1. **Logro del estudiante**

Entender y aplicar los algoritmos de almacenamiento de archivos físicos y [acceso concurrente].

1. **Enunciado**

En grupos máximo de cuatro integrantes, elegir un dominio de datos conformado por al menos dos archivos planos con datos reales (https://www.kaggle.com/datasets). Luego sobre ellos implementar dos técnicas de organización de archivos con sus principales operaciones: inserción, eliminación y búsqueda.

1. **Requerimientos generales**
   1. Implementar dos de las siguientes técnicas de organización de archivos en memoria secundaria.
      1. Sequential File
      2. AVL File.
      3. ISAM-Sparse Index.
      4. Extendible Hashing.
      5. B+ Tree.
   2. Escoger uno entre Sequential File, AVL File y el ISAM, y otro entre el Extendible Hash y el B+Tree.
   3. Operaciones que se deben implementar:
      1. La búsqueda específica puede retornar mas de un elemento que coincide con la key

vector<Registro> search(T key)

* + 1. La búsqueda por rango retorna todos los registros que calzan entre dos llaves de búsqueda

vector<Registro> rangeSearch(T begin-key, T end-key)

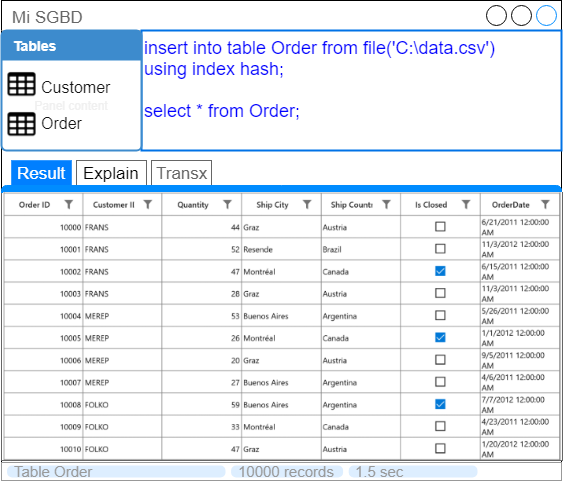
* + 1. Agregar un registro al archivo respetando la técnica de organización

bool add(Registro registro)

* + 1. Proponer un algoritmo de eliminación para cada técnica

bool remove(T key)

* 1. Implementar un parser de código SQL a sentencias ejecutables.
  2. Mostrar los resultados de forma amigable a usuario. Ver la siguiente GUI de referencia.



1. **Consideraciones de la implementación**
   1. ***En el Sequential Index La función add(Registro registro)*** *d*ebe utilizar un espacio auxiliar para guardar los nuevos registros. Cuando el espacio auxiliar llegue a K registros, aplique un algoritmo de reconstrucción del archivo de datos manteniendo el orden físico de acuerdo a la llave seleccionada. Asegúrese de mantener los punteros actualizados.
   2. **En el ISAM** partir construyendo el índice estático hasta de 3 niveles a partir de un conjunto de datos. Para nuevas inserciones se generan los overflow pages (encadenamiento de páginas).
   3. **En el ISAM** de debe definir el factor de bloque tanto en las páginas de datos como en las páginas del índice.
   4. La función **rangeSearch** debe usar el índice para buscar el begin-key y luego recorrer los registros de acuerdo la estructura del índice. Esta función no es soportada por las técnicas de hashing.
   5. Use adecuadamente los conceptos de programación orientado a objetos y programación genérica para que el programa soporte cualquier dominio de datos.
   6. Para la interfaz gráfica pueden usar QT, Visual Studio, wxWidgets u otra librería gráfica.
2. **Informe del proyecto**

* Archivo en Markdown / Wiki.
* Ortografía y consistencia en los párrafos.
* Al final del informe poner el **video explicando** el funcionamiento del programa, casos de uso, aspectos importantes de la implementación. El video no debe exceder los 10 minutos y deben participar todos los integrantes del grupo.
* Aspectos de evaluación e informe:

|  |  |
| --- | --- |
| **Item** | **Descripción** |
| * Introducción   (2 pts) | * Objetivo del proyecto. * Descripción del dominio de datos a usar. * Resultados que se esperan obtener. |
| * Técnicas Utilizadas.   (9 pts) | * Describa brevemente las técnicas de indexación de archivos que ha elegido. Explique procedimentalmente el proceso de inserción, eliminación y búsqueda. * Debe notarse con claridad el manejo de la memoria secundaria. * Se debe realizar un análisis comparativo teórico de las técnicas implementadas en base a los **accesos a memoria secundaria** tanto para las operaciones de inserción, búsqueda y eliminación. * Explicar claramente como se realizó el parser del SQL. |
| * Resultados Experimentales   (5 pts) | * Cuadro y/o gráfico comparativo de desempeño de las técnicas de indexación de archivos sobre el dominio de datos. **Tanto para inserción como para búsqueda**. * Considerar dos métricas: total de accesos a disco duro (read & write) y tiempo de ejecución en milisegundos. * Discusión y análisis de los resultados experimentales. |
| * Pruebas de uso y presentación   (4 pts) | * Presentar las pruebas de uso de la aplicación en interfaz gráfica amiga e intuitiva. * Recuerde mostrar la funcionalidad del aplicativo en el video. |

1. **Entregable**

* Los alumnos formarán grupos de máximo cuatro integrantes.
* El código fuente del proyecto será alojado en GitHub Classroom.
* En el Canvas subir solo el **enlace público** del proyecto.
* La fecha límite de entrega es el --------.