

TAREA PROGRAMADA 1

OBJETIVOS:

1. Implementar el Modelo Diccionario, utilizando 4 estructuras de datos genéricas: Lista Ordenada, Árbol de Búsqueda Binaria, Trie y Tabla Hash.
2. Comparar y evaluar las 4 estructuras de datos, utilizando criterios de eficiencia en términos del uso de espacio y del tiempo de ejecución.
3. Comparar y evaluar, utilizando criterios de eficiencia en términos del uso de espacio y del tiempo de ejecución, diferentes implementaciones finales de cada una de las estructuras de datos: la Lista Ordenada, por punteros simplemente enlazada y por arreglos; el Árbol de Búsqueda Binaria, por punteros y vector *heap*; el Trie, usando arreglos y punteros y la Tabla Hash, abierta.

ESPECIFICACIONES DEL PROYECTO:

El proyecto sólo podrá ser programado en Python o Pascal.

El Modelo Diccionario deberá tener las operaciones: *Init*, *Done*, *Clear*, *Insert* (no debe requerir que el elemento no exista), *Delete*, *Member* y *Print* (imprime todos los elementos del diccionario).

El tipo elemento de datos contenido en el diccionario será una *string* de letras de longitud 20. Las letras estarán en el rango: 'a'..'z' .

La función hash será la presentada en el libro de texto o una modificación de esta. Además, se deberá incluir una evaluación de la "aleatoriedad" o "uniformidad" de la misma.

La tabla hash deberá considerar el proceso de redistribución. Y su costo, en términos del tiempo de duración, deberá ser evaluado en forma independiente.

Se debe hacer una comparación para las siguientes parejas de estructuras de datos:

- L.O. por punteros vs L.O. por arreglos.
- A.B.B por punteros vs A.B.B por vector heap.
- Trie por punteros vs Trie por arreglos
- L.O. por punteros vs Tabla Hash (sin considerar redistribución)
- L.O. por punteros vs A.B.B. por punteros.

- L.O. por punteros vs Trie por punteros.
- Tabla Hash (sin considerar redistribución) vs Trie por punteros.

Tales comparaciones deberán incluir tiempos promedios y desviación estándar de hacer una inserción, un borrado y una búsqueda, en diccionarios pequeños (100 elementos), medianos (50 000 elementos) y grandes (1 000 000 de elementos). Además, debe hacer 10 corridas de cada uno. Las tiras de caracteres a insertarse, borrarse o buscarse serán generadas aleatoriamente. Además, deberá incluirse el tiempo de un *Print* y de un *Done*, para un diccionario grande.

Finalmente, debe presentar la cantidad de memoria utilizada por cada estructura de datos, para un diccionario pequeño, uno mediano y uno grande.

Tome en cuenta que sería interesante considerar la representación de diccionarios tales que, muchas tiras de caracteres comparten la misma parte prefija y tal parte prefija es muy larga o, por el contrario, diccionarios cuyas tiras de caracteres poseen partes prefijas muy distintas entre sí, o estas son muy pequeñas. Además, sería interesante considerar diccionarios en donde las inserciones no son del todo aleatorias, produciendo así, un A.B.B. desbalanceado.

Debe establecerse rangos para la cardinalidad del diccionario, es decir para N, que indiquen para qué valores de este, es más adecuado el uso de cada estructura de datos final.

Deben compararse los resultados de los tiempos obtenidos con los resultados teóricos (es decir, ordenes de duración).

Deben comentarse las limitaciones reales de lenguaje, herramientas de medición y máquina que restringen la investigación propuesta.

Para medir la cantidad real de espacio utilizada se puede utilizar alguna función de medición de memoria o algo similar y para medir los tiempos se pueden utilizar herramientas tales como algún perfilador.

El código fuente estará compuesto por 7 unidades (una para cada implementación final del Modelo Diccionario) y un programa (que servirá para probar y depurar cada implementación del Modelo y que luego se transformará en un programa que permita analizar tiempos y usar espacio).

El reporte final deberá contener: la especificación del Modelo (modelo y operaciones), la especificación de las 4 estructuras de datos genéricas y de las 7 estructuras de datos finales, los resultados de la evaluación (tiempos y espacio, mostrados en tablas y gráficos), las limitaciones, la comparación de los resultados reales con los teóricos, los rangos para N encontrados y una conclusión de la investigación realizada.

Este proyecto tarea podrá ser realizado por grupos de a lo sumo 2 personas y deberá ser entregado por etapas de la siguiente manera:

PRIMERA ETAPA

- especificación del Modelo Diccionario (modelo y operaciones)
- descripción de la E.D. genérica Lista Ordenada
- descripción de la E.D. Lista Ordenada por punteros simplemente enlazada
- descripción de la E.D. Lista Ordenada por arreglos
- implementación probada y depurada del Diccionario mediante una L.O. por punteros simplemente enlazada
- implementación probada y depurada del Diccionario mediante una L.O. por arreglos
- descripción de la E.D. genérica Tabla Hash
- descripción de la E.D. Tabla Hash abierta
- descripción de la función hash y la evaluación de su "aleatoriedad"
- descripción del proceso de redistribución y la evaluación de su tiempo de duración
- implementación probada y depurada del Diccionario mediante una Tabla Hash abierta
- el programa de prueba, el cual permitirá realizar las operaciones del diccionario mediante un menú

Fecha de entrega: 6 de octubre, 10:00 a.m.

SEGUNDA ETAPA

- descripción de la E.D. genérica Árbol de Búsqueda Binaria
- descripción de la E.D. A.B.B. por punteros
- descripción de la E.D. A.B.B. por vector *heap*
- implementación probada y depurada del Diccionario mediante un A.B.B. por punteros
- implementación probada y depurada del Diccionario mediante un A.B.B. por vector *heap*
- descripción de la E.D. genérica Trie
- descripción de la E.D. Trie por punteros
- descripción de la E.D. Trie por arreglos
- implementación probada y depurada del Diccionario mediante un Trie por punteros
- implementación probada y depurada del Diccionario mediante un Trie por arreglos
- el mismo programa de prueba anterior

Fecha de entrega: 20 de octubre, 10:00 a.m.

TERCERA ETAPA:

-Programa de análisis que permita obtener tiempos de duración y uso de espacio para las diferentes estructuras de datos y para diferentes tipos y tamaños de diccionarios.

-Reporte final: cuadros, gráficos, limitaciones, rangos para N, etc.

Fecha de entrega: 3 de noviembre, 10:00 a.m.

El código fuente y el código objeto deberá ser subidos a Mediación Virtual. La especificación del modelo, la descripción de las estructuras de datos y el reporte final también. Es responsabilidad de los estudiantes mantener un respaldo de los entregables.

Se podrá añadir o modificar especificaciones en clase según lo que se discuta ahí.

No existe posibilidad de ampliar los plazos para la entrega de cada una de las etapas, por lo tanto, es indispensable que se empiece a trabajar inmediatamente en la realización de este proyecto.