



Universidad de Costa Rica
Escuela de Ciencias de la Computación e Informática
Semestre I - 2021
Curso CI-0113 - Programación II
Profesor: Edgar Casasola Murillo
Fecha: Domingo 31 de julio 2022 11::55 pm

Tarea Programada 4

Comparación rendimiento map STL vrs ArbolRojoNegro

Objetivo de aprendizaje: Dominio de creación de contenedores reutilizables utilizando la técnica de Plantillas y comparación de rendimiento contra Biblioteca Estándar de Plantillas STL.

En esta tarea usted desarrollará una plantilla un árbol para almacenar pares llave-valor (Key,Value) , y será un árbol con inserción en la hojas manteniéndolo equilibrado con la técnica de balanceo de árbol Rojo-Negro. El árbol se debe comparar contra dos contenedores de la STL (un map y un vector con pares (Key,Value)) Si usted ordena el vector antes de buscar las palabras, el tiempo de búsqueda en el vector debe incluir el tiempo que tomó ordenarlo.

Deberá crear una plantilla de función para llevar a cabo las pruebas de rendimiento.

Para simplificar la tarea, haremos una comparación simple de la siguiente forma.

1. Cargar en un string[N] con las N palabras que utilizarán para cada ejecución del experimento. Esto dará igualdad de condiciones para cada uno de los contenedores en los diferentes experimentos. Las palabras serán tomadas en secuencia a partir de un archivo de texto que se les proporciona. Cada ejecución tomará cantidades diferentes según se indica en el punto 6.
2. En cada ejecución se deberá cargar en el contenedor que se evalúa, un par compuesto de una palabra y su siguiente palabra. Dicho de otra forma, para cada palabra se va a almacenar **solamente** “la primera aparición de una palabra que aparece inmediatamente después de ella en el texto”, si una palabra no aparece es porque no estaba en el texto, a la última palabra cargada en el vector se asocia una palabra vacía “”. Si una palabra ya está en el contenedor se descarta, porque ya tiene otra palabra que fue su primera aparición como consecutiva.
3. Para cada ejecución con su ArbolRN: Calcular el tiempo de búsqueda promedio de todas las palabras alternando cada palabra existente con una palabra inválida construida agregando la terminación “NolsE” a la palabra original. Lo que quiere decir es que se debe buscar una palabra que sí está y luego la palabra inválida construida que se sabe que no está. La búsqueda deberá seguir el orden en que aparecen en el vector de lectura, sin ordenar y sin eliminar repetidas. No se debe imprimir nada ni guardar en un archivo para que el tiempo de entrada/salida no afecte los cálculos del tiempo de búsqueda en el contenedor.
4. Con el map de la STL se repite el proceso del paso 3 (usar método emplantillado para tal efecto).
5. Con un vector de la STL que almacene pares del tipo (K,V) llave/valor, se repite el proceso del paso 3 (recuerde que si lo ordena previamente deberá sumar ese tiempo al tiempo total antes de promediar).

6. Escriba en pantalla el tiempo promedio de búsqueda de los tres contenedores al final de una ejecución con:
- A) Contenedores con 1.000.000 palabras y su primera palabra siguiente.
- A1) Con esos contenedores calcular los primeros 10 tiempos promedio de la siguiente forma: Inicie en 10.000 y terminando en 100.000, calcule el tiempo promedio cada 10.000 palabras y sus 10.000 palabras inválidas correspondientes.
- A2) Con esos contenedores calcular los primeros 10 tiempos promedio iniciando en 100.000 e incrementando en 100.000 palabras y sus 100.000 palabras inválidas correspondientes.
- La salida de cada experimento A1 y A2 será una tabla con 3 contenedores (filas) y 10 columnas de tiempos promedios para cada uno. En la documentación muestre un gráfico comparativo usando el graficador de alguna hoja de cálculo. (LibreOffice o Excel).
- B) OPCIONAL POR 10% EXTRA SOBRE NOTA DEL RUBRO DE TAREAS
- B1. ¿Puede hacer el experimento con el total de palabras en el archivo datos.txt? (39.151.187 PALABRAS)
- B2. Con búsquedas en bloques de 3.900.000 palabras válidas y sus correspondientes inválidas.
- B3. Caso contrario ¿con cuanto fue el máximo que su grupo pudo calcular los 10 tiempos promedio?

Se adjunta un archivo para pruebas llamado datos.txt disponible en:

<https://drive.google.com/file/d/1bjrg6J0kqdlGAghQrocppAqEe-wWUq-5/view?usp=sharing>