

Algoritmos y Estructuras de Datos

Proyecto Segundo Parcial – 2017

Conteo y repetición de palabras para comprimir un archivo de texto

Una forma simple de comprimir un archivo de texto es reemplazar cada palabra por un código binario que la represente. Por ejemplo, la palabra casa se reemplaza por el binario 1001 que ocupa 4 bit en lugar de 4 bytes. Para que la compresión sea efectiva, se debe reemplazar la palabra más repetida por un código mínimo, y así en forma creciente en el tamaño de código hasta la palabra menos repetida. Para ello es necesario contar cuantas veces se repite cada palabra en el texto.

La tarea consiste en contar cuantas veces se repite cada palabra. Para ello deberá leer un archivo de texto de al menos 500 palabras con al menos 100 palabras distintas e identificar cuáles son las palabras que lo forman y cuantas veces están repetidas cada una de ellas. El archivo es de libre elección.

Para contar las palabras deberá armar un árbol binario de búsqueda donde cada nodo contenga la palabra y la cantidad de veces que figura repetida. Al leer una palabra del archivo, deberá buscarla en el árbol. Si el mismo no existe, será agregada con el número de repetición 1. Si ya existiera, le sumará uno a las repeticiones. Deberá hacer dos rutinas similares para la construcción del árbol, una recursiva y la otra iterativa.

Al finalizar la lectura del archivo y tener completo el árbol, deberá indicar la altura del árbol creado y generar un listado de las palabras ordenadas de mayor a menor, según en número de repeticiones que tenga, indicando palabra y número de repetición, de forma tal que sea la entrada de la generación de los códigos de compresión. Deberá usar el algoritmo de inserción y para evitar la duplicación de datos, los nodos de la lista no deberán tener los datos de las palabras y su cantidad de repeticiones, sino tener punteros a los nodos del árbol con esa información. Generar un archivo con el listado ordenado de palabras.

Opción de Promoción

Agregar el armado del árbol como AVL y el ordenamiento del listado ordenado con algoritmo heap-sort. Generar un menú de opciones donde el usuario elija como arma el árbol y como lo ordena. Obtener la altura del árbol binario de búsqueda y del AVL, e informar la cantidad de comparaciones realizadas en el ordenamiento por cada algoritmo.