

Algoritmos y Estructuras de Datos

Proyecto Segundo Parcial – 2019

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 cuales 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 contar con un nodo para cada palabra distinta del archivo que sea simultáneamente un nodo de lista y un nodo de árbol binario de búsqueda (abb), es decir, que contenga tres punteros, uno al siguiente nodo de la lista y dos para el abb, generando una estructura combinada de lista con abb. Además, el nodo tendrá como valor a la palabra y al contador de repeticiones que esta tenga. Para un ordenamiento adicional por cantidad de repeticiones necesitará un cuarto puntero a lista.

Al leer una palabra del archivo, debe verificar si ya existe previamente, en cuyo caso solo debe incrementar el contador de repetición de la palabra. Si no existe, debe crear un nuevo nodo e insertarlo en la estructura combinada, en lista, en forma ordenada alfabéticamente por medio del algoritmo de inserción y en el abb, cumpliendo sus propiedades.

Al finalizar la lectura del archivo deberá generar un ordenamiento por la cantidad de repeticiones de mayor a menor utilizando el algoritmo de quicksort sobre los nodos de la lista sin alterar el ordenamiento alfabético, utilizando el cuarto puntero de cada nodo. En resumen, de los cuatro punteros del nodo se usarán dos para el abb, uno para el ordenamiento alfabético y el último para el ordenamiento por repeticiones.

Deberá llevar un registro de la cantidad de comparaciones que necesitó para ordenar por inserción y por quicksort y generar un archivo con el listado ordenado de palabras según las repeticiones y el orden alfabético.

Opción de Promoción

Agregar el armado del árbol que cumpla con el balanceo del algoritmo AVL. Generar un menú de opciones donde el usuario elija como arma el árbol. Obtener la altura del árbol binario de búsqueda y del AVL.