

Universidad del Valley de Guatemala

Algoritmos y Estructura de datos – Sección 40

Juan Gualim – 24852

Miguel Rosas – 241074

Informe algoritmo para comprimir y descomprimir un archivo de texto utilizando un árbol de Huffman

Implementación del árbol de Huffman

El árbol se construye a partir de la frecuencia de cada carácter que se encuentra en el archivo de entrada. Se utiliza una cola de prioridad para asegurar que los nodos con menor frecuencia sean combinados primero. Cada nodo del árbol contiene un carácter para las hojas y una frecuencia acumulada para nodos internos. Este árbol genera los códigos binarios únicos y sin prefijos para cada carácter, lo que permite optimizar el espacio de almacenamiento.

Diseño de algoritmo de compresión:

- Se lee el archivo original y se calcula la frecuencia de cada carácter.
- A partir de estas frecuencias, se construye el árbol de Huffman.
- Luego se asigna a cada carácter un código binario mediante el recorrido del árbol.
- El texto original se codifica reemplazando cada carácter por su código.
- La cadena de binarios de resultado se guarda como archivo .huff.
- El árbol se guarda en archivo .hufftree usando un recorrido preorder.

Diseño algoritmo de compresión:

- Se reconstruye el árbol leyendo el archivo .hufftree.
- Se leen los bits del archivo .huff y se decodifican recorriendo el árbol desde la raíz hasta una hoja.
- Al llegar a una hoja, se recupera el carácter y se vuelve a la raíz.
- El proceso se repite hasta completar todos los bits significativos.

Decisiones clave para el diseño:

Se decidió almacenar la longitud original del texto en bits para evitar errores por ceros de relleno, primero si se guardaron ceros de rellenos y provocó que apareciera un carácter al “o” al final. También se guardan los caracteres especiales como secuencias de escape (`\n`, `\t`), ya sea tabulación o salto de línea, esto permitió mantener la estructura original del documento, ya que antes no se guardaban los saltos de línea o las tabulaciones, lo que hacía poco legible el documento.