

INSTITUTO TECNOLÓGICO DE COSTA RICA

SEGUNDO PROYECTO PROGRAMADO

Ariel Herrera

Saúl Zamora

profesor

M. Sc. Saúl Calderón Ramírez

September 5, 2016

I. INTRODUCCIÓN

En la actualidad, los sistemas de almacenamiento y comunicación digitales requieren de métodos optimizados para el uso de recursos (energéticos, temporales, de espacio, etc). Para satisfacer tal necesidad de manera efectiva, muchas disciplinas han formulado múltiples algoritmos para comprimir y descomprimir la información. La compresión de datos consiste en aplicar algún método que permita reducir el tamaño original de la información. Los algoritmos de compresión sin pérdida son capaces de aplicar una serie de pasos para construir la información comprimida, para luego, cuando la información original necesite ser accesada, se descomprime y recupera la información original completamente idéntica. Los algoritmos de compresión con pérdida en cambio, al implementar la descompresión de la información, no logran recuperar el 100% de los datos originales. El algoritmo de Huffman implementado en este proyecto fue propuesto por David A. Huffman en 1952 enfocado en la compresión sin pérdida de datos.

II. ANÁLISIS DEL PROBLEMA

La técnica de Huffman trabaja al crear un árbol binario de nodos, los cuales pueden ser hojas o nodos internos. Al principio, todos empiezan como hojas, las cuales contienen un símbolo, el peso (*frecuencia*) es opcional, y un enlace al nodo padre, lo cual facilita leer el código comenzando de las hojas. Los nodos internos contienen el peso del símbolo, dos enlaces a nodos hijos y un enlace opcional a un nodo padre.

Como una convención, el bit 0 representa el siguiente hijo izquierdo y el bit 1 el siguiente hijo derecho. Un árbol terminado puede crecer hasta tener (n) hojas y ($n - 1$) nodos internos. Un árbol de Huffman que omite los símbolos que no se usan, produce el código con el largo óptimo.

El proceso inicia con las hojas conteniendo los símbolos a representar, luego un nuevo nodo es creado con los nodos con menor probabilidad como hijos, tal que la probabilidad de dicho nodo es igual a la suma de las probabilidades de sus hijos. Con los nodos anteriores mezclados en uno (ya no son considerados), y considerando al nuevo nodo, el proceso se repite hasta que solo quede un nodo: el árbol de Huffman.

III. DISEÑO DE LA SOLUCIÓN

IV. PRUEBAS

V. REFERENCIAS

REFERENCES

- [1] Mamta Sharma. Compression using Huffman coding. *IJCSNS International Journal of Computer Science and Network Security*, 10(5):133141, 2010.